Effects of Belief and Logic on Syllogistic Reasoning

Eye-Movement Evidence for Selective Processing Models

Linden J. Ball,[1] Peter Phillips,[1] Caroline N. Wade,[1] and Jeremy D. Quayle[2]

[1]Lancaster University, UK, [2]Derby, Derbyshire, UK

Abstract. Studies of syllogistic reasoning have demonstrated a nonlogical tendency for people to endorse more believable conclusions than unbelievable ones. This *belief bias* effect is more dominant on invalid syllogisms than valid ones, giving rise to a logic by belief interaction. We report an experiment in which participants' eye movements were recorded in order to provide insights into the nature and time course of the reasoning processes associated with manipulations of conclusion validity and believability. Our main dependent measure was people's inspection times for syllogistic premises, and we tested predictions deriving from three contemporary mental-models accounts of the logic by belief interaction. Results supported recent "selective processing" theories of belief bias (e.g., Evans, 2000; Klauer, Musch, & Naumer, 2000), which assume that the believability of a conclusion biases model construction processes, rather than biasing the search for falsifying models (e.g., Oakhill & Johnson-Laird, 1985) or a response stage of reasoning arising from subjective uncertainty (e.g., Quayle & Ball, 2000). We conclude by suggesting that the eye-movement analyses in reasoning research may provide a useful adjunct to other process-tracing techniques such as verbal protocol analysis.

Keywords: belief bias, syllogistic reasoning, eye-movement analysis, inspection times, process tracing

Syllogisms are deductive reasoning problems that involve two premises and a conclusion, for example: "Some artists are beekeepers; No beekeepers are carpenters; *Therefore*, some artists are not carpenters." The premises feature three terms: two *end* terms (one per premise) and a *middle* term (appearing in both premises). A logically valid conclusion describes the relationship between the end terms in a way that is *necessarily* true. Statements that are merely consistent with the premises, but not necessitated by them, are invalid. Participants are required either to produce a logical conclusion from given premises or to evaluate the validity of a presented conclusion; the present research employs a conclusion-evaluation paradigm.

It has been found that responses vary systematically according to three main factors. One factor, termed *figure*, refers to the arrangement of terms within premises. There are four possible figures: A-B, B-C; A-B, C-B; B-A, C-B; and B-A, B-C (where A and C refer to the end-terms in the first and second premises, respectively, and B refers to the middle term). A second factor, *mood*, denotes the different quantifiers within the premises and conclusion. The four standard quantifiers are referred to using letters: A = *all*, E = *no*, I = *some*, and O = *some … are not*. The example syllogism above is an AB-BC IEO problem. Studies have revealed that different combinations of figure and mood have a marked effect on reasoning accuracy (e.g., Bara, Bucciarelli, & Johnson-Laird, 1995; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991).

The third factor influencing syllogistic reasoning concerns people's prior knowledge, and gives rise to a phenomenon termed *belief bias*, which forms the focus of the present research. There are, in fact, three key findings that derive from studies where the validity of syllogistic conclusions has been manipulated alongside conclusion believability (see Evans, Newstead, & Byrne, 1993, for a review). First, believable conclusions (e.g., *Some addictive things are not cigarettes*) are more readily endorsed than unbelievable ones (e.g., *Some millionaires are not rich*). Second, valid conclusions are more readily endorsed than invalid ones. Third, there is an interaction between logic and belief such that the effects of believability are

DOI: 10.1027/1618-3169.53.1.77

Table 1. Percentage of conclusions accepted as a function of logic and conclusion believability in (a) the three experiments reported by Evans et al. (1983), and (b) the present experiment.

a) Evans et al. (1983)	Believable	Unbelievable	*M*
Valid	89	56	72
Invalid	71	10	40
M	80	33	
b) Present Experiment			
Valid	75	63	69
Invalid	53	6	30
M	64	34	

Note. The Evans et al. (1983) results reported here have been taken from Evans, Newstead, and Byrne (1993).

stronger on invalid than valid problems (Table 1 presents typical data).

Several accounts have been forwarded to explain the processes underpinning believability and validity effects in syllogistic inference. It is the aim of the present paper to help arbitrate between these competing theories using inspection-time data derived from people's eye movements whilst they tackle syllogistic problems. There is a considerable literature dealing with factors that determine the acceptability of syllogistic conclusions, but few studies have employed time-based measures (but see Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003) or process-tracing techniques (an exception being use of think-aloud protocols by Evans, Barston, & Pollard, 1983,). Analyzing aspects of on-line attentional processing using eye-movement tracking seems to have the potential to offer valuable data relevant to assessing belief-bias theories.

