The Influence of Activation Level on Belief Bias in Relational Reasoning

Adrian P. Banks

Department of Psychology

University of Surrey

Reference:

Banks, A.P. (2013). The influence of activation level on belief bias in relational reasoning. *Cognitive Science, 37,* 544-577.

http://onlinelibrary.wiley.com/doi/10.1111/cogs.12017/pdf

Keywords: Relational reasoning; belief bias; ACT-R; dual process theory

ADDRESS FOR CORRESPONDENCE

Adrian P. Banks, Dept of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK. Email: a.banks@surrey.ac.uk

Abstract

A novel explanation of belief bias in relational reasoning is presented based on the role of working memory and retrieval in deductive reasoning, and the influence of prior knowledge on this process. It is proposed that belief bias is caused by the believability of a conclusion in working memory which influences its activation level, determining its likelihood of retrieval and therefore its effect on the reasoning process. This theory explores two main influences of belief on the activation levels of these conclusions. Firstly, believable conclusions have higher activation levels and so are more likely to be recalled during the evaluation of reasoning problems than unbelievable conclusions, and therefore they have a greater influence on the reasoning process. Secondly, prior beliefs about the conclusion have a base level of activation and may be retrieved when logically irrelevant, influencing the evaluation of the problem. The theory of activation and memory is derived from ACT-R (the Atomic Components of Thought – Rational) cognitive architecture and so this account is formalized in an ACT-R cognitive model. Two experiments were conducted to test predictions of this model. Experiment 1 tested strength of belief and Experiment 2 tested the impact of a concurrent working memory load. Both of these manipulations increased the main effect of belief overall and in particular raised belief-based responding in indeterminately invalid problems. These effects support the idea that the activation level of conclusions formed during reasoning influences belief bias. This theory adds to current explanations of belief bias by providing a detailed specification of the role of working memory and how it is influenced by prior knowledge.

1. Introduction

Prior knowledge has frequently been shown to influence deductive reasoning, an effect known as belief bias (Evans, Barston, & Pollard, 1983; Klauer, Musch, & Naumer, 2000). This paper investigates the role of working memory and retrieval in reasoning and demonstrates how prior knowledge can influence this, leading to belief bias. Specifically, it is proposed that the activation level of the representations involved in the reasoning process determines which representations are retrieved during the evaluation of conclusions to relational reasoning problems. In turn, this affects the judgement about the validity of the conclusion. As prior knowledge affects the activation level of memory traces, it can exert an influence on which representations are retrieved and used in reasoning, leading to bias in the judgment of validity. This account provides a description of a key role of working memory in relational reasoning, and a novel mechanism through which belief can influence reasoning. As such it adds to extant explanations of belief bias which do not currently provide a detailed specification of the role of working memory and how it is influenced by prior knowledge.

The theory of activation and working memory is derived from the Atomic Components of Thought – Rational (ACT-R) theory. This is a cognitive architecture that embodies a number of principles governing cognitive processing which have been shown to explain performance across a wide range of tasks (e.g. Anderson, 2007). The concept of activation level has been used to explain the role of working memory in a number of studies (e.g. Anderson & Reder, 1999; Anderson, Reder, & Lebiere, 1996) and is applied directly here to provide the basis for the theory of working memory in relational reasoning. ACT-R theory has also been formalized computationally, and so the account of working memory and belief bias presented here has been realized as a cognitive model which provides well-specified, quantitative accounts of the patterns of response in two novel reasoning experiments.

Studies of belief bias typically ask participants to evaluate whether a conclusion follows logically from the premises given. They are asked to assume the premises are true, that is to ignore their prior beliefs, and to accept only conclusions that necessarily follow from these premises, i.e. logically valid conclusions. An example of a problem is:

> Scott went to the South Pole before satellites first went into space
>
> The Falklands War took place before satellites first went into space
>
> The Titanic sank at the time Scott went to the South Pole
>
> Michael Jackson released "Thriller" during the Falklands War
>
> Therefore, the Titanic sank before Michael Jackson released "Thriller"

The conclusion here is believable, but it is not logically valid. It is consistent with some interpretations of the premises, but not all of them. As a result it is possible but not necessary, and so is *indeterminately invalid.* Problems where the conclusion is consistent with all possible interpretations of the premises are referred to as *determinately valid.* Problems where the conclusion is consistent with no possible interpretations of the premises are referred to as *determinately invalid.* Typically it is found that: (a) valid conclusions are accepted more than invalid conclusions; (b) believable conclusions are accepted more than unbelievable conclusions; and (c) the effects of belief are stronger on indeterminately invalid conclusions than determinately invalid or valid conclusions (Newstead, Pollard, Evans and Allen, 1992). A model of belief bias must account for these three phenomena.

The effects of belief on reasoning have been found to be influenced by a range of factors. The influence of belief is reduced when participants are given strong instructions to reason logically (Evans, Newstead, Allen, & Pollard, 1994) but increases when responses must be made quickly (Evans & Curtis-Holmes, 2005). Emotionally charged content can also reduce belief bias (Goel & Vartanian, 2011). Participants with lower working memory capacity, or a concurrent working memory load (De Neys, 2006; Barton, Fugelsang, & Smilek, 2009) or lower cognitive ability (Torrens, Thompson, & Cramer, 1999) demonstrate more belief bias, as do older participants (Gilinsky & Judd, 1994). Thinking dispositions such as actively open-minded thinking and need for cognition predict belief bias (Macpherson & Stanovich, 2007). Neuroimaging studies indicate that different brain regions are active when reasoning is influenced by belief compared to when the effects of belief are inhibited (Goel & Dolan, 2003). Repetitive transcranial magnetic stimulation of different brain regions selectively influence the degree of belief bias (Tsujii, Sakatani, Masuda, Akiyama, & Watanabe, 2011). Hence belief bias is influenced by a range of cognitive and dispositional factors, and appears to arise from the interaction of different systems in the brain.

Belief bias has been found to influence a number of different deductive reasoning tasks including relational reasoning, (e.g. Roberts & Sykes, 2003), conditional reasoning (Evans, Handley, & Bacon, 2009), and transitive reasoning (Andrews, 2010). Categorical syllogisms are the most common type of problem on which belief bias has been demonstrated (e.g. Evans et al., 1983; Klauer, Musch, & Naumer, 2000). This study will investigate belief bias in relational reasoning problems because they are relatively pure strategically (most theories propose that reasoning involves the construction of spatial representations) and not prone to variations in premise interpretation (Roberts, 2000).

A number of theories have been proposed to account for belief bias in deductive reasoning. Most contemporary accounts are framed in terms of dual process theory (e.g. De Neys, 2006; Evans, 2003; Stanovich & West, 2000). Dual process accounts suggest that belief bias can be explained in terms of an interaction between rapid, automatic, heuristic processes and slow, controlled, analytic processes. Typically, theories suggest separate heuristic and analytic processes which each may generate a response. The heuristic process involves responding according to prior belief whereas the analytic process involves logical inference. In problems where the heuristic and analytic processes give the same solution, known as 'no conflict' problems, the two routes support each other in providing the correct solution. In problems where the heuristic and analytic processes give different solutions, known as 'conflict' problems, the conflict between the two responses must be resolved in some way. Belief bias arises when the belief-based, heuristic process is either favoured over or influences the logical, analytic process. The exact nature of the interaction of these two processes is detailed in several competing theories.

The main explanations of belief bias are: the selective scrutiny model (Evans et al., 1983); the misinterpreted necessity model (Evans et al., 1983); the mental models account (Oakhill, Johnson-Laird, & Garnham, 1989); the selective processing model (Evans, Handley & Harper, 2001; Klauer, Musch & Naumer, 2000) and the metacognitive uncertainty account (Quayle & Ball, 2000).

The selective scrutiny model, proposed by Evans et al. (1983), suggests that the effect of belief occurs prior to reasoning. If the conclusion is believable then it will be accepted directly, if not then some reasoning will take place (the reasoning process is not specified). Hence there is more belief based responding in believable problems than unbelievable ones.

The misinterpreted necessity model (Evans et al., 1983) suggests that the effect of belief occurs due to a misunderstanding about the nature of logical necessity. In experiments, participants are asked to accept only conclusions which necessarily follow from the premises, i.e. the conclusion must be true if the premises are. Indeterminately invalid conclusions have the potential for confusion because they are possible; given the premises are true they might be true or they might not. In this instance participants should reject the conclusion, but they may not have understood that this is what is required. As they find the conclusions ambiguous they may respond with the only cue available – the believability of the conclusion. Hence belief has a larger effect on indeterminately invalid syllogisms.

The mental models account of belief bias (Oakhill & Johnson-Laird, 1985; Oakhill et al., 1989) suggests that people reason using mental models. A mental model is an integration of the information in the premises into a single representation which depicts one possible account of the relationship between the premise terms. The theory proposes that people construct an initial mental model of the premises and test if this is consistent with the conclusion given in the problem. If it is not, the conclusion is rejected. If it is, and the conclusion is believable, then it is accepted. Only if the mental model is consistent with the conclusion but is unbelievable are alternative mental models of the premises constructed to test this conclusion further. A response bias towards rejecting unbelievable conclusions has also been suggested which can explain belief bias in syllogisms for which there is only one possible mental model.

A more recent explanation of belief bias is the Selective Processing model (Evans, Handley & Harper, 2001; Klauer, Musch & Naumer, 2000). This suggests that a single mental model of the premises is constructed. If the conclusion is believable then a search is conducted for a mental model which confirms the conclusion. If this is found then the conclusion is accepted, else it is rejected. However, if the conclusion is unbelievable then a search is conducted for a mental model which disconfirms the conclusion. If it is found then the conclusion is rejected, else it is accepted. For valid syllogisms there are only mental models that confirm the conclusion and so these will be identified irrespective of their believability, leading to a logically valid response. For indeterminately invalid syllogisms there are both confirmatory and disconfirmatory conclusions which could be created and so a confirmatory mental model will be found to support believable conclusions and an disconfirmatory mental model to reject unbelievable conclusions. Evans (2007) adds a default response to accept believable conclusions and reject unbelievable conclusions which may on occasion be overridden by the analytic process described above.

