MODELLNG A FRACTIONATED SYSTEM OF DEDUCTIVE REASONING OVER CATEGORICAL SYLLOGISMS

Gregory Giovannini

A Thesis submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Arts

Graduate Program in Psychology, York University, Toronto, Ontario

April 2017

Abstract

The study of deductive reasoning has been a major research paradigm in psychology for decades. Recent additions to this literature have focused heavily on neuropsychological evidence. Such a practice is useful for identifying regions associated with particular functions, but fails to clearly define the specific interactions and timescale of these functions. Computational modelling provides a method for creating different cognitive architectures for simulating deductive processes, and ultimately determining which architectures are capable of modelling human reasoning. This thesis details a computational model for solving categorical syllogisms utilizing a fractionated system of brain regions. Lesions are applied to formal and heuristic systems to simulate accuracy and reaction time data for bi-lateral parietal and frontotemporal patients. The model successfully combines belief-bias and other known cognitive biases with a mental models formal approach to recreate the congruency by group effect present in the human data. Implications are drawn to major theories of reasoning.

Dedication

Dedicated to family of the present and future

Acknowledgments

Thanks to the Goel Reasoning Lab for support,

And Sashank Varma for provision and support with the 4CAPS architecture,
and guidance in statistically analyzing the effectiveness of computational models.

Table of Contents

List of Figures

Introduction

Reasoning is the cognitive activity of combining and processing given information to generate inferences. Inferences take one or more propositional statements comprised of the given information (the premises) to provide justification for accepting some conclusion. When the given information presents a complete picture of a situation this is ideally a deductive reasoning process. Deductive reasoning – sometimes referred to as top-down logic – applies general law-like rules to information in order to build down to or establish what must be true about a specific instance. In contrast, inductive reasoning (bottom-up logic) takes specific instances of information and attempts to build up to possible general rules. The conclusions of inductive reasoning are always open the possibility of being wrong due to generalization from an incomplete problem space. Deductive arguments, however, can be deterministically evaluated in terms of their validity. If a deductive argument is *valid,* the conclusion is a logical entailment of the premises which must be true assuming that the premises themselves are true.

As an example of inductive reasoning, if one has observed a large number swans to all be white then one may take these specific experiences to suggest a general rule that all swans are white. In deductive reasoning, a general rule may state that all swans are white. If a creature is a swan, it logically and unavoidably follows that the creature is white. This is true provided that the general rule is actually correct and not built from inductive reasoning processes.

Deductive reasoning possesses a special quality where its inferences are separable from its content. We can do this with the above argument by representing swans with X and the property of whiteness with Y. If All X are Y, then if some token creature Z is X it must also be Y. The ability to separate these conclusions from their content is what, according to Goel (2009), makes deductive reasoning a good candidate for being a self-contained higher-level cognitive reasoning module.

1.1 Single or Multiple Module Reasoning

Theories regarding the structure of reasoning processes in the brain can invoke single deductive reasoning modules, or a collection of modules. Mental logic (Rips, 1994) and mental models (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird, 1983) represent two prominent theories of single deductive reasoning modules. The two theories diverge in how they represent information, and the neural networks these representations would invoke. Mental logic theories suggest reasoners understand the inferential role of logical terms (all, some, no, if, and, or, etc.) and that a linguistic representation of these terms is what drives processing. As the procedure for mental logic involves rules of inference applied to syntactic strings, it should demonstrate engagement of left prefrontal and superior temporal brain regions for language-based information processing (Goel, 2009). In contrast, mental model theories suggest we build spatially-based mental representations of potential situations to comprehend and process logic problems. Results supporting a mental model theory would instead show recruitment of a visuospatial (parietal/occipital) network (Barbey & Barsalou, 2009; Goel, 2005).

If reasoning in the human brain is instead characterized as a collection of modules, this collection may cooperate to perform inductive heuristic-based reasoning on information present in the problem, as suggested by the *simple heuristics* (Gigerenzer & Todd, 1999) paradigm. However, it may instead be organized so that specific modules in the collection to respond to particular cues in a somewhat reflexive manner suggested by the *massive modularity* (Fodor, 1983; Carruthers, 2006) paradigm.

Under the *simple heuristics* view, all reasoning is performed by an interconnected collection of fast and frugal heuristics. Gigerenzer and Todd (1999) specify three sets of heuristics which together form a computationally cheap solution for any reasoning problem. The first set involves heuristics for guiding the search for alternatives of choice and their relevant information. The second involves heuristics for stopping the search procedure. The last involves heuristics for actually making the decision from among the alternatives found. *Simple heuristics* appeal to evolution in their genesis; it is said

evolution would “seize upon informative environmental dependencies […] and exploit them with specific heuristics” (Gigerenzer & Todd, 1999) which may either be new, or the result of recombining or nesting old heuristics.

Massive modularity suggests the reasoning mind is a fractionated collection of specialized reasoning modules tuned to specific tasks. Evolutionary psychologists supporting a highly modular view of the mind (Cosmides & Tooby, 1992; Pinker 1997; Sperber, 1994) suggest evolution incrementally added to this repertoire of reasoning modules: this includes modules for “semantic inference, communicative pragmatics, social exchange, intuitive numbers, spatial relations, naïve physics, and biomechanical motion” (Barbey & Barsalou, 2009). These modules apply a small number of inputs to a limited internal database in the generation of output. In processing only a limited range of input such modules are said to be *informationally encapsulated*. This is bears the consequence of *cognitive impenetrability*, meaning that we are not consciously aware of nor able to influence their processing. To exhibit these traits, strong modular views predict neural systems of reasoning to be highly localized.

These two divergent theories arise from a similar desire to overcome the problem of computational intractability. Simon (1983; 1991) introduces the notion of *bounded rationality* in stating that finding an optimal solution to decision-making problems (such as through pure deduction) is too expensive. We have limited time, memory, and processing ability with which to make decisions, and as such we cannot evaluate all possible alternatives. If finding the best solution is unfeasible, we must make due with approximate methods for quickly finding solutions that are good enough. Combining the words “satisfy” and “suffice”, Simon (1955) terms this approach *satisficing*. This represents the idea that once an alternative is found that is appealing enough to meet some *aspiration level* we stop our search and go with that alternative.