As a method for studying cognition, eye-movement tracking has been used extensively in reading research, where it is claimed there is a close association between visual fixations and the processes underlying text understanding (Liversedge, Paterson, & Pickering, 1998). The position of a fixation is taken to indicate the text currently being processed, and the length of a fixation is assumed to indicate ease of processing. The validity of an association between attentional focus and eye movements appears to be supported by a large body of research (Rayner, 1999). Although it is possible to move attention covertly 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 (He & Kowler, 1992; Sclingensiepen, Campbell, Legge, & Walker, 1986). Moreover, evidence suggests that attention *precedes* a saccade to a new location (Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995) and that attentional shifts and saccades are obligatorily linked (Deubel & Schneider, 1996).

Following the assumptions that pertain to reading research, it is possible that monitoring participants' eye movements during visually-based deduction could likewise provide insights into the nature and organisation of reasoning processes. We acknowledge that the association between eye-movements and the high-level thought processes underlying display-based reasoning and problem-solving is not well established in the literature and is, therefore, potentially contentious. We are encouraged, however, by the recent increase in published research using eye-movement tracking to study processes such as insight problem solving (e.g., Knoblich, Ohlsson, & Raney, 2001) and reasoning (e.g., Underwood, Jebbett, & Roberts, 2004), and the broadly converging evidence that such studies have produced for established theories. Our own eye-tracking research on Wason's four-card selection task (Ball, Lucas, Miles, & Gale, 2003) has similarly suggested that this novel method is sensitive to detecting differences in processing times for problem components that are predicted by contemporary reasoning theories.

If a link between eye-movements and reasoning processes can be accepted, then it follows that the analysis of eye-movements during reasoning may afford some advantages over other process-oriented techniques such as think-aloud protocol analysis. Since eye movements are fast and spontaneous they should provide a detailed record of the sequence and organisation of processing, thereby improving the completeness of the trace of the locus of attention when compared with thinking aloud. What, though, might eye-movement analyses reveal about the processes underlying belief bias effects? We return to this question in more detail below. Suffice it to say that our working assumption is that differences in the *inspection times* associated with aspects of syllogisms should be informative as to the processing demands of different problems. For example, syllogisms that are assumed to be more demanding to represent and process during conclusion evaluation would be expected to be associated with increased inspection times relative to less demanding syllogisms. Similarly, the additional scrutiny that some theorists claim is applied to premises when evaluating *unbelievable* conclusions should be detectable in eye movements in the form of longer inspection times for premise information.

The Mental Models Account of Belief Bias

One dominant explanation of belief bias effects is that proposed by Oakhill and Johnson-Laird (1985; Oakhill, Johnson-Laird, & Garnham, 1989), which is situated within Johnson-Laird's (e.g., 1983) general mental models account of syllogistic reasoning with *abstract* material. This general theory assumes that people begin reasoning by constructing an integrated representation of the premises (i.e., a *model*) that makes explicit a minimal amount of information concerning the logical relationships between premise terms. People then draw a conclusion that is consistent with this information, or, in the case where a conclusion is given, they evaluate whether the conclusion is consistent with this initial model. Finally, they test the consistency of the conclusion against alternative models of the premises. If a conclusion is found to be inconsistent with any model it is rejected, otherwise it is accepted. Some syllogisms (*one-model* problems) are relatively easy because they require the construction of only a single model to determine the conclusion's validity, thus placing a minimal load on working memory resources. Other syllogisms (*multiple-model* problems) are difficult because they place higher loads on working memory, requiring the construction of two or three models to guarantee a conclusion's validity.

The idea that syllogistic difficulty is closely associated with the number of models that need to be constructed has received empirical support (e.g., Bara et al., 1995; Johnson-Laird & Bara, 1984). The mental models account of *belief bias* assumes that prior beliefs are influential in determining whether or not participants will construct alternative models of the premises when testing a conclusion's validity. Conclusions that are true in an initial model will simply be accepted if they are *believable*, but will be tested meticulously against alternative and potentially falsifying models if they are *unbelievable*. In this way, the occurrence of a logic by belief interaction with multiple-model syllogisms is explained. This theory also explains why a logic by belief interaction is often not observed with one model problems (e.g., Newstead, Pollard, Evans, & Allen, 1992), as these problems have no counterexample models such that the construction of an initial model should be sufficient to determine a correct response (but see Gilinsky & Judd, 1994, for opposing evidence).

The models account of belief bias is a form of *selective-scrutiny model* (see Evans et al., 1983) – or what Roberts and Sykes (2003) term a *selective falsification* account – as it is only unbelievable conclusions that enhance people's reasoning by motivating a more rigorous logical analysis of the syllogistic argument. This account seems readily to lend itself to the derivation of inspection-time predictions in relation to belief-bias effects with multiple-model problems. In particular, reasoners should inspect the premises of problems with unbelievable conclusions reliably *longer* than those with believable conclusions. This is because reasoners will be more likely to search for counterexample models in the former case, which would, presumably, require premise re-reading to refresh and revise the mental representation of the information they contain.