Finally, Quayle and Ball (2000) propose a metacognitive uncertainty account. This explanation is developed from mental models theory. It suggests that when people use multiple mental models in order to solve a problem, greater demands are placed on the limited capacity of their working memory and there is a feeling of uncertainty about the accuracy of reasoning. As a result, people fall back upon a belief heuristic and accept or reject conclusions according to how believable they are. Indeterminately invalid syllogisms require more mental models to be constructed to evaluate the conclusion than valid syllogisms, and so place greater load upon working memory. Therefore there will be more metacognitive feelings of uncertainty with invalid syllogisms and more belief-based responses.

The aim of this paper is to consider the role of working memory in reasoning and specifically to propose a fuller account of how working memory and retrieval might be affected by prior knowledge and therefore influence the reasoning process, leading to belief bias. The review of current theories of belief bias highlights that the role of working memory is often not considered in detail.

In the selective scrutiny model, the impact of belief is unrelated to any memory required during reasoning aside from recognizing the conclusion as believable. There is no role of working memory in this explanation. In the misinterpreted necessity model belief bias is mainly determined by the logical form of the problem, it is not influenced by the memory requirements of the task. Working memory plays no part in this explanation either. The explanation of both the mental models account and the selective processing account is in terms of which mental models are constructed. They do not describe how working memory could influence belief bias directly. The metacognitive uncertainty account does consider the role of working memory, specifically how working memory capacity limitations cause people to switch from analytic to heuristic reasoning. However, this theory is about the limitations of working memory but not the processes involved in working memory. It describes what occurs at the point working memory limits are exceeded, but it does not describe how prior knowledge could influence the retention and retrieval of memories throughout the whole reasoning process. A full account of the role of working memory in reasoning would add to these theories by offering an explanation of working memory processes throughout reasoning.

The remainder of the paper will describe and test the theory of how working memory processes during relational reasoning contribute to belief bias. Firstly, as this study is based in ACT-R theory, the relevant details of ACT-R will be outlined; in particular working memory within ACT-R. Secondly, the model of relational reasoning that has been developed within the ACT-R cognitive architecture will be described. The model is then used to generate predictions about the influence of strength of belief and working memory load on reasoning. Two novel experiments are conducted to test these predictions and assess how well the model fits these data. Finally, the implications of these findings and the model for existing theories of belief bias are discussed.

2. ACT-R, activation levels, and belief bias

2.1 Overview of ACT-R theory and its relevant features

The Atomic Components of Thought – Rational theory (ACT-R) is a cognitive architecture which aims to provide a general theory of cognition (Anderson, 2007). It explains how information is encoded, retrieved and processed. These general architectural principles apply to any cognitive task and have been used to model cognition in a wide variety of domains from simple memory and arithmetic tasks through to more complex behaviour such as car driving and scientific discovery (e.g. Anderson, 2007; Anderson & Lebiere, 1998; Anderson & Matessa, 1997; Anderson, Reder, & Lebiere, 1996; Salvucci, 2006).

According to ACT-R theory, cognition emerges through the interaction of several independent modules in the brain. These modules have specialised functions. The declarative memory module governs the storage and retrieval of declarative information. The problem state (or imaginal) module is responsible for the storage of intermediary mental representations that are used during thinking. The control (or goal) module contains information about the current goal when completing a task. The procedural module controls how each of these modules interact. There are also perceptual and motor modules, such as vision, but these are not used in this model and so they will not be discussed further.

Procedural memory controls the interaction of modules through a series of if-then production rules that comprise procedural memory. The conditions of each rule are matched against the modules, and if one is met then it fires and executes its actions. Typically, these actions involve the further encoding, retrieval or processing of information and so the cognitive process is advanced. The cycle then continues and another production rule (or possibly the same one) will match and fire and so on, until the process completes.

The conditions of each production rule are matched against various module *buffers*. Buffers provide an interface to each of the modules by holding *chunks* that are the output of the current processing of that module. A chunk is a symbolic representation of facts that have been encoded. Only one chunk can be held in a buffer at any one time. Depending on the function of the module, different types of chunk will be placed in these buffers. For example the chunk in the declarative memory buffer is the last item retrieved from declarative memory; the chunk in the problem state buffer is the intermediary representation that is being used in the current task. Hence production rules typically match a pattern of the buffer contents from several modules and determine the action that should be taken given that situation, moving it along a step. Different patterns will trigger different rules and therefore different actions.

In addition to these symbolic representations, ACT-R also defines several equations that determine the speed and accuracy of access to these representations. The value that is of relevance to the model discussed here is the *activation level* of a declarative memory chunk. Activation determines the probability that a chunk in declarative memory will be retrieved. If activation is high, above a threshold which is estimated from the data, the chunk will be successfully retrieved. If not, it will be forgotten.

Activation levels are determined by three factors: the chunk's base level activation, spreading activation, and a noise component. The model presented here explores the role of base level activation on the reasoning process and so this component will be described in detail. Basel level activation is determined by Equation 1 below. This determines the activation of chunk *i*, which has been encountered *n* times; t_j is the time since the j^{th} presentation of the chunk. The parameter *d* determines the decay rate of the base level activation and is typically estimated to be 0.5, as it is here.

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \qquad (1)$$

This means that base level activation is learned over a period of time and increases every time the chunk is retrieved from memory or is encountered again. Base level activation also decays over time. Hence the activation level fluctuates according to how frequently and recently a chunk is used. Essentially, using a chunk raises its base level activation; but without use the activation of a chunk will decay. The more frequently and recently a chunk has been encountered, the greater its activation level and so the more likely it is to be recalled.

Spreading activation is determined by the chunk's association with other chunks currently in the buffers. A chunk in a buffer may serve as a cue to chunks in declarative memory if they share information, raising their activation level. Finally, the noise component adds some random variability to the retrieval process. These two components of activation are not critical to the key predictions of this model. Details of the equations governing these and the remaining subsymbolic processes are in Anderson (2007).

*2.2 ACT-*R *& working memory*

Working memory actively holds information in mind so that it can be processed during reasoning. However, whilst working memory is used in several accounts of reasoning (e.g. Johnson-Laird, 1983; Rips, 1994), there is considerable debate as to the structure and processes involved in working memory (e.g. Baddeley, 2000; Baddeley & Hitch, 1974; Cowan, 2000; McElree, 2001). ACT-R does not have a working memory store that is distinct from long term memory. Instead, the function of working memory is served by two components (Anderson, 2005; Daily, Lovett, & Reder, 2001; Lewis & Vasisth, 2005; Borst, Taatgen, & van Rijn, 2010). Firstly, the buffers to the control module, declarative memory module and the problem state module each have a capacity of one item. They serve as the focus of attention for that module as the item in the buffer can be used for immediate processing. Secondly, the items in declarative memory have varying levels of activation. Those with high activation levels would be readily retrieved if required and so they are accessible memories that could be used during reasoning. Hence working memory is comprised of the buffer contents which are directly accessible as the focus of attention and a further subset of active items in declarative memory that can rapidly become the focus of attention through retrieval from declarative memory.

Items are removed from the focus of attention when they are replaced in the buffer by another item. The activation of the items in declarative memory decays over time and they will eventually fall below the retrieval threshold, but perceiving the stimulus again or retrieving the memory, for example through rehearsal of the item, brings it back into the focus of attention and raises its activation. These are the processes that account for forgetting and maintenance of items within working memory.

A range of evidence supports ACT-R's theory of memory (e.g. Anderson, 2007). The very tight constraint on the capacity of items in the focus of attention is similar to other theories of working memory such as McElree (2001) and the notion of working memory as the activation of items in long term memory is similar to Cowan (2000).

Whilst ACT-R does not have a distinct working memory module, the term is used here to describe the active memories and those in the focus of attention. This is a useful term functionally even if there is no simple architectural mapping because it is this subset of items that general theories of reasoning propose are involved in the reasoning process. It is through the general architectural mechanisms of ACT-R that describe which memories are active that the model presented here explains belief bias.

2.3 Outline of the ACT-R model

The goal of this paper is to provide a more detailed account of the role of working memory in relational reasoning and how it may contribute to belief bias. As discussed above, the majority of theories of belief bias have focused on the selective construction and manipulation of mental models and the biased conclusions drawn from them. This paper is not arguing that the selective processing of reasoning materials does not contribute to belief bias. Rather, it is claimed that these theories are incomplete by not fully examining the role of working memory. This model therefore focuses on the role of working memory rather than the construction of mental models during reasoning. However, the model does require mental models to be constructed and conclusions drawn in order that they may be stored in working memory. The ACT-R model therefore has a construction stage in which mental models are constructed based on the premises and conclusions that are drawn from these mental models, referred to here as *initial conclusions*. This stage of the process is based on previous research on reasoning about relations, in particular Schaeken, Van der Henst, and Schroyens (2007).

The novel contribution of this model lies in the retrieval stage of the process. This differs from earlier accounts of reasoning. In this stage, the initial conclusions are retrieved from working memory and compared with the conclusion presented in the problem in order to evaluate it and provide a response. (For clarity, the conclusion presented in the problem will be referred to here as the *presented conclusion*). If they confirm the presented conclusion then it is accepted, if not it is rejected. This is a key role of working memory, influencing the retention and retrieval of the initial conclusions which are used to evaluate the problem. It is in this retrieval stage that the activation level of the initial conclusions determines the likelihood of their retrieval and it is through this process that belief exerts its influence. The construction and retrieval stages are described in detail below[1].