Gigerenzer and Todd (1999) incorporate this *satisficing* heuristic among others in the decision-making process to limit the problem space – the amount of information considered – and arrive at cheap yet effective decisions. The massively modular approach limits the problem space using specialized modules with limited interconnectivity to respond to particular cues and information in a particular way

without analyzing all possible information. However, in an attempt to reduce computational complexity and facilitate fast decision-making, these two views have seemingly given away the ability to try and find exact solutions altogether – the capacity for deductive reasoning. The massively modular account fragments and isolates the reasoning mind to the point where deliberate analytical investigation becomes insupportable. Meanwhile, a reasoning system built purely from heuristics seems only capable of supporting a sense of intuition guided by environmental cues. Whether our reasoning is driven by rigid and reflexive responding in a cognitively impenetrable way, or through purely inductive intuition, neither of these approaches appear to be coherent with our normative views of rationality. Namely, that a rational choice is not simply a selection, but a selection for a reason (Bermudez, 2002), which implies reasoning is a thoughtful process, unlike that of an eye-blink reflex (Goel, 2009).

1.2 Dual-system theory

Dual-system theory (Evans 2003; Evans & Over, 1996; Stanovich 2004) provides space for the collective module approaches under the branding of system 1, and deductive reasoning processes like mental models or mental logic under system 2. System 1 provides a collection of parallel-processes which have been considered fast, automatic, or associative. This system can contain rigid and evolutionarily-specified processing modules similar to the informationally encapsulated Fodorian modules. It can also encompass the fast and associative heuristics processing system (De Neys, 2006). Formal rules, stimulus discrimination, and decision-making choices practiced to the point of automaticity (Kahneman & Klein, 2009) can also be said to be a part of System 1. As such, later clarifications of dual-system theory (Evans & Stanovich, 2013; Stanovich 2011) suggest system 1 is considered as a plurality of autonomous systems.

System 2, by contrast, is a slower serial process that is more often rule-based. It is far more limited in its processing capacity in requiring considerable conscious thought, effort, and working memory resources. It is through this system that true logic-based reasoning approaches are executed. The two systems are sometimes supposed to have a *default-interventionist* (Evans 2007) structure, which specifies that System 1's intuitive responses are the default upon which System 2 may or may not intervene. Other theorists (Barbey & Sloman, 2007) suggest a *parallel-cooperative* structure with each system providing its input with a following conflict resolution process. The current variation in dual-system theory is quite large; in its most explicit state, each system may be ascribed numerous attributes including the characterization of system 1 as the evolutionarily "old mind", and system 2 as the "new mind". At its more conservative side, System 1 represents various autonomous processes while System 2 requires substantial working memory resources and hypothetical thinking (e.g. thought experiments/model building) – the latter demand which Stanovich (2011) calls *cognitive decoupling*.

1.3 Current Neuropsychological Data

Current neuropsychological findings from brain imaging data (Goel et al., 2000) have shown a number of divisions in the networks recruited when reasoning. These differences emerge depending upon whether or not the content of problem is familiar to reasoners, whether the content is in conflict with held beliefs or not, and differing when the reasoning problem is presented in an informationally complete or incomplete manner. These findings undermine the acceptance of a single reasoning system like mental models or mental logic, and suggest a diverse collection of brain regions involved in logical reasoning.

Common methodology employed in past research investigating the neural basis of reasoning, such as having subjects solve puzzles or perform other tasks while undergoing brain imaging, is good at identifying neural systems involved in reasoning, but it does little explain the interactions between these systems. A double-dissociation in lesion studies can tell us that some brain area is important to some task A but not B, and that another area is important to task B but not A; this brings the suggestion that these areas support some different underlying mental function (Dunn & Kirsner, 2003) – though precisely how an area contributes, or the time-scale or steps of the function generally can be difficult to determine. These questions remain open.

Computational modelling is one way of representing a complex interaction of different modules to answer these questions. Different models can more precisely describe how these areas may function, and their results can be compared to neuropsychological data or the data of other computational models. The present thesis provides a computational model of deductive reasoning along the lines indicated by imaging data, and tests its predictions by comparing how it performs (in terms of accuracy and reaction time) under conditions of simulated lesion with that of archival lesion data from the Vietnam head injury study (VHIS).

Imaging studies have shown a division in the brain networks recruited depending upon whether the material is familiar or not, otherwise known as the content effect (Goel et al., 2000). In the cited study, eleven subjects solved categorical syllogisms (a type of deductive reasoning problem) by indicating the problems to be valid or invalid;

propositional statements possessed familiar and meaningful content (e.g. “All swans are white”) or unfamiliar content (e.g. “All X are Y”). A left-lateralized frontal-temporal language system (BA 21/22/47) appears to preferentially process familiar and conceptually-coherent material (see Figure 1), while a bilateral parietal visuospatial system (BA 7/40) processes unfamiliar or content-free material (see Figure 2). In both familiar and unfamiliar conditions the left prefrontal cortex is also active, evidencing its value to general reasoning process.

It has been described that the validity of deductive reasoning problems are independent of their content. As the activation patterns in the brain diverge depending upon the content of the syllogisms, this suggests purely deductive processes – such as through the use of only mental models or mental logic processes – fail to fully describe reasoning over these types of problems. Furthermore, while mental models theory predicts visuospatial systems to be necessary and sufficient conditions for deductive reasoning and mental logic theories predicts linguistic systems to be necessary and sufficient, these predictions fail to hold across conditions of familiar and unfamiliar content. Dual-system theories thus far would still hold by relegating familiar content activations to System 1 heuristic operations and the bi-lateral parietal recruitment to System 2 rule-based operations.

Reasoning with familiar material

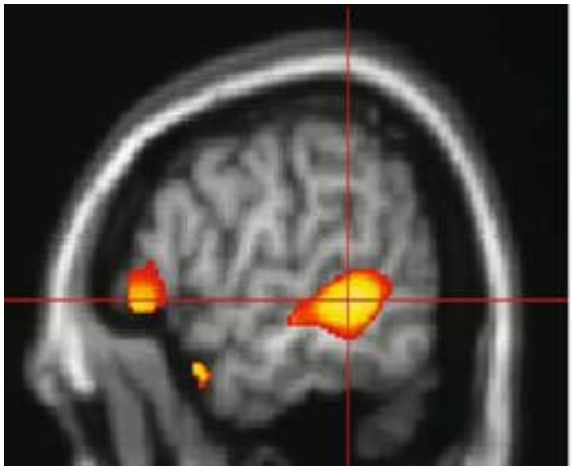


Figure 1

Reasoning with unfamiliar material

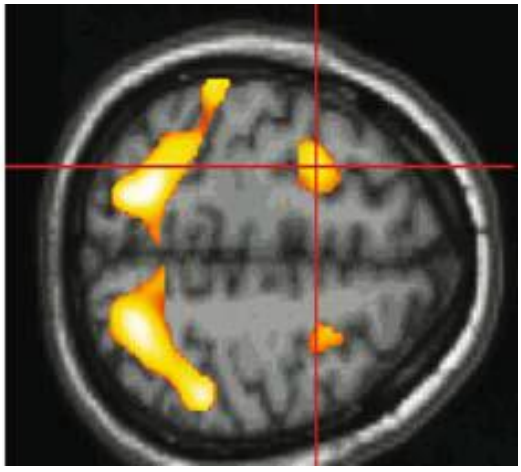


Figure 2

(reproduced from Goel et al., 2000)