In addition to a difference in *total premise inspection time* between unbelievable and believable problems, it is also possible to make a more detailed inspection-time prediction about the precise temporal location of the expected inspection-time effect, that is, it should arise *after* a conclusion has been inspected (what we subsequently refer to as the *post-conclusion-inspection stage*). This is because the increased processing that people are claimed to engage in for certain problems is, according to the models account, directly contingent on the believability status of this conclusion.

The Metacognitive Uncertainty Account of Belief Bias

A fundamental assumption of Johnson-Laird's general models account of syllogistic inference is that limited working memory capacity constrains the number of models that can be manipulated during reasoning. What is curious about the specific models account of belief bias, however, is the lack of any mention of working memory constraints as a limiting factor in reasoning with belief-oriented materials. In an attempt to address this apparent omission, Quayle and Ball (e.g., 1997, 2000) forwarded an alternative models-based account of belief bias that places limited processing capacity at the centre of belief-bias phenomena. They propose that the tendency to respond in accordance with belief is determined by the relative demands placed on working memory by different *types* of syllogism. Syllogisms that place unmanageable loads on working memory result in participants being unable to test the truth of putative conclusions against models of the premises. In this instance, it is argued that participants will be *uncertain* of a conclusion's logical status, and will default to a belief-based response.

This so-called *metacognitive uncertainty* explanation of the logic by belief interaction in the conclusion-evaluation paradigm hinges on the observation that with multiple-model syllogisms, correct responses can be given after construction of a single mental model for valid problems, whilst invalid problems require consideration of more than one model (cf. Garnham, 1993; Hardman & Payne, 1995).

Quayle and Ball (2000) argue that invalid problems will, therefore, place greater demands on working memory than valid ones, often causing participants to fall back on a belief heuristic (so explaining the belief by logic interaction). Quayle and Ball (2000) also provide empirical support for both a distinction between the processing demands of valid versus invalid conclusions, and the idea that the logic by belief interaction can be attributed to working memory limitations. They showed that: (1) people were less *confident* in responses to invalid than valid conclusions, and (2) the logic by belief interaction disappeared with those possessing superior working memory capabilities, whilst remaining dominant with those having inferior working memories.

The metacognitive uncertainty account is closely allied to a previous theory of belief bias proposed by Evans et al. (1983), the *misinterpreted necessity* model. This model likewise emphasises uncertainty as a determinant of belief bias, with people defaulting to a belief-based decision in cases where a conclusion is *possible* but not necessitated by the premises (i.e., when the syllogism is *indeterminately invalid*). The key differences between this account and the metacognitive uncertainty account are that the latter specifically invokes a mental models mechanism as an explanation of people's basic reasoning competence, and additionally clarifies the importance of a valid-invalid processing distinction in models-based reasoning.

Like the standard mental models theory, the metacognitive uncertainty account also lends itself to inspection-time predictions. The concept of a valid-invalid processing distinction suggests that people should inspect the premises of problems with invalid conclusions reliably *longer* than those with valid conclusions, as the former require additional processing before reasoners can ascertain the conclusion's logical status. In addition to a difference in total premise inspection time between invalid and valid problems, it is also possible to predict the temporal locus of the inspection-time effect: It should arise in the post-conclusion-viewing stage, as at this point the logical status of valid conclusions is not in question but is still undecided for invalid conclusions. These predictions are distinct from those of the standard models theory of belief bias.

Selective Processing Accounts of Belief Bias

In a recent, influential paper on belief bias, Klauer, Musch, and Naumer (2000) presented a multinomial threshold model in an attempt to disentangle the respective contributions of reasoning processes and response biases to conclusion acceptance. Their meta-analysis of 22 previous studies revealed that existing data were structurally too sparse to enable discrimination between different theories of belief bias such as the mental models account and the misinterpreted necessity account. Klauer et al. therefore presented four new studies, conducted to provide richer data to enable more penetrating tests using the multinomial model. None of the aforementioned theories, however, were consistent with the pattern of results that derived from the studies of Klauer et al. (note that although the metacognitive uncertainty model was not explicitly assessed, its equivalence to the misinterpreted necessity model in terms of parameter values would render it similarly deficient in accounting for the data solely via a response-bias mechanism).