2.3.1 Construction stage

Most theories of relational reasoning suggest that people reason using mental models: spatial representations in which the information in the premises is integrated into a single representation which depicts one possible layout of the premise terms given the relationships between them (e.g. Schaeken, Johnson-Laird, & d'Ydewalle, 1996). Each mental model

[1] The model code can be downloaded from http://www.surrey.ac.uk/psychology/people/dr_adrian_banks/ or by contacting the author.

represents a layout consistent with the premises. Reasoners initially attempt to construct a single mental model (Goodwin & Johnson-Laird, 2005). The theory presented here suggests that they initially construct a mental model that integrates all of the information into a single representation. For example, here is a relational reasoning problem with letters rather than believable events:

A happens before B
C happens before B
D happens at the same time as B
E happens at the same time as C
Therefore, E happens before D

In this case the relationship between A and C is indeterminate, so a mental model is constructed in which A and C occur before B and D but their relative order is not specified. This is similar to the 'isomeric' mental models described by Schaeken et al. (2007). The model in this case is:

A	B
C	D
E	
Prior to Time$_2$	Time$_2$

Time is represented spatially, running from left to right. An initial conclusion can be drawn from this model by focusing on the two elements mentioned in the conclusion (E and D) and retaining their relative position in a new representation. This leads to the initial conclusion that E happens before D.

Sometimes the conclusion can refer to elements that are represented indeterminately in the mental model. For example:

A happens before B

C happens before B

D happens at the same time as A

E happens at the same time as C

Therefore, E happens before D

This leads to the following model:

A	
D	B
C	
E	
Prior to Time$_2$	Time$_2$

Focusing on the two elements in the conclusion, D and E, does not lead to an unambiguous new representation because their relative order is indeterminate. In this case further mental models are constructed of each possibility:

C	A	B		A	C	B
E	D			D	E	
Time_1	Time_2	Time_3		Time_1	Time_2	Time_3

The initial conclusion drawn from the first mental model is that E happens before D, but the initial conclusion drawn from the second mental model is that D happens before E. This is an indeterminately invalid problem because the presented conclusion is consistent with one but not both of the layouts. That is, the conclusion is possible but does not necessarily follow from the premises.

Each mental model is represented in a single chunk in ACT-R. Each chunk has slots that hold (a) the elements in the mental model and (b) the relative time at which these occur. The first premise is encoded into the chunk by filling the first two slots. Hence the premise 'A occurs before B' would form the chunk:

Model0-0

TermA	A
TimeA	1
TermB	B
TimeB	2
Relation	Before

The remaining elements in the premises are then attached to further slots within the chunk. Each new premise in these problems contains an element that already exists in the mental model and one new element. The new element is placed in the next free 'Term' slot. The relative time of the event is represented in the 'Time' slot. If the premise states that the new element occurs before the existing element, then the number placed in the time slot is one less than the existing element. If the premise states that the new element occurs at the same time as the existing element, then the number placed in the time slot is the same as the existing element. If the premise states that the new element occurs after the existing element, then the number placed in the time slot is one greater than the existing element. Hence the time slot keeps the relative time of the occurrence of event.

The initial 'isomeric' model constructed from the indeterminate problem directly above is:

Model0-0

TermA	A
TimeA	1
TermB	B
TimeB	2
TermC	C
TimeC	1
TermD	D
TimeD	1
TermE	E
TimeE	1
Relation	Before

The relationship between D and E is indeterminate, and so mental models are constructed of each possibility. As before, elements from each premise are placed into slots within the chunk. The process is slightly altered in order to make explicit each possible arrangement of the elements. This time, if the premise states that the new element occurs before the existing element, then the number placed in the time slot is the same as the existing element and the number in the time slot of the existing element is increased by one. As before, if the premise states that the new element occurs at the same time as the existing element, then the number placed in the time slot is the same as the existing element. If the premise states that the new element occurs after the existing element, then the number placed in the time slot is one greater than the existing element.

This leads to two models:

Model1-0

TermA	A
TimeA	2
TermB	B
TimeB	3
TermC	C
TimeC	1
TermD	D
TimeD	2
TermE	E
TimeE	1
Relation	Before

Model2-0

TermA	A
TimeA	1
TermB	B
TimeB	3
TermC	C
TimeC	2
TermD	D
TimeD	1
TermE	E
TimeE	2
Relation	Before

The final stage in the process is to draw a conclusion from the mental models. This is completed by creating a new chunk representing the two elements in the conclusion and the relationship between them. In the problems used in this experiment, the conclusion always linked the elements in the slots of TermD and TermE. These two elements and their relationship are therefore placed in the slots of a conclusion chunk. If TimeD was less than TimeE then TermD was placed before Term E and vice versa. The two conclusions created from the two mental model chunks directly above are:

Conclusion0-0		Conclusion1-0	
Term1	E	Term1	D
Term2	D	Term2	E
Relation	Before	Relation	Before

Overall, the construction stage involves the creation of mental models from which initial conclusions are drawn. These are then held in working memory until the retrieval stage when they are used to evaluate the presented conclusion.

2.3.2 Retrieval stage

The problem is evaluated by recalling the initial conclusions drawn in the construction stage from working memory and comparing them to the presented conclusion. One initial conclusion is retrieved with determinate problems. If this conclusion confirms the presented conclusion then it is accepted, if it disconfirms it then it is rejected. More than one initial conclusion is retrieved with indeterminate problems (because more than one was drawn in the construction stage). If all of the conclusions that are retrieved confirm the presented conclusion then it is accepted, but if one is retrieved that disconfirms it then it is rejected. During the retrieval of multiple confirmatory conclusions, the fact that the conclusion is possible is held in the control buffer, only rejecting the conclusion when a disconfirmatory conclusion is retrieved. If, when no more conclusions are retrieved, the evaluation is still possible then the conclusion is accepted. If no conclusions are successfully retrieved then the response is a guess, with an equal chance of accepting or rejecting the presented conclusion.

If all of the conclusions drawn in the construction stage, and no logically irrelevant beliefs, are retrieved then the logically valid response will always be given. However, as the activation levels of the conclusions in memory decay over time it is likely that some will be forgotten. Which conclusions are retrieved and which are forgotten is not random, it is influenced by a number of factors including prior knowledge about the conclusions. The role of working memory in belief bias lies in explaining how prior knowledge influences which conclusions are retrieved.

The likelihood that one of the initial conclusions will be retrieved successfully from working memory is determined by its activation level. In ACT-R theory, an item in declarative memory has an activation level which decays over time but is raised whenever that item is experienced again. Essentially, the new encounter with that item merges with the existing memory trace for that item and raises its activation level. If an initial conclusion is believable, that is, the initial conclusion is the same as the prior belief in declarative memory, it will merge with the prior belief to create a single memory of higher activation. However, if an initial conclusion is unbelievable, that is, the initial conclusion is different to the prior belief in declarative memory, it will not merge with it. As a result its activation level will not be as high. In other words, believable initial conclusions will have higher activation levels than unbelievable initial conclusions because they merge with prior beliefs and are therefore more likely to be retrieved.

This simple explanation, that believable initial conclusions are more likely to be successfully retrieved from working memory during the reasoning process than unbelievable initial conclusions, can account for the main belief bias phenomena. The main effect of belief on reasoning is explained through the greater likelihood that believable initial conclusions will influence the evaluation of the problem as a result of the greater likelihood that they will be retrieved and used in the evaluation. The interaction of belief and logical validity is explained through increased belief-based responding to indeterminate problems. This interaction is more complex. Two initial conclusions are constructed with indeterminate problems, as described above, one of which will be believable and the other unbelievable. The believable initial conclusion confirms the presented conclusion in the believable condition, so this conclusion is more likely to be accepted because the believable initial conclusion is the one most likely to be retrieved. However, in the unbelievable condition, the believable initial conclusion disconfirms the presented conclusion, so this conclusion is more likely to be rejected because the believable initial conclusion is the one most likely to be retrieved. In determinate problems only one initial conclusion is drawn, and so this effect does not arise. As a result, the influence of belief is greater in indeterminate problems than determinate problems leading to an interaction of logical validity and belief.

Additionally, there is a second influence of prior beliefs. They can be unintentionally retrieved even when logically irrelevant. As a result they may be used in the evaluation process instead of or as well as the initial conclusions leading to an increase in the believable response irrespective of whether it is logically correct or not. This contributes to the main effect of belief.

In summary, this explanation shows how the merging of initial conclusions with prior belief means that believable initial conclusions are more active and more likely to be remembered. Also, the occasional unintentional retrieval of logically irrelevant prior beliefs can influence responses. Together, these explain the major belief bias effects. Having outlined the account of belief bias it is now possible to test formally how well it fits the data using an ACT-R model of the process.

3. Experiment 1: The effect of belief strength on belief bias

The ACT-R model described above suggests that the activation level of the initial conclusions formed during reasoning will influence belief bias. The first experiment will test the effect of high activation levels. If initial conclusions have a high activation, they are more likely to be retrieved. According to ACT-R theory one situation in which this will occur is if the prior beliefs have high activation levels. When they merge with the believable initial conclusions the resulting believable initial conclusion will have a higher activation level too. This will lead to increased likelihood of retrieving the believable initial conclusions and greater belief bias. The activation level of a prior belief relates to how strong that belief is. Strong beliefs have a high level of activation, so it is predicted that belief bias will be greater with conclusions which participants have stronger beliefs about.

To establish how strongly beliefs are held, participants were asked to rate the believability of the conclusions used in the problems on a seven point scale ranging from 'completely unbelievable' through to 'completely believable'. Extreme scores on this scale indicate a strongly held belief (either they certainly belief it or they certainly don't) whereas more moderate scores indicate a weak belief. An extreme believable rating indicates that the conclusion is strongly believed and the activation of that prior belief is high. An extreme unbelievable rating indicates that the opposite of that conclusion is strongly believed and has high activation.