Another aspect to the effect of content on logical reasoning is that subject performance is significantly higher when the deductive validity of the problem is consistent with held beliefs, and considerably lower when it is not. A valid conclusion suggesting “All men are mortal” would be congruent with held beliefs, while a valid conclusion stating “All men are evil” would, hopefully, contradict held beliefs. Inhibitory or incongruent trials, where the validity of the problem does not match held beliefs, show different patterns of activation depending upon the success of the subject. When conflict between logical inference and belief is detected, belief-bias responding must be inhibited and formal reasoning mechanisms are to be given preference. Conflict detection is associated with activation of right lateral/dorsal lateral prefrontal cortex (BA 45, 46) (see Figure 3; Goel & Dolan, 2003), whereas belief-biased responding is associated with ventromedial prefrontal cortex activation.

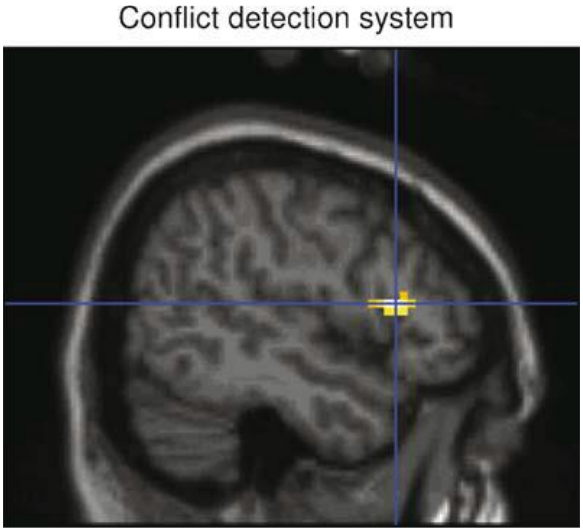


Figure 3

The final aspect of the deductive reasoning system to be discussed involves a hemispheric asymmetry in reasoning with complete versus incomplete information. Goel et al. (2006) utilized a 3-term transitive inference task to test neurological patients with focal unilateral frontal lobe lesions. A double dissociation was found where patients with lesions to the left prefrontal cortex were selectively impaired on complete (determinate) trials, and patients with right prefrontal cortex lesions were impaired in incomplete (indeterminate) trials (see Figure 4). Figure 5 displays the behavioural results for this study (Goel et al., 2006). This functional dissociation between left and right prefrontal

cortex is difficult to reconcile with previously explored theories of deductive reasoning; they do not necessitate neuronal system differences depending upon complete or incomplete information.

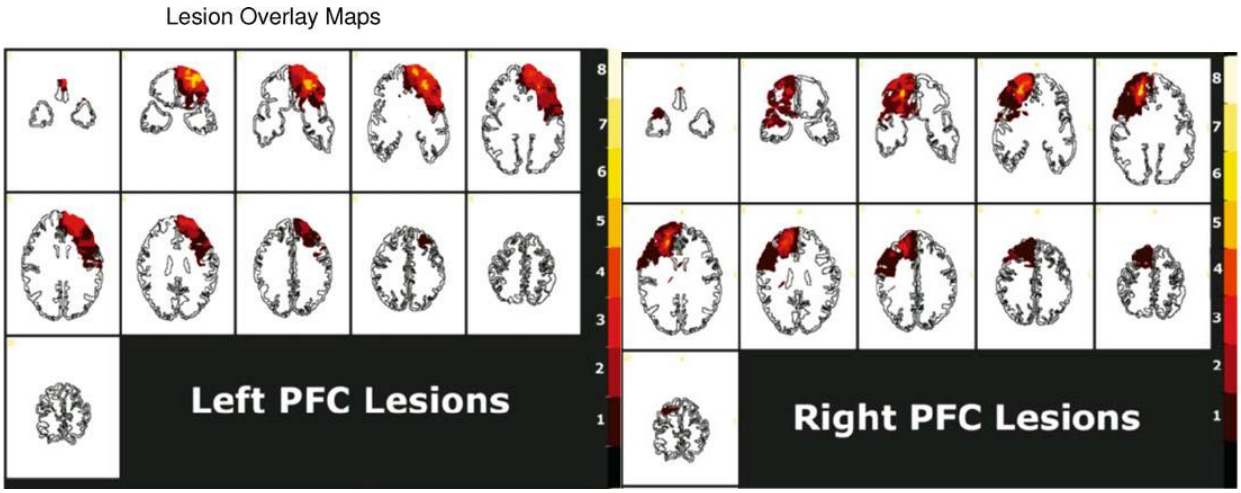


Figure 4 (reproduced Goel et al., 2006)

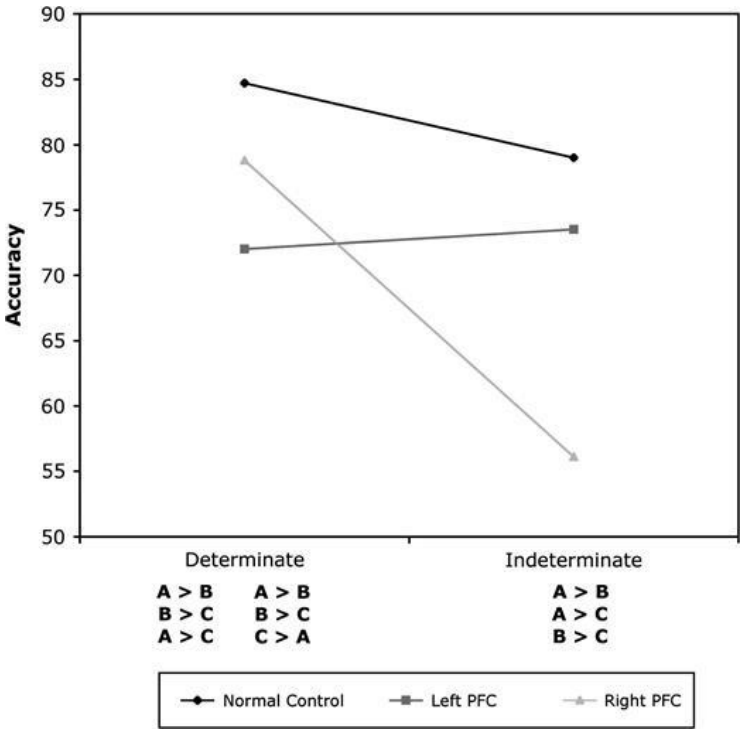


Figure 5 (reproduced Goel et al., 2006)