In addressing the limitations of current accounts of belief bias, Klauer et al. present a new theory whereby people typically construct only a *single* mental model when evaluating syllogisms, with the believability of a given conclusion biasing the process of *model construction* rather than influencing the search for alternative models. Thus, it is assumed that participants perform: (1) *positive* (confirmatory) tests for believable conclusions (attempting to construct a model in which the premises and conclusion hold true) which, if possible, will lead to conclusion acceptance (see Cherubini, Garnham, Oakhill, & Morley, 1998); and (2) *negative* (disconfirmatory) tests for unbelievable conclusions (attempting to construct a model of the premises that excludes the conclusion and integrates its logical opposite), which, if possible, will lead to conclusion rejection. This account explains the large belief effects observed for indeterminately invalid conclusions, since for these there are potentially both confirmatory and disconfirmatory models, and either can be found depending on what is sought. A final, important component of the account of Klauer et al. (their "Assumption 5") is that reasoners will have considerable difficulty in forming an integrated semantic representation of an unsound set of premises, defined as a set of premises that logically entails an unbelievable conclusion, and, in particular, when strong conflicts with prior knowledge arise in the process of integrating premises.

The theory of Klauer et al. (2000) gains good support from a series of innovative experiments they present (note, too, that a very similar theory was independently proposed by Evans, 2000 – see also Evans, Handley, & Harper, 2001 – along with compelling supporting data). What, then, would these recent *selective processing* theories, using the Evans' (2000) terminology, predict in relation to inspection-time effects for multiple-model syllogisms in a conclusion-evaluation paradigm? Of critical relevance here is Assumption 5 of Klauer et al., which typically applies to syllogisms where logic and belief are in conflict, that is, to valid syllogisms with unbelievable conclusions and to invalid syllogisms with believable conclusions. Because people will find it difficult to construct coherent semantic representations in these cases, then increased total premise inspection time should ensue as reasoners strive to develop integrated premise models. In contrast, constructing a model of the premises for valid and believable and invalid and unbelievable conclusions should be relatively easy. Overall, this would lead to the prediction of an interaction between logic and belief in total premise inspection times – as distinct from the predictions pertaining in the case of the standard models-based theory or the metacognitive uncertainty theory.

It is also possible to derive a detailed prediction about the temporal locus of the inspection-time effect from the selective processing accounts of Evans et al. (2001) and Klauer et al. (2000). The account of Evans et al. (2001; see their Figure 5) explicitly details a *backward reasoning* mechanism that is driven directly by the presented conclusion (see Morley, Evans, & Handley, 2004, for evidence for such backward processing). The idea that people reason from conclusions to premises means that the predicted logic by belief interaction in premise inspection times should dominate at the post-conclusion inspection stage. Klauer et al. (2000) acknowledge that a backward reasoning process would explain the strong conclusion-based focus observed in think-aloud data (Evans et al., 1983), although they maintain that an order assumption is not required to account for their data. Nevertheless, obtaining inspection-time evidence from the analysis of eye-tracking data would enable a useful clarification concerning this important theoretical issue.

Method

Participants

Sixty-eight students at Lancaster University participated in the experiment for payment. None had received formal instruction in logic, and all were tested individually.

Materials

While participants tackled computer-presented syllogisms their eye movements were recorded using an LC Technologies EyeGaze system that determines gaze direction using the pupil-centre/corneal-reflection method. The tracker consists of a standard desktop computer running Windows NT/2000, an infrared camera mounted beneath the monitor, and software to process the resulting data. The tracker is accurate to within 0.45° of visual angle, which, at 50 cm from the screen, covers approximately 0.38 cm (12.8 pixels). Eye movements are sampled at 60 Hz. Although the tracker can tolerate head motion of around 3 cm in all directions, participants used a chin-rest to minimise data loss. Fixations were detected at a conventional threshold of 100 ms or above.

Two forms of three-model syllogisms were used in the A-B, B-C figure with A-C conclusions. One form was valid and in the IEO mood, the other was invalid and in the EIO mood. Invalid conclusions were indeterminate (i.e., consistent with premises but not necessitated by them). A set of potential conclusions which were false by definition (e.g., "Some tulips are not flowers") was chosen, together with a set of believable conclusions (e.g., "Some animals are not cats"). The conclusions were devised to be believable when terms were in one order, but unbelievable with reversed terms. To assess believability a larger set of potential conclusions was pre-rated by 30 participants on a seven-point scale ranging from –3 (*totally unbelievable*) to +3 (*totally believable*). Conclusions receiving the most extreme and consistent ratings were used in the study (see Table 2).

Half of the valid and invalid syllogisms were presented with believable conclusions, and half with unbelievable conclusions. Three, valid one-model syllo-

Table 2. Mean believability ratings of conclusions used (B = believable; U = unbelievable).