In this experiment, it is predicted that in problems that have conclusions with stronger prior beliefs, the believable initial conclusions will have higher activation levels than when prior beliefs are weaker, leading to a larger main effect of belief with strong belief problems. In addition, there will be an increase in unintentional retrievals of logically unrelated prior beliefs because of their higher activation level, also increasing the main effect of belief. The increase in belief-based responding in the indeterminate condition will be accentuated because of an increased likelihood of retrieving the believable initial conclusion. This confirms the presented conclusion in the believable condition and so it will be accepted more in this condition, but it disconfirms the presented conclusion in the unbelievable condition and so it will be rejected more in this condition. As only the indeterminate problems generate both confirmatory and disconfirmatory initial conclusions, the effect of strong beliefs will be greater in this condition than in the valid condition, leading to an interaction of belief and logic.

A search of the current literature found that this prediction has not previously been reported in belief bias research and so, if supported, would be a novel finding. If this theory is correct, then it should be possible to explain the pattern of results across the strong and weak belief conditions simply by increasing the initial activation of the prior belief in the ACT-R model to simulate stronger believability of the problems.

3.1 Method

3.1.1 Participants

Thirty-four undergraduate psychology students participated in return for course credit.

3.1.2 Design

A within subjects, three factor design was used. Problems varied in the believability of their conclusion (*believable* or *unbelievable*), the strength of this belief (*strong* or *weak*) and their validity (*valid, determinately invalid* or *indeterminately invalid*). A full factorial design resulted in twelve problem types. The dependent variable was the number of conclusions accepted in each condition. Problems were presented in a different random order to each participant.

3.1.3 Materials

Participants completed twenty-four problems, two of each type. The problems used were all of the form described above. The content of the problems was based on those used by Roberts and Sykes (2003) and all described temporal relations between a number of historical events. Two premises were believable and two were unbelievable in each problem in order to control for premise believability. For example:

> The University of Cambridge was built before the English Civil War
> The Cold War occurred before the English Civil War
> Graham Bell invented the telephone during the English Civil War
> The Berlin wall was dismantled during the Cold War
> Therefore, Bell invented the telephone before the Berlin wall was dismantled.

A second set of problems was created by reversing the relational term for each conclusion. Thus the believable valid problem in the first set becomes an unbelievable determinately invalid problem in the second set, and so on. This technique counterbalances the strength of belief in the believable and unbelievable conditions and avoids discrepancies arising if the conclusions chosen for one condition are stronger in belief overall than the other. Half of the participants used the first set of problems and half used the second set.

Finally, participants completed a questionnaire in which they rated the strength of their belief in the conclusions of the problems. This rating was used to allocate the problems to the strong and weak belief conditions. Conclusions with more extreme scores were allocated to the strong condition whereas conclusions with more moderate scores were allocated to the weak condition. Exactly half of the problems were categorised as strong and half were categorised as weak.

The effectiveness of this manipulation was subsequently confirmed by asking 28 participants to independently rate each of the conclusions for believability on a seven point scale ranging from 'completely unbelievable' through to 'completely believable'. These ratings are presented in Table 1. The problems in the strong belief condition had significantly more extreme ratings than those in the weak belief condition for all of the problem types, with the exception of the Believable Indeterminately Invalid condition. These problems fell short of significance but nonetheless the trend in the ratings was in the expected direction. Therefore the strong beliefs are indeed stronger than the weak beliefs.

--------------------Insert Table 1 about here--------------------

3.1.4 Procedure

Participants completed the task individually. They were given the following written instructions: 'This is an experiment to test peoples' reasoning ability. You will be given 24 problems in total. For each problem you will be shown four 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 statements are, in fact, true. If, and only if, you judge the conclusion necessarily follows from the statements, you should press 'd' on the keyboard, otherwise press 'k'. Please answer all the questions as accurately and quickly as you can. Please press the spacebar when you are ready to move on to the next problem.' Participants were given a practice problem and their solution was discussed to ensure that they had understood the task. Then they completed the experiment. The practice and experimental materials were presented using a computer. Finally, the questionnaire assessing strength of belief of the conclusions was completed.

3.2 Results and discussion

Responses to the forward and reversed problems were combined and the percentage of conclusions accepted calculated for each condition. These responses are presented in Fig. 1. A 2x3x2 within subjects ANOVA showed a main effect of validity F(2,66) = 127.18, $p<0.001$, $\eta^2=0.79$, a main effect of belief F(1,33) = 6.30, $p=0.02$, $\eta^2=0.16$, and an interaction of belief and validity F(2,66) = 4.43, $p=0.02$, $\eta^2=0.12$. Thus the main belief bias effects were found: valid conclusions were accepted more than invalid conclusions, believable conclusions were accepted more than unbelievable conclusions, and the effect of belief on accepting conclusions differed according to the validity of the problem. This interaction is examined in more detail below. There was no significant main effect of strength of belief F(1,33) = 0.04, p=0.85, η^2=0.001, as expected, because the predicted increase in acceptance of believable conclusions is balanced by the decrease in acceptance of unbelievable conclusions. There was a three way interaction between strength of belief, validity and belief F(2,66) = 9.73, p<0.001, η^2=0.23, suggesting that the interaction of belief and logic was greater in problems with stronger beliefs.

It was predicted that the effect of belief bias would be greater in problems with conclusions that elicited stronger beliefs. Therefore the percentage of conclusions accepted for valid and indeterminately invalid problems were examined separately for strong and weak beliefs. In the strong belief condition there was a main effect of validity F(1,33) = 123.44, p<0.001, η^2=0.79, a main effect of belief F(1,33) = 13.36, p=0.001, η^2=0.29, and an interaction of validity and belief F(1,33) = 14.23, p=0.001, η^2=0.30. Thus the predicted belief bias effects were found. Believable conclusions were accepted more often, especially in the indeterminately invalid condition. In the weak belief condition there was a main effect of validity F(1,33) = 70.78, p<0.001, η^2=0.68, but no main effect of belief F(1,33) = 0.36, p=0.55, η^2=0.10, nor an interaction of validity and belief F(1,33) = 0.80, p=0.38, η^2=0.02. Although Fig. 1 suggests a tendency to accept believable conclusions more than unbelievable conclusions in the weak condition, this effect was not strong enough to lead to a significant difference. Therefore the predictions of this experiment were supported. Belief bias is greater in problems with conclusions eliciting a stronger belief.

--------------------Insert Figure 1 about here--------------------

3.2.1 ACT-R model

Four parameters were estimated from the data and the same values were used in all of the conditions across both experiments. These were: the noise in the equation governing activation level was estimated at 0.6; the retrieval threshold which was estimated 0.5; and the noise in the equation governing utility of productions was estimated at 0.1. (The utility of a production rule is a value governing how likely it is to fire in the event that more than one rule matches a given condition). Finally, the initial activation level of the prior belief chunk was set to a lower value in the weak belief condition than the strong belief condition. Otherwise the ACT-R defaults were used for all parameters.

These parameters influence the model in specific ways. Noise in the activation level of declarative memories influences the likelihood that an initial conclusion is retrieved randomly rather than because it relates to the conclusion being evaluated. A high level of noise therefore leads to random responding and a low level of noise leads to less random, more logical responding. Retrieval threshold influences whether an initial conclusion is successfully retrieved or whether it is forgotten. If this parameter is too high, then initial conclusions will not be retrieved, leading to guessing. If this is too low then all initial conclusions will be retrieved leading to increased logical responding. The data falls between these two extremes – it is neither entirely random nor entirely logical. Hence there appears to be a moderate amount of noise in the process, but not enough to mask systematic responses. Future research could model individual differences in logical responding by manipulating these parameters, but these data are only analysed at a group level and so a moderate position is required to best fit the mean response. The parameter determining noise in the utility of productions plays a small but nonetheless necessary part in this model. This is influential in only the situation in which no initial conclusions are retrieved and the model's response is to guess. In order for this guess to be random there must be some random noise in the selection process of a valid or invalid response. Therefore setting this parameter to any value apart from zero provides the necessary random component.

One goal of this experiment was to provide a test of the predictions based on activation levels during reasoning. The predictions of the ACT-R model are presented alongside the experimental data in Fig. 1. The fit between the model and data was good, with an R^2 of 0.95. The model demonstrates a tendency to accept believable conclusions more than unbelievable conclusions, and this effect is more marked in the strong belief condition. Valid conclusions are more likely to be accepted than invalid conclusions. The effect of belief on indeterminately invalid conclusions is evident in the strong belief condition, but much reduced in the weak belief condition. Thus the model explains the main belief bias phenomena, and also reflects the experimental predictions of the influence of belief strength on belief bias.

The fit of the model is not perfect, however. In addition to the large preference for believable conclusions in the indeterminately invalid condition, the model also predicts in Fig. 1 that there will be a small preference for believable conclusions over unbelievable conclusions in the valid and determinately invalid conditions. This was not found in the data. However this trend is a common finding in other belief bias experiments, e.g. Roberts and Sykes (2003), and its absence may be a result of noise in these data rather than a poor prediction by the model. The preference for believable conclusions in determinate problems is typically small and often non-significant.

In sum, the only difference between the model of strong belief and weak beliefs is the strength of initial activation of the prior belief chunk. All other parameters are the same.

This suggests that changing the activation level of the prior belief is a sufficient and parsimonious explanation of belief bias effects when it influences reasoning through the process outlined in the current model.

4. Experiment 2: The effect of concurrent working memory load on belief bias

Experiment 1 investigated the impact of initial conclusions that participants have strong beliefs about, modelled in ACT-R using high activation levels. Experiment 2 will test the opposite effect, the impact of reducing the probability of recall by lowering activation levels of the initial conclusions. This will be achieved by introducing a secondary task which, according to ACT-R theory, will lower the activation levels of the initial conclusions and therefore influence belief bias.