Goel (2009) puts forth a new framework for conceptualizing the deductive reasoning system to better explain current neuropsychological data, and break down the dichotomous implications of many dual mechanism theories. Termed as a

fractionated system of deductive reasoning, it combines a left prefrontal cortex general pattern completer with right prefrontal cortex systems for conflict detection and uncertainty maintenance; it includes a left frontal-temporal system for heuristic or conceptual processing, and a bilateral parietal system for formal operations (see Figure 6). These divisions may of course not be the complete list of what may be found for the deductive reasoning system, but it is a step towards a more dynamic and interactive connection of systems than can be provided by any other account – including dual mechanism theories. In featuring a number of systems able to inhibit or facilitate the activity of others, a much greater degree of variability in performance can be generated. This is particularly important for explaining deductive reasoning processes for solving categorical syllogisms, as a study by Bucciarelli and Johnson-Laird (1999) found the best participant to be correct on 95% of problems, and the worst to be correct only 25% of the time.

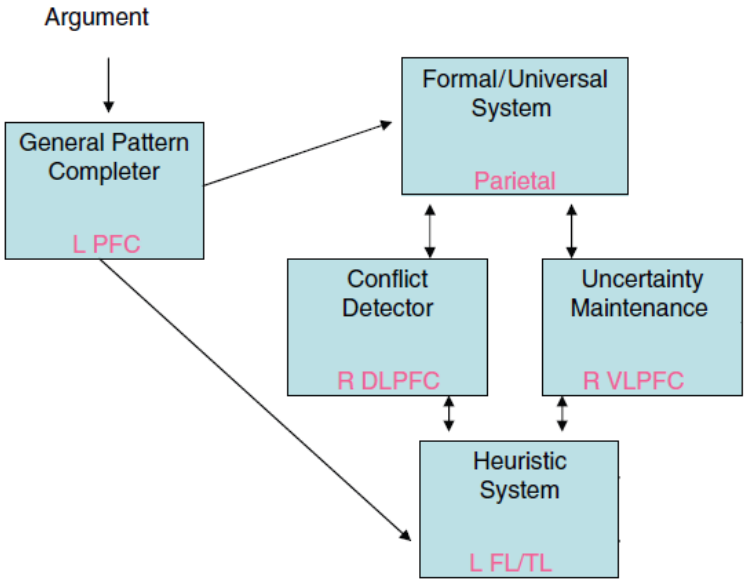


Figure 6

Categorical syllogisms are the type of reasoning problem for which control and patient data will be used to test the computational model. They are one type of deductive reasoning problem as they provide an informationally complete picture of the problem space. Categorical syllogisms consist of two premises and a conclusion. Each premise establishes relations between two terms (x and y) in one four 'moods' (A, E, I or O) where A = All x are y, E = No x are y, I = Some x are y, and O = Some x are not y. A

& E are universal moods describing how all x relate to y, while I & O are particular moods describing only how some x relate to y. Each premise must contain one 'end term' (a or c) and one 'middle term' (b), where the middle term is used to logically connect the premises so that the conclusion can state some relation between the two 'end terms' using one of the four syllogistic moods. There are also four figures which represent the possible orders in which the terms may occur in the premises. This allows for a total of 256 distinct forms for the categorical syllogism. In the example below, the premises and conclusion are set in the A mood. The middle term (B – the term present in both premises) allows us to determine relations between pigeons (C) and flight (A). If all birds are capable of flight, and all pigeons are birds, then all pigeons can fly.

Example Categorical Syllogism:		Abstract Terms
Premise 1 (Mood A):	All birds can fly	All B are A
Premise 2 (Mood A):	All pigeons are birds	All C are B
Conclusion (Mood A):	All pigeons can fly	All C are A

The goal of this thesis is to create a computational model of this fractionated deductive reasoning system for solving categorical syllogisms. Accomplishing this would provide some evidence for such a system to be computationally feasible in the explanation of human performance. This is done by modelling the performance of neurological patients as well as healthy controls on syllogistic tasks. In simulating fluctuations of performance due to brain damage or belief-bias it could be argued a fractionated reasoning system can support deductive reasoning processes. The performance of patients and controls from archival VHIS data and that provided by the computational model will be statistically compared on a number of factors.

Limitations with the patient data restrict the model versus data comparisons to control groups, frontotemporal lesion groups (heuristic system), and bilateral parietal lesion groups (formal system). Overall differences in accuracy and reaction time due to lesion group will be examined. The formal system will employ the mental models approach, which bears a strong prediction that single model problems will be solved faster and easier than multiple model problems (Johnson-Laird, 1983). This approach

will be provided some verification by having the computational model's estimation of which problems are single or multiple model problems tested to see if this distinction yields significantly different accuracies and reaction times for model and human data. Explaining the difference between the two is difficult without explaining the process, so this discussion is reserved for section 4.1.

As the formal and heuristic systems are the primary focus of this investigation, the effects of a conclusion's validity being congruent or incongruent with held beliefs (the congruency effect) is of primary importance. As an example, the proposition 'Some males are children' would be congruent with a person's held beliefs, while the statement 'All dogs can fly' would be incongruent with belief. Belief-bias will be examined to see if the strength of its effect changes due to lesion grouping.

Lastly, the wide variability of comparable individuals' performance on categorical syllogisms suggests the possibility of differences in cognitive processing style, with some reasoners being slower and more deliberately analytical, while others' decision-making is more quickly decided by heuristics and biases. Previous research suggests that potential interaction effects of congruency with the dependent variables of accuracy and reaction time may be moderated by overall differences in cognitive style between participants (Stupple, Ball, Evans, & Kamal-Smith, 2011). Highly logical subjects are thought to be more likely to inhibit belief bias effects and take more time to solve belief-logic conflict problems, especially for problem with conclusions that are believable but invalid. The reaction time prediction investigated by Stupple et al. (2011) will be investigated for the human data only, as the computational model will provide no implementation for differences in cognitive processing style. Programming such a device would demand numerous complicated assumptions beyond the scope of the present study.