		M	*SD*
B	Some animals are not cats	2.90	0.41
U	Some cats are not animals	–2.43	1.53
B	Some metals are not steel	2.77	0.57
U	Some steel is not metal	–2.47	1.16
B	Some flowers are not tulips	2.87	0.35
U	Some tulips are not flowers	–2.23	1.74
B	Some men are not kings	2.47	1.25
U	Some kings are not men	–2.33	1.39
Filler Items			
B	All chemists are scientists	0.90	2.02
B	Some athletes are women	2.80	0.41
U	No popes are catholic	–2.30	1.84

Note. All the items were rated on a seven-point scale that ranged from –3 (*totally unbelievable*) to +3 (*totally believable*).

gisms (two with believable conclusions, one with an unbelievable conclusion) acted as filler items and were in the A-B, B-C figure, and AAA, AEE and IAI moods. All syllogisms were presented individually with a good degree of screen separation between syllogistic components. The response words "yes" and "no" appeared as clickable radio buttons below each syllogism, and were preceded by the following reminder: "Does the conclusion above necessarily follow from the first two statements?."

Design

A within-participants design was used with two factors: logic (valid vs. invalid conclusions), and believability (believable vs. unbelievable conclusions). The block of target and filler syllogisms was preceded by a single practice syllogism with a believable conclusion. The order of the problem types was varied using a four by four balanced Latin square design, with the restriction that the filler items appeared in the same (i.e., 2nd, 4th, and 6th) position after the practice item for all participants. The thematic content of syllogisms was rotated over the different problem types, producing four sets of materials. These sets were distributed evenly and randomly amongst participants.

Procedure

Each participant sat approximately 50 cm in front of the display screen. Once the eye-tracker had been calibrated, the following on-screen instructions were presented:

This is an experiment to test people's reasoning ability. You will be given eight problems in total. For each problem you will be shown two statements and you are asked if a certain conclusion (given below the statements) may be logically deduced from them. You should answer this question on the assumption that the two statements are, in fact, true. If, and only if, you judge the conclusion necessarily follows from the statements, you should select "yes" using the mouse, otherwise select "no." Please take your time and be sure you have the right answer before giving your response. You can then click on the button labelled "next" to obtain another problem.

Having read the instructions participants self-paced their progression through the eight computer-presented problems, with eye-movements being concurrently monitored and logged. Once participants had left the experimental laboratory the raw eye-movement data were processed to permit: (1) definition of specified "regions of interest" (i.e., bounded areas around syllogistic components); (2) correction for patterns of eye-movement drift for each target syllogism, where necessary; and (3) extraction of component-based inspection-time measures (based on cumulative fixation times within regions of interest), to be used in subsequent analyses. An alpha level of .05 was set for all statistical tests.

Results and Discussion

Conclusion Acceptances

The mean percentage frequencies of participants endorsing conclusions are presented in Table 1 for each type of syllogism, and indicate standard belief-bias effects. Wilcoxon signed-ranks tests revealed significant effects of believability, $z = -4.83$, $p < .0005$, two-tailed, with participants accepting more believable than unbelievable conclusions, and Logic, $z = -5.35$, $p < .0005$, two tailed, with more valid than invalid conclusion being endorsed. There was also a significant interaction between logic and believability, $z = -3.33$, $p = .001$, two tailed, such that the effect of believability was greater on syllogisms leading to invalid than valid conclusions.

Total Premise Inspection Times

Our first analysis explored the effects of logic (valid vs. invalid conclusions) and believability (believable vs. unbelievable conclusions) on *total premise inspection times* (see Table 3) using a repeated-measures ANOVA. An examination of the distributions of inspection times within our various experimental conditions revealed that the data were slightly positively skewed, though typically within the +2.5 cut-off recommended by Tabachnick and Fidell (1996) as being legitimate for parametric analysis with small samples. As a precautionary measure, however, we pursued analysis of variance on both the natural data and the log-transformed data (subsequent to the addition of a constant of 1.0). Both the untransformed and transformed analyses produced an identical pattern of significant and nonsignificant effects, with similar effect magnitudes, although the transformed data typically yielded statistical findings that were even more conclusive. For ease of interpretation, however, we only report the results for the untransformed analyses throughout this paper.

In terms of the mental models prediction that unbelievable conclusions should lead to more premise processing than believable ones, the analysis revealed a significant effect, but in the *opposite* direction to that predicted, $F(1, 67) = 4.10$, $MSE = 57.852$, $p = .047$. As for the metacognitive uncertainty account, there was no evidence for a predicted effect of logic on this processing measure, $F(1, 67) = 1.07$, $MSE = 42.438$, $p = .304$. Finally, the Logic × Believability interaction predicted by selective processing accounts was seen to be highly reliable, $F(1, 67) = 10.42$, $MSE = 41.875$, $p = .002$, and was precisely as expected (i.e., longer processing times for the premises of invalid believable and valid unbelievable conclusions relative to the other problem types).