Within ACT-R, activation is raised whenever a chunk is used but will then decay over time. When the activation level of a chunk falls below a threshold it will not be retrieved from working memory. Therefore, the longer the reasoning process continues, the more likely it is that initial conclusions will not be successfully retrieved. This experiment used a concurrent working memory load to delay the reasoning process in order to investigate the effect that this has on belief bias. A random five digit number was used for this. It was presented to participants before they attempted to solve each problem. They maintained this number in working memory whilst they read and responded to the reasoning problem, and then recalled the number.

Adding a concurrent working memory load to the experiment requires a participant to switch between two tasks: rehearsing the working memory load and completing the reasoning task. Salvucci and Taatgen (2008) present a theory of concurrent multitasking that proposes some constraints when completing two tasks simultaneously. They suggest that whilst a person has multiple cognitive resources, each resource executes its processes serially. Rehearsal of the working memory load draws upon the same resource as reasoning, specifically the procedural memory module. Therefore when the working memory load is being rehearsed, reasoning must wait. As a result, the initial conclusions decay more when there is a working memory load.

The implication of reducing the activation levels of the initial conclusions is that they are less likely to be retrieved. For every problem, if all of the initial conclusions and no logically irrelevant beliefs are retrieved then the response given will be logically valid. But if the activation levels of all initial conclusions are reduced then many will not be retrieved. If no initial conclusion is retrieved then the response will be a guess. Therefore the concurrent working memory load is predicted to increase guessing and reduce both logic and belief-based responding in all conditions. Additionally, there is predicted to be an increase in belief-based responses in the indeterminate condition. In these problems two initial conclusions are drawn and should be retrieved to ensure logical responding. The first initial conclusion to be retrieved will typically be the most active one, the believable initial conclusion. The effect of concurrent working memory load will be to delay this process so that initial conclusions will have decayed further. Therefore it is less likely that the second initial conclusion will be successfully retrieved with the working memory load than without. As explained above, the believable conclusion confirms the presented conclusion in the believable condition but disconfirms it in the unbelievable condition, leading to increased acceptance of believable presented conclusions and rejection of unbelievable ones. By increasing the likelihood that only the believable initial conclusion is retrieved, the working memory load will accentuate this effect. In sum, the effect of working memory load will be to increase guessing across all conditions and, in addition, to increase belief-based responding in the indeterminate condition.

The effect of a concurrent working memory load on belief bias in syllogistic reasoning was first studied by De Neys (2006). De Neys found that working memory load reduced logical responding in problems where belief and logic conflicted but not in problems where they were consistent. This result was explained in terms of a general dual process theory (e.g. Evans, 2003; Stanovich & West, 2000) rather than using one of the more specific models of belief bias discussed above. As such, the more detailed interactions of different problem types are not specified, the theory only refers to the more general situation when heuristic and analytic reasoning processes either converge or conflict.

De Neys's explanation is that heuristic processes are automatic and will be unaffected by working memory load whereas the analytic processes require working memory resources and will be hampered by the load. In no-conflict problems the belief-based response is logically correct, so drawing upon the belief heuristic will generate the correct response. As the heuristic system does not use working memory resources this performance will be unaffected by the working memory load. In conflict problems the belief-based response is logically incorrect, so using the heuristic will generate the incorrect response. Without a working memory load the analytic system can override the heuristic system and generate the correct response, but with a load the analytic system is hindered and performance will be impaired. Hence working memory load affects conflict problems more than no-conflict problems.

This is an appropriate level of analysis for the conclusions drawn by De Neys, but it is not sufficiently detailed to allow the data to discriminate between different models of belief bias. In fact, when the effect of belief bias on different problem types is considered, De Neys's explanation leads to different predictions compared to other models of belief bias, including the activation level account developed here. Therefore the current experiment can discriminate between different theories of belief bias.

If De Neys's account is correct, then loading working memory will cause a shift towards belief based responses on conflict problems in all conditions: valid, determinately invalid and indeterminately invalid because the shift to heuristic reasoning processes will occur equally in all of them. In contrast, the activation level account described above predicts that loading working memory will cause a shift towards belief based responses only with indeterminately invalid problems. Valid and determinately invalid problems will not be affected as much. To test these hypotheses the conflict vs. no-conflict conditions will be compared for each problem type (valid, determinately invalid, and indeterminately invalid) to test if concurrent working memory load affects belief bias in all of problems or just indeterminately invalid problems.

4.1 Method

4.1.1 Participants

Thirty-two undergraduate psychology students participated in return for course credit.

4.1.2 Design

A within subjects, three factor design was used. Problems varied in the believability of their conclusion (*believable* or *unbelievable*), their validity (*valid, determinately invalid* or *indeterminately invalid*) and whether there was a concurrent working memory load or not (*load* or *no load*).

The dependent variable was the number of conclusions accepted in each condition. Problems were presented in a different random order to each participant. The order of the working memory load condition was counterbalanced. Half of the participants completed the working memory load condition before the no load condition and the other half completed the no load condition first.

4.1.3 Materials

The problems used were the same as those used in Experiment 1. Two equivalent sets were created and were used in the load and no load conditions, alternating between subjects to counterbalance any effects of problem content. Forward and reversed problem conclusions were used to counterbalance the different beliefs used in the problems across the conditions, as in Experiment 1.

4.1.4 Procedure

Participants completed the task individually. They were given the same practice trial and written instructions as in Experiment 1, with an extra practice before the working memory load condition and these additional instructions: ‘Additionally, during twelve of these problems you will be asked to remember 5 random numbers between 1 and 9 whilst you are solving the problem. When you have solved it, recall these numbers out loud for the experimenter to note down. There will be a different set of numbers for each problem’.

4.2 Results and discussion

Responses to the forward and reversed problems were combined and the percentage of conclusions accepted calculated for each condition. These responses are presented in Fig. 2. A 2x3x2 within subjects ANOVA showed a main effect of validity $F(2,62) = 63.22$, $p<0.001$, $\eta^2=0.67$, no main effect of belief $F(1,31) = 0.98$, $p>0.05$, $\eta^2=0.03$, and an interaction of validity and belief $F(2,62) = 5.64$, $p=0.01$, $\eta^2=0.15$. Thus the expected effects of validity and the interaction with belief were found, but unexpectedly there was no main effect of belief. Examining Fig. 2 it seems that the reason for this was that unbelievable valid conclusions were accepted slightly (but not significantly – see below) more frequently than believable valid conclusions. This cannot be attributed to content effects of the problems because of the counterbalancing of the two sets of problems with forward and reversed conclusions.

As expected, there was no main effect of working memory load because the predicted decrease in accepting valid conclusions was balanced by the predicted increase in accepting invalid conclusions. However this hypothesis is testable if the data are recoded using the percentage of logically correct responses instead of the percentage of conclusions accepted. A main effect of working memory load was found $F(1,31) = 4.27$, $p=0.047$, $\eta^2=0.12$. The logical response is given less often when reasoning with a concurrent working memory load. This suggests that the working memory load manipulation was effective.

It was predicted that the effect of belief bias would be greater with the working memory load than without. The three-way interaction of working memory load, validity and belief was non-significant $F(2,62) = 0.47$, $p=0.63$, $\eta^2=0.15$. However, planned comparisons were conducted to compare the effects of working memory load on the percentage of conclusions accepted for valid and indeterminately invalid problems. In the load condition there was a main effect of validity F(1,31) = 19.36, $p<0.001$, $\eta^2=0.38$, a marginally significant effect of belief F(1,31) = 3.53, $p=0.07$, $\eta^2=0.10$, and a significant interaction of validity and belief F(1,31) = 8.91, $p=0.01$, $\eta^2=0.22$. It was also predicted that this interaction of valid and invalid problems would be present only with indeterminately invalid problems, not the determinately invalid problems. Comparing the valid and determinately invalid problems in the load condition, a main effect of validity was found F(1,31) = 48.02, $p<0.001$, $\eta^2=0.61$, but there was no main effect of belief F(1,31) = 0.09, $p=0.77$, $\eta^2=0.003$, and no interaction of validity and belief F(1,31) = 0.90, $p=0.35$, $\eta^2=0.03$. Thus the predicted effects of working memory load were found. Believable conclusions were accepted more often, specifically in the indeterminately invalid condition.

In the no load condition when comparing valid and indeterminately invalid problems there was a main effect of validity F(1,31) = 40.63, $p<0.001$, $\eta^2=0.57$, but no main effect of belief F(1,31) = 0.03, $p=0.87$, $\eta^2=0.001$, nor an interaction of validity and belief F(1,31) = 2.05, $p=0.16$, $\eta^2=0.06$. Comparing valid and determinately invalid problems in the no load condition found a main effect of validity F(1,31) = 181.19, $p<0.001$, $\eta^2=0.85$, no main effect of belief F(1,31) = 1.34, $p=0.26$, $\eta^2=0.04$, and no interaction of validity and belief F(1,31) = 0.24, $p=0.63$, $\eta^2=0.01$. Therefore the predictions of this experiment were supported. When valid problems are compared with indeterminately invalid problems, there is an effect of belief bias when there is a working memory load but no effect without.

--------------------Insert Figure 2 about here--------------------

The predictions of the activation level account were contrasted with the explanation suggested by De Neys. To test these predictions the responses for conflict and no-conflict conditions were compared for each problem type. Six post hoc t tests were required to do this, and so a Bonferroni correction was used giving a significance level of p<0.008. In the no working memory load condition there was no significant difference between believable valid and unbelievable valid problems (t(31) = -1.09, p=0.28); there was no significant difference between believable determinately invalid and unbelievable determinately invalid problems (t(31) = -0.49, p=0.63); and there was no significant difference between believable indeterminately invalid and unbelievable indeterminately invalid problems (t(31) = 1.14, p=0.26). In the working memory load condition there was no significant difference between believable valid and unbelievable valid problems (t(31) = -1.07, p=0.29); there was no significant difference between believable determinately invalid and unbelievable determinately invalid problems (t(31) = 0.37, p=0.71); but there was a significant difference between believable indeterminately invalid and unbelievable indeterminately invalid problems (t(31) = 2.98, p=0.006). These results do not support the predictions derived from De Neys's paper that working memory load will reduce logical responding in all conflict problems. Instead the effect seems to be present only in the indeterminately invalid problems. Determinate conflict problems are unaffected by a working memory load. These data suggest that the influence of working memory load on this task is more complex than initially thought and cannot be fully explained by a proposing shift from analytic to heuristic processing in all problems. However, this finding fits the activation level account of belief bias which predicted that the effect would only occur in indeterminate problems.