1.4 What is 4CAPS

The computational model presented in this thesis is built on top of 4CAPS and written in LISP – a programming language popular for use in the design of artificial intelligence systems. The Collaborative Activation-based Production System (CAPS) is designed to model high-level cognitive functions at the cortical level. It has been used for simulating sentence comprehension, mental rotation, and problem-solving tasks such as the Tower of London and the Tower of Hanoi (Just & Varma, 2007; Varma, 2006). 4CAPS models the cortical constraints of information processing across multiple brain areas. It does this using a hybrid symbolic-connectionist architecture. The symbolic aspect is apparent in its use of production systems to represent cognitive processes. Production systems consist of an 'if-then' condition-action pair. If a collection of variables are storing particular values, then a number of actions will take place. A visual cat-detection module may have conditions that if it is furry, has a tail, and has pointed ears, then signal it is a cat. Working memory elements (WMEs) are used by 4CAPS to store a list of variables to be searched through by productions. When all the conditions of a production are satisfied it may modify WMEs or generate new WMEs. All productions are fired in parallel, where one check of all productions is considered to be one program cycle. Production systems are excellent for performing deductive tasks, though on their own there is little room for variability.

It is the connectionist aspect which provides these additional degrees of freedom. Individual WMEs are to be understood as a neural cluster supporting some cognitive representation (e.g. a word or object) which possesses its own firing rate. This activity level, typically varying from 0 to 1, can be increased or decreased similar to excitation and inhibition as a result of modification by conditionally satisfied productions. The spreading of activation from one WME to another (or many-to-many, etc.) will often have a stronger effect if the initial set of WMEs are at a higher activity level. To continue using the example of a cat detection module, the presence of the various cat-like features mentioned may have an additive excitatory effect on a cat-detector WME, while unlikely features being present (like having a hard shell) would fire productions that

inhibit this cat-detector WME. If the sum of these activations were above some minimal threshold, another if-then production may signal the presence of a cat (see Figure 7).

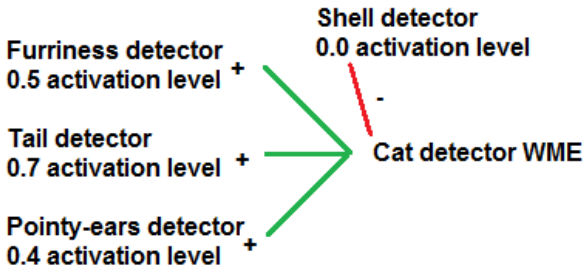


Figure 7

How these activation levels change do not depend only upon other working memory elements. 4CAPS is intended to model the interaction between multiple brain areas to facilitate cognitive functions. That being the case, the total level of activation within any particular center cannot exceed the cortical supply or capacity of that area. When this capacity is exceeded, the activation level of all WMEs will be scaled down to match cortical supply. Since the conditions of productions can include not just the content of a WMEs values – it may also require the WMEs be at some minimal activation level to be detected – a scaling down of activation may stop some productions from firing when they would have otherwise. This is how 4CAPS can simulate human errors, forgetting information, or a slowing down of processing under cases of computational overload.

All centers of the brain in a 4CAPS system have their own activation capacity. This capacity can be reduced to simulate impairment by a brain lesion. This will result in the center having to work harder and be more likely to introduce errors. Different brain areas can also share computational resources in situations of high computational demand. However, when a different brain center is forced to take on the typical computations of another it may not be as efficient in its processing. WMEs can be organized under classes where each brain area can have its own specified degree of specialization for processing these classes. An area with a specialization level of 0.5 will consume twice the activation to perform the same activity as an area with a 1.0 specialization level. For the sake of simplicity and the avoidance of unwarranted assumptions, the modules of this current computational model are functionally distinct.

While the areas certainly influence each other's processing through excitation and inhibition, one region will not attempt to take on the functional role of another if it becomes overloaded.

In summary, 4CAPS is a cognitive neuro-architecture for modelling cortical constraints on information processing. Its mechanisms combine deterministic production systems with analog activation levels. Group differences can be represented by variations in the processing capacities of brain areas and the effectiveness of their intercommunication. These differences alter the connection weights between nodes with the result that WMEs may fail to reach critical thresholds. This architecture allows 4CAPS to simulate human error under cases of computational overload (Varma, 2014).

Methods

2.1 Study Population

The human data for this thesis comes from archival data collected from male patients and controls as part of the Vietnam head injury study (VHIS – for more detail see Raymont, Salazar, Krueger, & Grafman, 2011). All patients received penetrating head injuries during service at Vietnam in the 1960s. To be included in the selection for the patient groups subjects needed to be above a minimal threshold of damage in key Brodmann areas (BA) identified by previous research; demonstrating a 10% proportion of damage in a single critical BA, or combined across the few relevant BAs for that system (see Appendix A). To be included in a particular functional grouping, patients were also required not to possess significant damage in another key functional system. Parietal patients, for example, could not be significantly damaged in key left-frontotemporal regions. For patient groups requiring damage to only the left hemisphere, those with damage to the right hemisphere were excluded. In accordance with the findings of Goel et al. (2000) on neural dissociations due to familiar and unfamiliar content, the parietal group required damage in either or both hemispheres in BA 7 and BA 40 (Goel et al., 2000), while the frontal-temporal group needed to show damage in key left-hemisphere areas (BAs 21/22/47).

The number of patients surviving selection for uncertainty maintenance (BAs 44/47) and conflict detection systems (BAs 45/46) was extremely low due to overlapping damage in other key areas, this was especially true for damage being present in both of these systems as they are close in proximity. Due to these selection difficulties, statistical comparisons of performance provided below was restricted to comparisons of the formal and heuristic systems with a primary focus on differences between congruent and incongruent syllogisms. Patients who did not respond to 33% or more of the syllogisms were also excluded. The demographics for the subject groups (see Appendix B) concerning age, level of education, and measures of memory and intelligence showed no significant differences among subject groupings. Figures 8 and 9 show overlay images illustrating the areas of damage for the final frontotemporal and parietal

groups, where lighter colored regions show areas of greater damage overlap across patients.