Although we had made no a priori predictions concerning *overall inspection times* for different syllogisms (i.e. premise inspection times plus conclusion inspection times), our eye-movement data enabled an ANOVA to be conducted to examine effects of Logic and Believability on global processing effort. The analysis revealed effects that were equivalent to those observed with the pure premise inspection-time measure: The effect of Logic was nonsignificant, $F(1, 67) = 2.57$, $MSE = 75.083$, $p = .114$, whereas the effect of believability and the Logic × Believability interaction were reliable, $F(1, 67) = 6.49$, $MSE = 102.952$, $p = .013$, and $F(1, 67) = 11.33$, $MSE = 70.793$, $p = .001$, respectively.

Premise Inspection Times Arising Subsequent to Conclusion Viewing

Our second analysis aimed to clarify the temporal locus of belief-based effects on reasoning by determining whether viewing the conclusion influenced the amount of premise processing that subsequently occurred. To examine this issue, we derived a *post-conclusion-viewing* measure of premise inspection times (see Table 3), and pursued a two-way ANOVA on the resulting data using the factors logic and believability.

Table 3. Mean inspection times, in seconds, for premises, conclusions, and overall, as a function of logic and believability

	Believable		Unbelievable	
Inspection Time	Valid	Invalid	Valid	Invalid
Premises: Pre-conclusion	4.27	4.27	4.12	4.03
Premises: Post-conclusion	6.36	9.71	7.21	5.56
Premises: Total	10.62	13.97	11.33	9.57
Conclusion	4.37	6.14	4.00	3.97
Overall	14.99	20.11	15.32	13.54

Notes. Pre-conclusion denotes the time spent inspecting the premises prior to the first fixation upon the conclusion, and *Post-conclusion* denotes the time spent inspecting the premises after the first fixation upon the conclusion.

From the perspective of the standard models theory, people should engage in increased processing of premises to unbelievable relative to believable conclusions during a post-conclusion viewing stage, since with the former they need to flesh out initial premise representations during the search for falsifying models, whilst with the latter they are posited merely to accept the conclusions as valid. Although the ANOVA revealed a marginally significant effect of believability on this inspection-time measure, $F(1, 67) = 3.35$, $MSE = 55.161$, $p = .072$, the effect was, again, in the inverse direction to that predicted by models theory (believable conclusions promoted increased processing during post-conclusion viewing relative to unbelievable conclusions). As for the metacognitive uncertainty account, this hinges on a processing distinction between the demands of valid versus invalid syllogisms, and predicts increased processing times for the premises of invalid problems on the post-conclusion-viewing measure relative to valid problems. The ANOVA, however, revealed no logic effect, contrary to the predictions of this account, $F(1, 67) = 1.401$, $MSE = 34.705$, $p = .241$.

The selective processing model of Evans et al. (2001; see also Morley et al., 2004) is clear in specifying the dominance of backward processing (i.e., conclusion-to-premise reasoning) in the syllogistic evaluation paradigm, and would, therefore, predict a Logic × Believability interaction arising at the post-conclusion viewing stage. Although the selective processing model of Klauer et al. (2000) is neutral as to the role of backward versus forward processing, this theory could readily accommodate evidence either way. As it transpired, the ANOVA revealed a highly reliable Logic × Believability interaction arising on the post-conclusion-viewing measure, $F(1, 67) = 12.12$, $MSE = 35.128$, $p = .001$. As is evident from Table 3, it is those problems where conclusion validity and believability are in conflict that give rise to increased premise processing at the post-conclusion-viewing stage relative to problems where validity and believability coincide. It is also evident from Table 3 that premises are only inspected a minimal amount prior to the conclusion being viewed; indeed, quantitative aspects of premise processing during this pre-conclusion-viewing stage are effectively equivalent across problems types. An ANOVA conducted on premise inspection times arising prior to the first viewing of the conclusion revealed no effect of believability, $F(1, 67) = 0.546$, $MSE = 4.917$, $p = .463$, or Logic, $F(1, 67) = 0.016$, $MSE = 6.251$, $p = .889$, and, importantly, no Logic × Believability interaction, $F(1, 67) = 0.016$, $MSE = 6.825$, $p = .901$. This latter null effect – when viewed alongside the Logic × Believability interaction evident in the post-conclusion-viewing measure – suggests that theories emphasising conclusion-to-premise reasoning in the evaluation paradigm may provide a better fit to the data than those that characterise processing as involving a forward reasoning mechanism.