4.2.1 ACT-R model

The fit of the ACT-R model to the data was also tested. The concurrent working memory load was modeled by placing a chunk containing the number in declarative memory. Productions were added to retrieve this chunk and vocalize the number. These productions compete with the productions for reasoning. When the working memory load production wins, it is retrieved and vocalised which blocks reasoning during this time. As a result reasoning is delayed and the activation levels of the initial conclusions in declarative memory decay.

The predictions of the ACT-R model are presented alongside the experimental data in Fig. 2. The fit between the model and the data was good, with an R^2 of 0.95. The no load condition is equivalent to Experiment 1, and so as before there is a tendency to accept more valid conclusions and more believable conclusions. The tendency to accept believable conclusions is greater in the indeterminately invalid condition. These effects are accentuated in the working memory load condition; in particular the effect of belief on indeterminately invalid conclusions is greater than in the no load condition.

5. General discussion

Gilhooly suggests that 'In life, memory and thinking are inextricably intertwined' (Gilhooly, 1998, p. 7). The purpose of this paper has been to explore this link to explain belief bias in relational reasoning. A novel mechanism is proposed in which belief bias can occur through the influence that belief has on the retrieval of conclusions from working memory. This account suggests that when evaluating relational reasoning problems, mental models are constructed of the premises from which initial conclusions are drawn. These are then retrieved from working memory in order to evaluate the conclusion presented in the problem. The likelihood of a conclusion being retrieved is determined by its activation level, which is raised if the conclusion matches a previously held belief. Therefore believable initial conclusions have a greater influence on the reasoning process.

This explains the main effect of belief in belief bias studies because believable conclusions are more likely to be used when evaluating the problem, and also because prior beliefs will sometimes be retrieved even when unrelated to the logical task. It also explains the interaction of logic and belief in indeterminately invalid problems because in these problems two alternative models are created, but the believable one is more likely to be remembered. The believable conclusion will confirm a believable problem but disconfirm an unbelievable problem, leading to an interaction between logic and belief. However, this interaction does not arise with determinately invalid problems because only disconfirming conclusions can be drawn, irrespective of belief.

The theory of activation and its influence on recall is drawn from ACT-R theory. This theory has been used as an effective explanation underpinning cognition in a wide range of tasks, and so it is a reasonable extension of the theory to test if it underpins reasoning too. This approach also means that the theory proposed here was realized as an ACT-R computational model which was tested against belief bias data. The fit of the model predictions to the experimental data was good throughout, providing support for the sufficiency of this explanation. The activation level account that has been proposed here adds to extant theories of belief bias by specifying the role that working memory plays in more detail and identifying the contribution that it can make to belief bias effects.

Two novel hypotheses were suggested by the idea that the activation level of the initial conclusions formed during reasoning will influence belief bias. Firstly, according to this account, the strength of belief should influence belief bias as it would affect the activation of the believable conclusions. Experiment 1 compared problems with strong and weak beliefs and as expected strong belief problems showed a greater main effect of belief and a greater interaction of logic and belief when comparing valid and indeterminately invalid problems. More importantly for this account of belief bias, these effects were replicated in the ACT-R model by only adjusting the strength of the prior belief. As predicted, fitting this single parameter reproduced all of the changes in belief bias suggesting a parsimonious explanation of these effects. Secondly, as activation decays over time, it was predicted that a concurrent secondary task that loads working memory would increase the effects of belief bias as the task would slow the reasoning process and only the most active memories, i.e. the believable ones, would be retrieved. Experiment 2 compared problems completed with and without a concurrent working memory load and as expected the working memory load condition increased the effects of belief on reasoning. These data were also replicated using a similar ACT-R model. Overall, good support was found for the activation level account of belief bias through the fit of the model to reasoning data and the findings of novel experiments suggested by this theory.

A detailed explanation of how working memory can influence belief bias has not been discussed in previous theories. As a result, the main theoretical contribution of this paper lies in the account of how prior knowledge influences the retrieval of initial conclusions from working memory, referred to here as the retrieval stage, and how this determines the evaluation of reasoning problems. This extends previous belief bias research and also provides a good account of the experimental data collected in this paper.

5.1 Implications of the experimental findings for the major accounts of belief bias

The fit of the ACT-R model to the experimental data was good, implying that the mechanism for explaining belief bias through the activation level of conclusions offers an effective explanation of the data. That this account is sufficient on its own to describe the data is interesting, suggesting that activation levels in working memory do have an important role to play in reasoning and belief bias. However, this is not to claim that existing theories have been refuted by these experiments as the aim of this paper is to extend current theories of belief bias through considering the role of working memory rather than replace them. Nonetheless, the current theories do vary in their ability to explain the pattern of results found here.

The selective scrutiny theory predicts that belief bias occurs because believable conclusions are accepted without reasoning but unbelievable problems do involve reasoning. This distinction is predicted to take place irrespective of strength of belief and so this model cannot account for the findings of Experiment 1. Whilst accepting believable conclusions is unlikely to be greatly affected by a working memory load, reasoning with unbelievable conclusions might be. However the reasoning mechanism in this model is not specified, and so it is not possible to predict how it would be affected by a working memory load and what the implications would be for patterns of response. Therefore this model cannot account very satisfactorily for the findings of Experiment 2. The misinterpreted necessity theory predicts that belief bias occurs through a misunderstanding of logical necessity. This is their general understanding about the meaning of logical necessity rather than task specific knowledge. Therefore it is not likely to be altered by strength of belief or working memory load in the experiment. As a result, these theories do not account for the findings in this study very well.

The mental models account can explain the larger belief bias effects found with stronger beliefs in Experiment 1. If belief in the first model is strong it is plausible that it is more likely that participants will accept the conclusion without further analysis. If belief is weak then it is plausible that participants are less likely to accept the conclusion immediately and continue to flesh out all of the mental models, reasoning as they do with belief-neutral materials to generate the logical response. The results of Experiment 2 do not fit the mental models account as well, in particular the increased belief-based responding found with indeterminately invalid problems when completed with a concurrent working memory load. The mental models explanation of this interaction lies in fleshing out the mental models of problems with believable conclusions. A working memory load would make this difficult to do because limited working memory capacity would be exceeded. Therefore the mental models account predicts that this interaction would decrease, not increase as was found. Overall, the mental models account can explain the findings of Experiment 1 but is harder to reconcile with the findings of Experiment 2.

The selective processing account has been proposed by both Klauer et al. (2000) and Evans et al. (2001). This model explains the main effect of belief well. It suggests a default heuristic response is made but this can be overridden by an analytic response. Increasing the strength of belief in a problem could reduce the likelihood that the analytic process intervenes, explaining the main effect of belief in Experiment 1. Introducing a working memory load could prevent analytic processing and again reduce the likelihood that the analytic process intervenes, explaining the main effect of belief in Experiment 2. The explanation of the interaction of belief and logic is less clear, however. This interaction is explained through analytic, not heuristic processing. Therefore if the effect of the manipulations in Experiments 1 and 2 is to increase heuristic responding and reduce analytic intervention then this model predicts that the interaction would decrease, not increase as was found in both experiments.

There is an alternative explanation of these findings though. It may be that belief strength was sufficiently low in Experiment 1 that the items are treated as relatively belief-neutral materials and so no interaction would be expected with weak beliefs, unlike the strong belief materials. The effect in Experiment 2 can be explained with the modified version of the selective processing model (Stupple, Ball, Evans, & Kamal-Smith, 2011). In this account there is an initial step in which participants who respond with high logical accuracy construct all of the mental models whereas those with moderate logical accuracy construct only one mental model, following the process of seeking confirmation for believable problems and disconfirmation of unbelievable problems. This means that those with high logical accuracy do not demonstrate an interaction of logic and belief – they respond correctly in all conditions – but those with moderate logical accuracy do. It could be that the increased interaction here is a result of some participants constructing all of the mental models with no working memory load, but being limited to constructing a single mental model by the working memory load, which is the process that leads to the interaction. Overall, the selective processing model accounts for the main effects of belief bias well, and the modified selective processing model accounts for the effects of the working memory load.

The metacognitive uncertainty account can explain the findings. In Experiment 1 the complexity of the problems with strong and weak beliefs was the same and therefore similar levels of metacognitive uncertainty would be expected. However, it may be that if belief is weak then the cue for default responding will not have a large effect. As a result, there will be fewer belief-based responses with weak beliefs than strong. The metacognitive uncertainty account also provides a good explanation of the findings in Experiment 2. A concurrent working memory load would ensure that participants were more likely to exceed their working memory capacity when reasoning, and so more metacognitive uncertainty and therefore more belief bias would be expected with this manipulation. The effect would be particularly strong with indeterminate problems where metacognitive uncertainty is predicted to be higher anyway.

In summary, the selective scrutiny and misinterpreted necessity accounts of belief bias offer no explanation of these findings. The mental models account explains the findings of Experiment 1 but does not explain the effect of working memory load in Experiment 2 as well. The selective processing account explains the increased main effect of belief but only the modified selective processing account adequately explains the increased interaction of belief and logic with a working memory load. The metacognitive uncertainty account can explain these findings. In comparison to these theories, the explanation presented here about the effect that these experimental manipulations have on raising and lowering the activation levels of the initial conclusions in working memory fares well.