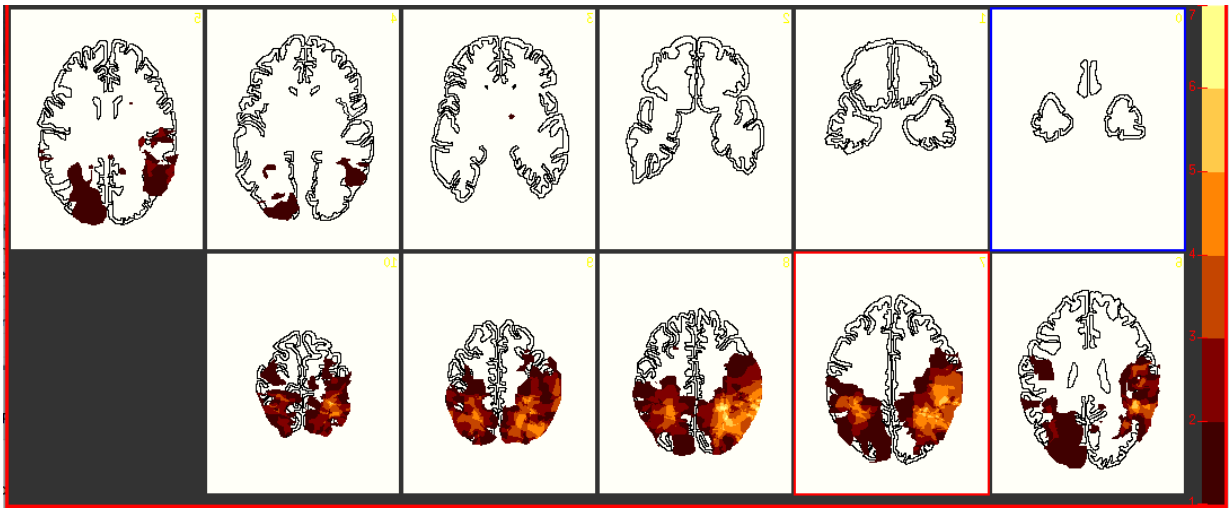


Figure 8 Bilateral Parietal Lesion Group

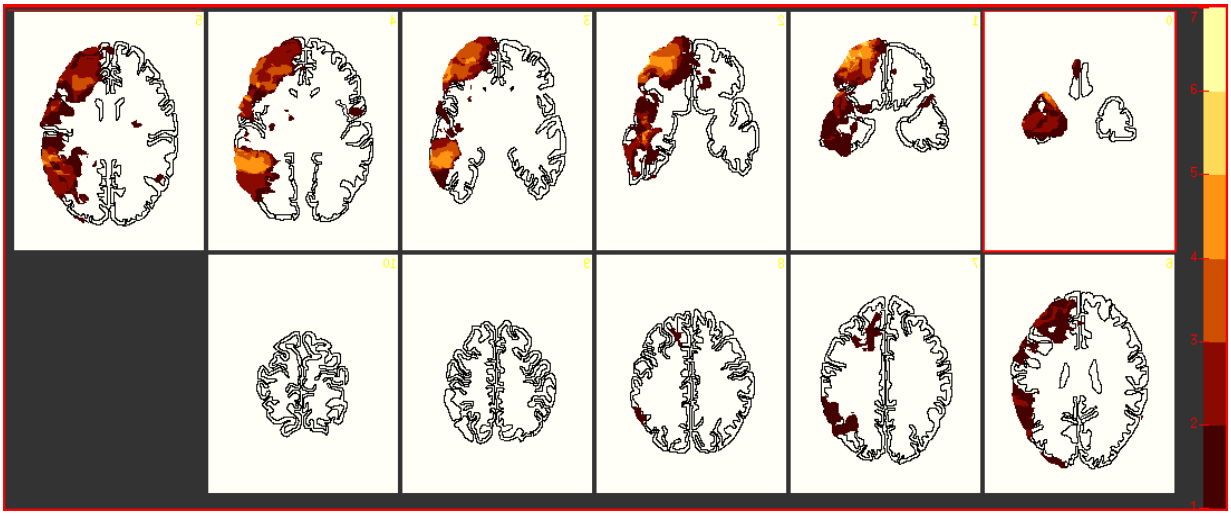


Figure 9 Left Frontotemporal Lesion Group

2.2 Original Task

In the task relevant to this study, brain damaged patients and healthy controls solved 19 content-free and content-imbued categorical syllogisms. While the responses in the content-free conditions were used in adjusting parameters of the model, the thesis focuses on belief-bias, and as such the subsequent analyses only concern the content-imbued data. Participants were instructed to determine whether or not the given conclusion followed logically from the two premises (i.e. if it was logically valid) with the assumption that the premises are true. Participants were told to press the 'C' key if they believed the conclusion to be valid, and the 'M' key if they thought it to be invalid. For each trial, the premises and conclusion were presented simultaneously, and they remained on screen until a response was indicated, which advanced them to the next syllogism. There was no limit to the amount of allowed time for a particular trial. Two blocks of problems were presented, one consisting of syllogisms and the other an equal number of operators (which are not syllogisms), where the trials within each block were randomized to prevent order effects. Afterwards, ratings indicating the general believability of the conclusions solely on the basis of their content were also collected on a scale from 1 (very unbelievable) to 5 (very believable).

Human Data

3.1 Human Data Results

In an initial analysis of the data, two separate one-way ANOVAs were conducted to examine the basic relationship between lesion group (IV) and one of the two dependent variables: these variables consisted of reaction time measured in milliseconds, and accuracy as a proportion ranging from zero to one. The arcsine transformation [arcsine($\sqrt{x}$)] was applied to all accuracy data in this study; this transformation is one method for addressing the problems associated with proportions over count based data, particularly those where values may fall below 0.3 or above 0.7. Where such extreme values may occur, in a variable bounded between zero and one, problems may emerge such as in the interpretation of confidence intervals, which may extend beyond this range and become meaningless. This transformation pulls out the ends of the distribution and provides a new range of π; it also provides correction for the occasionally observed departures from homogeneity of variance in the untransformed accuracy data. Key statistics such as F ratios or their associated p-values are derived from transformed variables, though graphs and descriptive statistics such as means or standard errors are expressed in untransformed terms for ease of interpretation.