General Discussion

Klauer et al. (2000) suggest that further research on belief bias effects in syllogistic inference might benefit from employing process-tracing methodologies (e.g., think-aloud protocol analysis) to provide more direct tests of the assumptions of current theories, including their own account. The present experiment progressed the suggestion of Klauer et al. in the novel direction of deriving eye-movement measures of online attentional processing whilst people tackled computer-presented syllogisms involving belief-oriented conclusions. Eye-movement tracking is a methodology that has been favoured by many reading researchers, where it is claimed that fixation durations on words are highly informative of the psycholinguistic processes driving text comprehension (e.g., Rayner, 1999). Few studies, however, have monitored eye-movements with the aim of advancing an understanding of reasoning mechanisms, despite the potential for this method to reveal detailed and important information about the nature and dynamics of inferential processes (see Ball et al. 2003).

The present experiment involved monitoring the cumulative duration of people's fixations on different syllogistic components (i.e., premises and conclusions), with the key assumption being that increased inspection times on such components are a reflection of increased processing effort (e.g., aimed at deriving coherent representations of premise information). Our rationale for pursuing such inspection-time measures was that they should allow us to arbitrate between different predictions that are derivable from three contemporary theories of belief bias in syllogistic reasoning.

First, the standard models account (e.g., Oakhill & Johnson-Laird, 1985; Oakhill et al., 1989) predicts that reasoners should inspect the premises of problems with unbelievable conclusions for longer than those with believable conclusions. This is because unbelievable conclusions are supposed to motivate a more penetrating logical analysis of premises than believable conclusions. Moreover, such an inspection-time imbalance should be especially marked during the period *after* the conclusion has first been examined, since increased processing of problems with unbelievable conclusions should be directly contingent upon the reasoner's assessment of the conclusion's believability status. Our inspection-time results, however, provided no support for these models-based predictions, with the evidence showing processing effects in the opposite-to-predicted direction. Such findings seem to be difficult to reconcile with a standard models account of belief bias.

Second, the metacognitive uncertainty account (Quayle & Ball, 2000) – a development of the misinterpreted necessity model of Evans et al. (1983) – proposes that belief-based responding is a default strategy that prevails with invalid problems because of their inherent processing demands, which leads to a state of subjective uncertainty (see Klauer et al., 2000, for extensive discussion of response-bias accounts of belief bias based around the concept of reasoning uncertainty). The metacognitive uncertainty account of Quayle and Ball predicts that there should be an inspection-time imbalance between the premises for valid and invalid conclusions, with the latter promoting increased processing effort (especially at a post-conclusion-viewing stage) as they require consideration of more mental models than the former (see Quayle & Ball, 2000). Again, however, our inspection-time data failed to support the predictions of this account: No reliable inspection-time imbalance was identified between the premises of valid and invalid problems, either overall or at a post-conclusion-viewing stage of processing. The lack of evidence for a valid-invalid processing distinction undermines a central tenet of the metacognitive uncertainty account, and suggests that a pure response-bias mechanism engendered by states of subjective uncertainty is unable to account fully for the logic by belief interaction in syllogistic inference (see Klauer et al., 2000).

The third perspective on belief bias that we examined was the selective processing framework advanced by Klauer et al. (2000) and Evans et al. (2001). Theories within this framework propose that reasoners construct and assess *single* models of premises, with believable conclusions promoting confirmatory testing and unbelievable conclusions promoting disconfirmatory testing. The logic by belief interaction in conclusion acceptance rates is explained as an effect arising during the reasoning process itself, rather than a bias located at the response stage. For example, the account explains the large magnitude of belief bias with *indeterminately invalid* conclusions as a consequence of: (1) the confirmatory testing of a believable conclusion uncovering a supporting model (leading to logically incorrect conclusion acceptance); and (2) the disconfirmatory testing of an unbelievable conclusion uncovering a falsifying model (leading to logically correct conclusion rejection).

A further, central assumption of the theory of Klauer et al. that was of particular importance to our inspection-time predictions was their "Assumption 5," which stipulates that reasoners will have considerable difficulty in constructing integrated models for unsound pairs of premises – where the premises logically entail an unbelievable conclusion – and, in particular, when strong conflicts with prior knowledge arise in the process of integrating the premises. In belief-bias studies this assumption typically pertains to those syllogisms where the validity and believability statuses of conclusions are in opposition (i.e., conclusions that are valid yet unbelievable or invalid but believable). An example of a premise pair from the present study where prior definitional knowledge may thwart premise integration is as follows:

Some tulips are inexpensive

No inexpensive things are flowers

As can be seen, integrating these premises into a meaningful model requires the reasoner to imagine that tulips are not flowers, a self-contradictory concept. Such problems should, therefore, be associated with increases in premise processing times relative to those that have sound premises that concur with prior knowledge. These assumptions entail the prediction of a cross-over logic by believability interaction on the premise inspection-time measure. Our pattern of findings support this prediction.