5.2 Comparison of the selective processing model and the activation level account of belief bias

Of the belief bias theories discussed, the activation level account presented here has greatest synergy with the selective processing model. Selective processing suggests that people attempt to construct a confirming model with believable problems and a disconfirming model with unbelievable problems and that this explains the interaction of logic and belief that is found in belief bias experiments. The activation level account suggests something similar, but for different reasons. The explanation here is that people tend use believable initial conclusions more frequently when solving both types of problem. However, a believable conclusion will confirm a believable problem and disconfirm an unbelievable problem. Hence the effect is the same as in selective processing and the interaction of belief and logic is explained in a similar way, but the explanation of why people do this is simpler. The selective processing model suggests that people sometimes use a disconfirming strategy, which is comparatively unusual thinking where a confirmation bias is more commonly found (e.g. Nickerson, 1998). It also proposes a search through models to find a confirming or disconfirming one, but this process is unspecified and is potentially complex. The proposal here is a less complex – it suggests that people generate a range of models and infer a range of conclusions (both confirming and disconfirming), but the unbelievable ones decay away and are forgotten, leaving believable conclusions which are confirming in the case of believable problems and disconfirming in the case of unbelievable problems.

Furthermore, the activation level approach naturally extends to explain the effects addressed by the modified selective processing model. The modified model adds another stage to the process in which people may be motivated to construct all the models. This is seen as a distinct step preceding whether they search for confirming or disconfirming models or not. This multi-step model accounts for the experimental data well, but leaves unexplained why some people decide to construct all models, or one model, or no models. The account presented here suggests that people tend to try and construct all of the models. Those with a large working memory capacity will remember all of the conclusions that follow from these models (generally leading to logically correct responses), those with a moderate working memory capacity will remember the believable conclusion (which will either confirm or disconfirm depending on the problem type, leading to an interaction of logic and belief) and those with a small working memory capacity will remember only their prior beliefs (leading to purely belief based responses). This replicates the effect of the multiple stages in the modified selective processing model, but with only mechanism – activation and decay in memory. Furthermore, this mechanism is based on wider architectural principles which have support indicating that they underpin cognition in a range of tasks. It is reasonable to assume therefore that they underpin reasoning too.

5.3 Dual process theory and the influence of activation levels on belief bias

Most contemporary accounts of belief bias are framed within dual process theory (e.g. Evans, 2008). This broad reasoning framework suggests that many reasoning phenomena, of which belief bias is a prime example, can be explained in terms of an interaction between rapid, automatic, heuristic processes (often referred to as System 1 e.g. (Stanovich & West, 2000) or Type 1 processes (Evans, 2009)) and slow, controlled, analytic processes (often referred to as System 2 or Type 2 processes). A debate exists about how these two systems or types of thinking interact. One possibility is that both processes occur in parallel, with conflict resolution required if they suggest different responses (e.g. Sloman, 1996; Handley et al., 2011). Another possibility is that the heuristic response is given as a default but a more analytic process may intervene subsequently (e.g. Evans, 2006). The selective processing model of belief bias is an example of the latter.

Extending this debate about the architecture of dual process theories, Evans (2009) proposes a distinction between type 1 and type 2 systems, each consisting of multiple processes rather than the simple distinction between just two types of process. Type 1 systems are comprised of purely type 1 (heuristic, implicit) processes whereas type 2 systems have some type 2 (analytic, explicit) processes but these may be supported by type 1 processes. Working memory is the defining feature of type 2 systems – if the system uses working memory to complete a task then it is a type 2 system, even if some type 1 processes are recruited in its operation. This theoretical proposal is of relevance here because it accurately describes a number of features of the ACT-R model outlined in this paper. Specifically, the principles governing memory are type 1 processes, but they are used to support the construction and use of mental models to draw conclusions, which is a type 2 process.

The type 1 process of relevance here is the mechanism through which activation levels are determined. The changes to activation level which influence memory are implicit; they reflect the environment that the person experiences and so are only indirectly influenced by the person through the choices they make about their environment. But these activation levels do influence what is remembered and then used explicitly in reasoning, specifically here in the evaluations of the conclusions. The evaluation is therefore a type 2 process, but the conclusions that are remembered and used are determined by a type 1 process. Therefore these two systems interact cooperatively during reasoning rather than competing to generate a response as implied by parallel dual process theories.

There is a further interesting implication of activation levels cueing certain conclusions for evaluation in this way. The level of activation of the conclusions follows from ACT-R theory's base level learning equation (Equation 1). This is based on the rational analysis of memory developed by Anderson and Schooler (1991). They observed that a memory system cannot efficiently provide unlimited access to all the items in memory, given the costs associated with retrieval. Therefore memory performs optimally by making available the items that are most likely to be useful. They demonstrated that this likelihood is based on past usage of a memory, and the base level learning equation influences the activation level to reflect this likelihood. Hence memory is optimised to retrieve what is most likely to be useful.

Therefore, based on the rational analysis of memory, the first conclusion retrieved is the one that is most likely to be relevant on the basis of previous usage. What is described as a bias based on the task set by the experimenter is actually a rational response for a memory system optimised to recall the information most likely to be relevant in a given situation. If a conclusion has been encountered often in the past it is more likely to be useful in the future, e.g. Schooler and Anderson (1997). The experiment has created the relatively artificial situation in which experiences from before the experiment (prior beliefs) should be ignored, but if the environment in which the reasoning was taking place was typical for a person, then these same memory processes would initially retrieve the conclusion that is most likely to be relevant. As Evans and Over (1996) point out, in typical reasoning situations it is rational (in terms of meeting your personal goals rather than following a normative system) to use all of your relevant beliefs when reasoning. The activation level model presented here exemplifies this principle.

The possible explanation of belief bias raised here also has a connection to the work of Oaksford, Chater and colleagues who have applied a rational analysis approach to reasoning (e.g. Oaksford & Chater 1994; 2007). Their approach has typically been to explore rational analysis accounts of reasoning itself, leading to computational level explanations of reasoning. The explanation here differs in that it is based on the rational analysis of memory, and how this is used in reasoning. Therefore it is an algorithmic level of explanation of the basis of belief bias which is still underpinned by rational principles of the more general cognitive architecture which are applied when completing a reasoning task. Given the similarities in these approaches, exploring the links between them might be an interesting prospect.

5.4 Limitations and future work

The goal of this paper has been to extend current explanations of belief bias by investigating the role of working memory in reasoning, in particular the effect of believability on the activation level of initial conclusions and their influence on conclusion evaluation. As a result, a large part of the explanation has rested on working memory and the retrieval stage rather than the more common explanations in relational reasoning which rely on mental model construction. In part this is deliberate, the claim made here is that the operation of working memory in relational reasoning and belief bias has been comparatively overlooked and a more detailed account has been provided. But the claim is not that reasoning is solely about memory. A number of aspects of this work could be developed further to provide a fuller account of belief bias in relational reasoning.

One area, although not emphasized, is the detailed account of the construction of mental models and the process of drawing initial conclusions from them that has been presented in this paper. This model has been developed in line with existing research, e.g. Schaeken et al. (2007), but could be subject to specific research itself. A more detailed account of the mental models constructed could be developed, for example drawing upon the selective processing account of mental model construction and combining it with the account of working memory in reasoning developed here.

A second area is to develop the model to account for the chronometric data that has provided new insight into existing theories (e.g. Ball, Phillips, Wade, & Quayle, 2006; Stupple & Ball, 2008). Within ACT-R theory activation level not only influences the likelihood of recall, as has been tested here, but also the speed of recall. Therefore this approach could naturally be extended to test chronometric predictions.

A third area is to consider other forms of reasoning. Markovits and colleagues have conducted several studies of causal conditional reasoning that demonstrate the role of retrieval (e.g. Quinn & Markovits, 1998; Quinn & Markovits, 2002; Grosset, Barrouillet & Markovits, 2005). Although this research is not based on the ACT-R theory used here, there is a parallel suggesting that retrieval processes are important when reasoning with concrete materials. Future work could extend this hypothesis to other forms of reasoning, such as syllogistic reasoning.

A final area of research is to extend the reasoning tasks beyond the standard reasoning tasks and their interpretations that are commonly used to investigate belief bias. This paper has a greater focus on the effects of belief on logical reasoning than on logical reasoning itself, and it may well be that participants' interpretation of the task is not as straightforward as is typically assumed here and within this literature. Furthermore, whilst the task used here is typical for this literature, there are alternatives that have not been sufficiently explored. For example, Stenning and van Lambalgen (2008) discuss defeasible logic. The ACT-R architecture is well suited to exploring this domain of reasoning using the mechanisms discussed in this paper.

5.5 Conclusion

A novel account of the role of working memory in relational reasoning and a mechanism through which it influences belief bias is proposed, based on the activation level of conclusions formed during reasoning. Prior beliefs influence the reasoning process by either merging with believable conclusions, raising their activation level and increasing the likelihood that they will be retrieved during reasoning, or when prior beliefs are retrieved directly and used to evaluate the presented conclusion. The theory governing the activation level of conclusions in working memory was drawn from ACT-R theory and so this account of belief bias was formalized in an ACT-R computational model. This model provided a good fit to existing data, reproducing the main belief bias phenomena in relational reasoning. It also led to novel predictions which were tested in two experiments. These predictions were confirmed and the ACT-R model provided a good fit to these data, supporting the idea that the activation level of conclusions formed during reasoning influences belief bias. This adds to current explanations of belief bias which do not provide a detailed specification of the role of working memory and how it is influenced by prior knowledge. How people retain and recall the initial conclusions during reasoning and the influence of prior knowledge on this process provides a good explanation of belief bias in relational reasoning.

Acknowledgements

I would like to thank Niels Taatgen, Leendert van Maanen, Keith Stenning, David Peebles and all the anonymous reviewers for helpful comments that greatly improved this article. I would also like to thank Jessica Fielding for assistance with data collection.

References

Anderson, J.R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29,* 313-341.

Anderson, J.R. (2007). *How can the mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, New Jersey: LEA.