The preliminary one-way ANOVA tests investigated the effect of membership to a lesion group on performance. Lesion group was divided into three levels: the control group, those with bilateral parietal lobe lesions, and those with left frontotemporal lesions. These one-way ANOVAs were performed on all syllogisms- across congruence for content-imbued syllogisms. The accuracies of subjects did not differ significantly due to lesion group, $F(2, 69) = 0.59$, $p = .56$. The effect of group membership on reaction time was marginally significant, $F(2, 69) = 2.67$, $p = .077$, $\eta p^2 = .07$. This effect showed a trend of increased reaction time for the frontotemporal lesion group ($M = 24904$, $SE = 2033$) compared to the control group ($M = 19658$, $SE = 1206$).

The results of the initial task by group analyses are graphically summarized by figures 10 and 11 using untransformed values. Other known factors may obscure differences, such as variation in a syllogism's representational complexity, and the effects of congruence (or incongruence) between belief and a syllogism's deductive

validity. Therefore, despite limited results at a coarse level of analysis, differences due to lesion group are probed further.

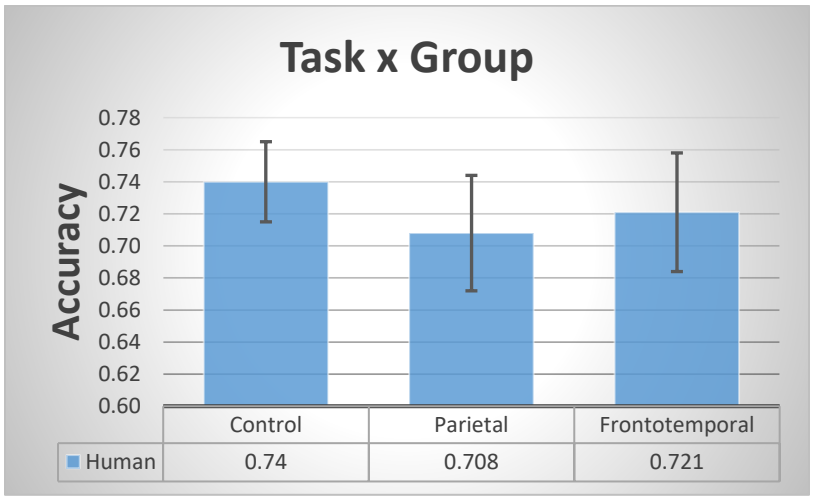


Figure 10

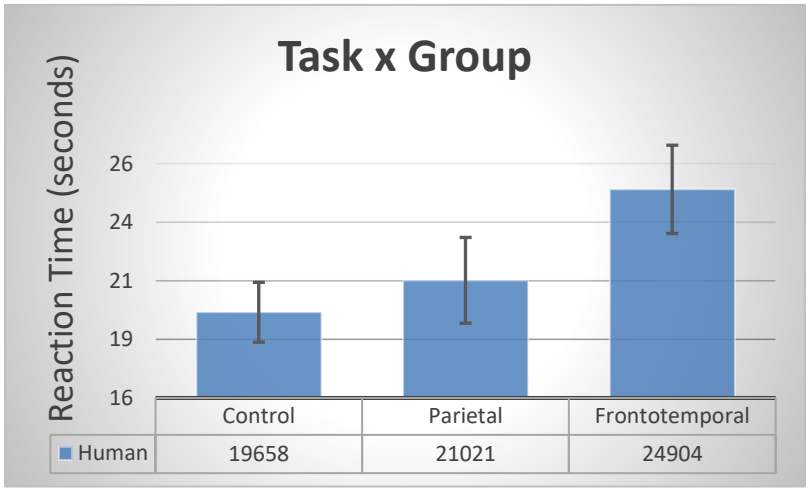


Figure 11

Two-factor 2x3 repeated measures ANOVAs were conducted; this consisted of a between-subjects IV of lesion group with the same three levels, and a within-subjects factor of model complexity. Model complexity consisted of two levels, single or multiple modelling, where single model problems can typically be solved from an immediate evaluation of the premises, and multiple model problems typically require representation of a number of alternate situations to properly evaluate a conclusion's deductive validity. Determining whether a problem was a single or multiple model problem was decided by the computational model's representational output for the task syllogisms.

The first two-factor ANOVA, for the dependent variable of accuracy, demonstrated a significant main effect of model complexity on problem accuracy, $F(1, 69) = 45.21$, $p < .001$, $\eta p^2 = .40$. Humans were more accurate with single-model problems ($M = .78$, $SE = .01$), than for multiple-model problems ($M = .63$, $SE = .02$). The second two-factor ANOVA, conducted for the dependent variable of reaction time, also showed a significant main effect of model complexity on reaction time, $F(1, 69) = 8.57$, $p = .005$, $\eta p^2 = .11$. Humans took less time to solve single-model problems ($M = 21072$, $SE = 962$), than they needed to solve multiple-model problems ($M = 23219$, $SE = 1150$).