Our analyses also allowed us to determine the temporal locus of this premise-based interaction effect: It arose during a stage of premise inspection *after* conclusions had been looked at rather than at a stage of premise inspection prior to the reasoner viewing the conclusion. This finding suggests that a backward, conclusion-to-premise reasoning mechanism may dominate in syllogistic inference (i.e., people's reasoning is directly guided by the presented conclusion). Evidence for such backward reasoning supports the selective-processing proposals of Evans et al. (2001). It also concurs with the observation of conclusion-driven processing in think-aloud data (Evans et al., 1983) and with recent findings (Morley et al., 2004) showing a marked dissociation between figural-biased responding in a conclusion-generation paradigm (where forward-reasoning seems to be deployed because of the lack of any guiding conclusion) and belief-biased responding in a conclusion-evaluation paradigm.

In summary, our eye-movement data support central assumptions of recent selective processing accounts of belief bias forwarded by Klauer et al (2000) and Evans et al. (2001). The selective processing framework emphasises the critical role of a conclusion's believability in determining the nature of the processing that is subsequently applied to premises, that is, reasoning is directed toward the construction of models that either support or refute presented conclusions, contingent on their believability status. Our data concur with the predictions of Klauer et al of increased processing effort being required for problems where logic and belief disagree, and well as the proposals of Evans et al. for conclusion-driven reasoning in the syllogistic-evaluation paradigm.

Finally, we would argue for the potential value of eye-movement analysis in reasoning research as an adjunct to existing process-tracing techniques such as verbal protocol analysis. This study has demonstrated how an analysis of participants' eye movements can provide insights into the nature and time course of reasoning processes that can help facilitate discrimination between competing accounts of reasoning phenomena. We acknowledge that the application of eye-movement analysis in reasoning research is underdeveloped and much remains to be done to acquire a deeper theoretical understanding of the link between fixation durations, attentional processing, and high-level reasoning mechanisms. We trust, however, that our study has illustrated the kind of data that can be acquired through the employment of eye-movement tracking in studying deduction, and that it has demonstrated how such data may contribute converging evidence to augment our understanding of the dynamics of on-line reasoning processes.

Acknowledgements

We thank Karl Christoph Klauer, Maxwell Roberts, and an anonymous reviewer for valuable comments on previous versions of this paper.

References

Ball, L. J., Lucas, E. J., Miles, J. N. V., & Gale, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology, 56A,* 1053–1077.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology, 108,* 157–193.

Cherubini, P., Garnham, A., Oakhill, J. V., & Morley, E. (1998). Can any ostrich fly? Some new data on belief bias in syllogistic reasoning. *Cognition, 69,* 179–218.

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 (2000). Thinking and believing. In J. Garcìa-Madruga, N. Carriedo, & M. J. González-Labra (Eds.), *Mental models in reasoning*. Madrid: UNED.

Evans, J., St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11,* 295–306.

Evans, J. St. B. T., & Handley, S. J., & Harper, C. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 54A,* 935–958.
Evans, J., St. B. T., Newstead, S. E., & Byrne, R. M. J (1993). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates.
Garnham, A. (1993). A number of questions about a question of number. *Behavioural and Brain Sciences, 16,* 350–351.
Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging, 9,* 356–371.
Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 48A,* 945–975.
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. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16,* 1–62.
Johnson-Laird, P. N., & Byrne, R. M. J (1991). *Deduction*. Hove: Lawrence Erlbaum Associates.
Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35,* 1897–1916.
Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852–884.
Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory and Cognition, 29,* 1000–1009.
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.
Morley, N. J., Evans, J., St. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 57A,* 666–692.
Newstead, S. E., Pollard, P., Evans, J., St. B. T., & Allen, J. (1992). The source of belief bias in syllogistic reasoning. *Cognition, 45,* 257–284.
Oakhill, J., & Johnson-Laird, P. N. (1985). The effect of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology, 37A,* 553–570.
Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31,* 117–140.
Quayle, J. D., & Ball, L. J. (1997). Subjective confidence and the belief bias effect in syllogistic reasoning. In M. G. Shafto and & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 626–631). Mahwah, New Jersey: Lawrence Erlbaum Associates.
Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 53A,* 1202–1223.
Rayner, K. (1999). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.
Roberts, M. J., & Sykes, E. D. A (2003). Belief bias and relational reasoning. *Quarterly Journal of Experimental Psychology, 56A,* 131–154.
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.
Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.ition). New York: Harper & Row.
Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin and Review, 10,* 184–189.
Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology, 57A,* 165–182.

Received May 30, 2003
Revision received March 3, 2005
Accepted March 17, 2005

Linden J. Ball

Department of Psychology
Lancaster University
Lancaster, LA1 4YF
UK
Tel. +44 1524 593 470
Fax: +44 1524 593 744
E-mail: L.Ball@lancaster.ac.uk