Anderson, J. R. & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review, 104,* 728-748.

Anderson, J.R., & Reder, L.M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128,* 186-197.

Anderson, J.R., Reder, L.M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221-256.

Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396-408.

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory & Cognition, 38,* 928-940.

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

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Ball, L.J., Phillips, P., Wade, C.N., & Quayle, J.D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53,* 77-86.

Barton, K., Fugelsang, J., & Smilek, D. (2009). Inhibiting belief demands attention. *Thinking & Reasoning, 15,* 250-267.

Borst, J.P., Taatgen, N.A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal Experimental Psychology: Learning, Memory, and Cognition, 36,* 363-382.

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

Daily, L.Z., Lovett, M.C., & Reder, L.M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25,* 315-353.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428-433.

Evans, J.St.B.T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454-459.

Evans J.St.B.T. (2006). The heuristic-analytic theory of reasoning: extension and evaluation. *Psychonomic Bulletin Review, 13,* 378-95.

Evans, J.St.B.T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement.* Hove: Psychology Press.

Evans, J.St.B.T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255-278.

Evans, J.St.B.T. (2009). How many dual-process theories do we need? One, two, or many? In J.St.B.T. Evans & K. Frankish (eds.), *In two minds: Dual processes and beyond* (pp. 33-54). Oxford University Press: Oxford.

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., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning,11,* 382-389.

Evans, J. St. B. T., Handley, S. J., & Bacon, A.M. (2009). Reasoning under time pressure: A study of causal conditional inference. *Experimental Psychology, 56,* 77-83.

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., Allen, J.L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6,* 263-285.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction.* Hove: Psychology Press.

Evans, J.St.B.T., & Over, D.E. (1996). *Rationality and reasoning.* Hove, UK: Psychology Press.

Gilhooly, K.J. (1998). Working memory, strategies, and reasoning tasks. In R.H. Logie & K.J. Gilhooly (Eds.), *Working memory and thinking* (pp. 7-22). Hove, UK: Psychology Press.

Gilinsky, J.S., & Judd, B.B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging, 9,* 356-371.

Goel, V., & Dolan, R.J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87,* B11-B22.

Goel, V., & Vartanian, O. (2011). Negative emotions can attenuate the influence of beliefs on logical reasoning. *Cognition and Emotion, 25,* 121-131.

Goodwin, G.P., & Johnson-Laird, P.N. (2005). Reasoning about relations. *Psychological Review, 112,* 468-493.

Grosset, N., Barrouilet, P., & Markovits, H. (2005). Chronometric evidence for memory retrieval in causal conditional reasoning: The case of the association strength effect. *Memory & Cognition, 33,* 734-741.

Handley, S.J., Newstead, S.E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 28-43.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, UK: Cambridge University Press.

Klauer, K.C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852-884.

Lewis, R.L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29,* 375-419.

Macpherson, R., & Stanovich, K.E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences, 17,* 115-127.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817-835.

Newstead, S.E., Pollard, P., Evans, J.St.B.T., & Allen, J.L. (1992). The source of belief bias in syllogistic reasoning. *Cognition, 45,* 257-284.

Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175-220.

Oakhill, J.V., & Johnson-Laird, P.N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology, 37A,* 553-569.

Oakhill J., Johnson-Laird P.N., & Garnham A. (1989). Believability and syllogistic reasoning. *Cognition, 31,* 117-40.

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

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford: Oxford University Press.

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.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68,* B93-B101.

Quinn, S., & Markovits, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking and Reasoning, 8,* 179-191.

Rips, L.J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Roberts, M.J. (2000). Strategies in relational inference. *Thinking and Reasoning, 6,* 1-26.

Roberts, M.J., & Sykes, E.D.A. (2003). Belief bias and relational reasoning. *Quarterly Journal of Experimental Psychology, 56A,* 131-154.

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors, 48,* 362-380.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60,* 205-234.

Schaeken, W., Van der Henst, J., & Schroyens, W. (2007). The mental models theory of relational reasoning: Conclusions' phrasing, and cognitive economy. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental model theory of reasoning: Refinements and extensions* (pp.129-150). Mahwah, New Jersey: LEA.

Schooler, L.J., & Anderson, J.R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology, 32,* 219-250.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3-22.

Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.

Stupple, E.J.N., & Ball, L.J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection time evidence for a parallel-process model. *Thinking & Reasoning, 14,* 168-181.

Stupple, E.J.N., Ball, L.J., Evans, J. St.BT., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology, 23,* 931-941.

Torrens, D., Thompson, V.A., & Cramer, K.M. (1999). Individual differences and the belief bias effect. *Thinking & Reasoning, 5,* 1-28.

Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: An rTMS study. *Neuroimage, 58,* 640-646.

Table 1

Means and standard deviations of the believability of conclusions used, and the difference in ratings between Strong and Weak conditions

	Strong		Weak			
	M	S.D.	M	S.D.	t(27)	p
Believable Valid	6.50	0.65	5.72	0.89	4.29	0.001
Unbelievable Valid	1.11	0.28	1.60	0.55	-4.68	0.001
Believable Determinately Invalid	6.55	0.79	5.95	0.83	3.29	0.003
Unbelievable Determinately Invalid	2.24	0.88	2.82	0.94	-2.76	0.01
Believable Indeterminately Invalid	6.28	0.99	5.99	1.09	1.56	0.13
Unbelievable Indeterminately Invalid	1.30	0.56	1.71	0.74	-3.76	0.001

Fig. 1. Percentages of conclusions accepted by condition for strong and weak beliefs, and model predictions. The error bars represent standard error.

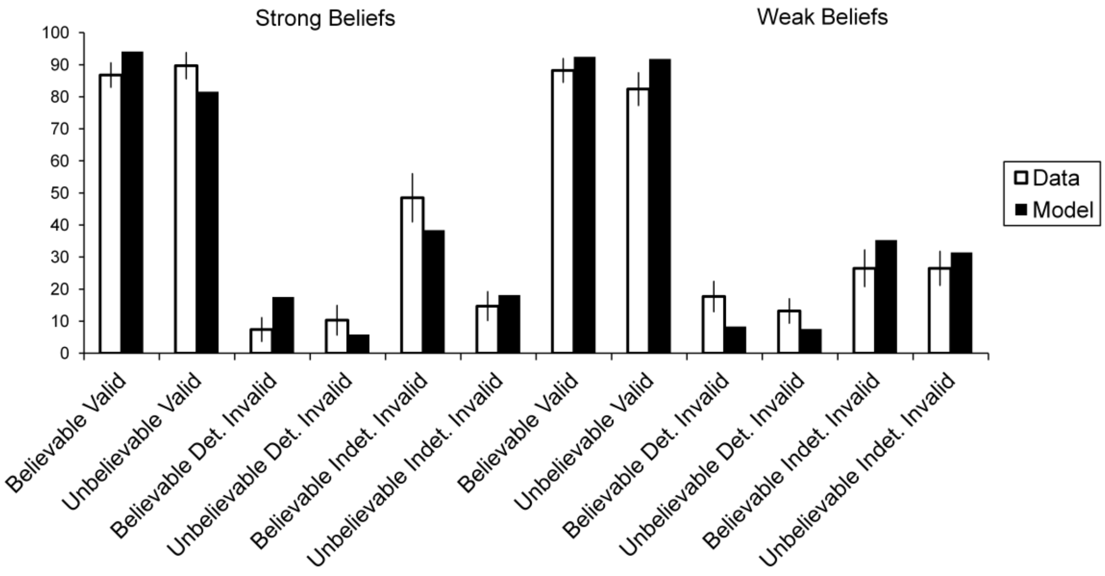
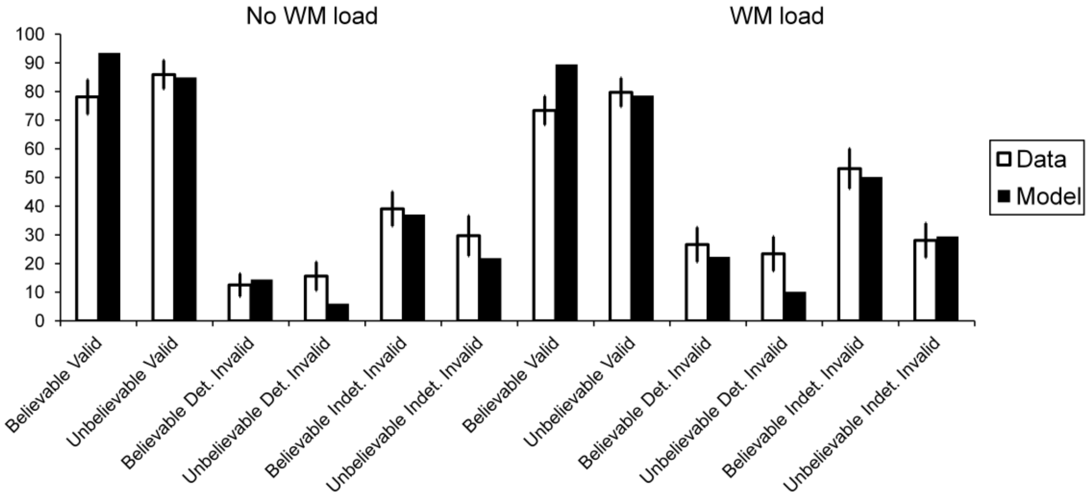


Fig. 2. Percentages of conclusions accepted by condition with and without a working memory load, and model predictions. The error bars represent standard error.


No WM load
WM load
□ Data
■ Model
Believable Valid
Unbelievable Valid
Believable Det. Invalid
Unbelievable Det. Invalid
Believable Indet. Invalid
Unbelievable Indet. Invalid
Believable Valid
Unbelievable Valid
Believable Det. Invalid
Unbelievable Det. Invalid
Believable Indet. Invalid
Unbelievable Indet. Invalid