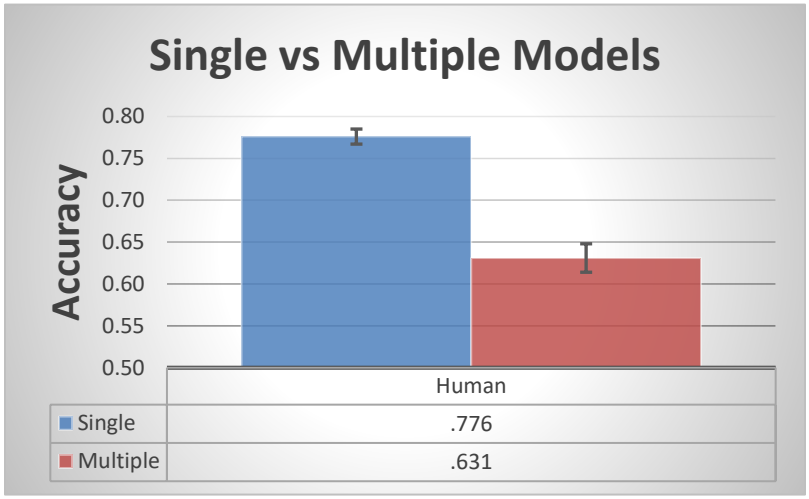


Figure 12

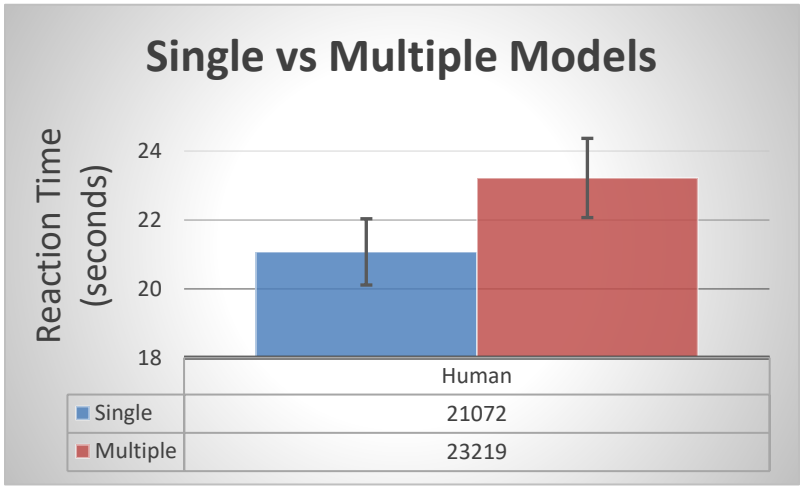


Figure 13

The next set of analyses investigated the relationship between congruency (IV) and lesion group (IV), on accuracy and reaction time (DVs) in two sets of 2x3 two-factor repeated measures ANOVAs. For the two levels of congruency, congruent and incongruent syllogisms, a problem is congruent in one of two cases: if the conclusion is

believable and deductively valid, or the conclusion is unbelievable and deductively invalid. On the contrary, a problem is incongruent when the believability of the conclusion and its deductive validity are in disagreement.

The first 2x3 two-factor repeated measures ANOVA investigates the DV of accuracy. A significant main effect of congruency on problem accuracy was observed, $F(1, 69) = 86.96$, $p < .001$, $\eta p^2 = .56$. Subjects demonstrated higher accuracy levels with congruent syllogisms ($M = .85$, $SE = .02$) than with incongruent syllogisms ($M = .61$, $SE = .03$). A significant interaction between congruency and group on accuracy was observed, $F(2, 69) = 3.23$, $p = .046$. Finding the source of this interaction was difficult, so post-hoc analysis using the LSD test was used to find the group differences most likely to account for the interaction. The frontotemporal group showed the strongest differences for the congruent condition, displaying mean accuracies which trended lower than the parietal lesion group ($p = .18$), and the control group ($p = .094$). The frontotemporal group appears to be impaired on congruent syllogisms ($M = .80$, $SE = .03$) compared to the control group ($M = .87$, $SE = .02$), and the parietal lesion group ($M = .89$, $SE = .03$). For the incongruent condition, the parietal lesion group ($M = .54$, $SE = .05$) showed a trend of impaired performance compared to frontotemporal group ($M = .65$, $SE = .05$, $p = .15$), and the control group ($M = .62$, $SE = .04$, $p = .18$) in terms of accuracy. Figure 14 provides a depiction of this interaction for the human data.

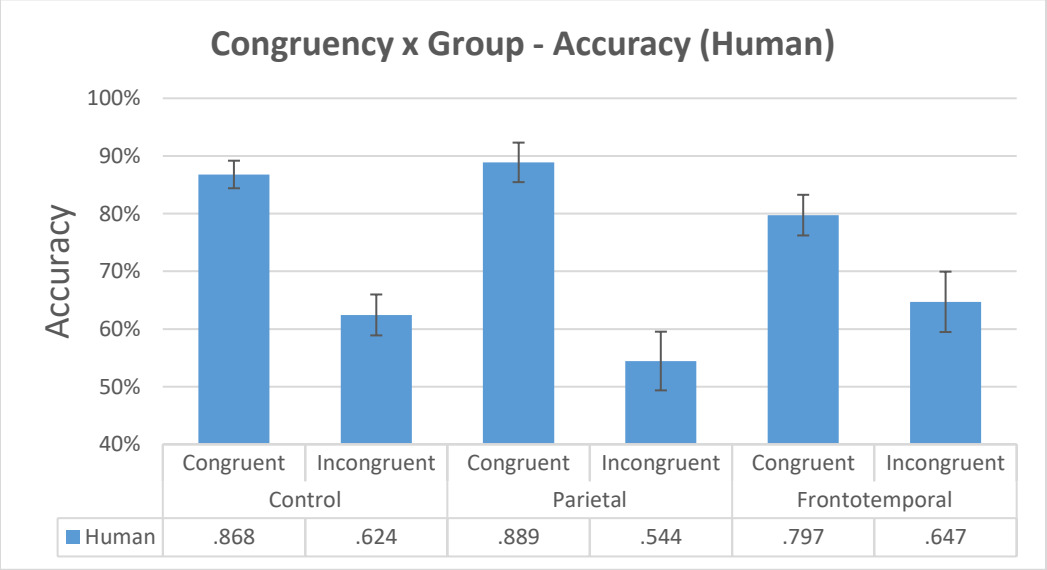


Figure 14

The second two-factor repeated measures ANOVA, utilizing the DV of reaction time, displayed a significant main effect of congruency on reaction time, $F(1, 69) = 19.75$, $p < .001$, $\eta p^2 = .22$. Belief-congruent syllogisms were solved faster ($M = 20115$, $SE = 1004$) than belief-incongruent syllogisms ($M = 23429$, $SE = 1080$). No significant interaction was found between congruency and lesion group for the DV of reaction time, $F(2, 69) = 1.07$, $p = .349$.

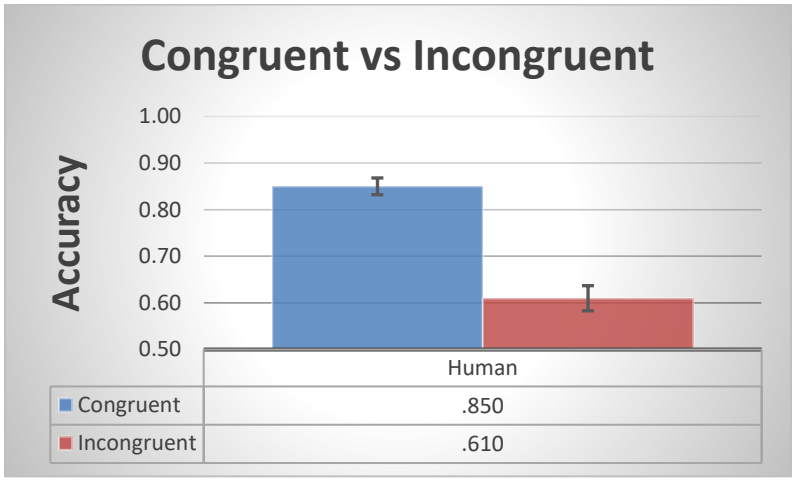


Figure 15

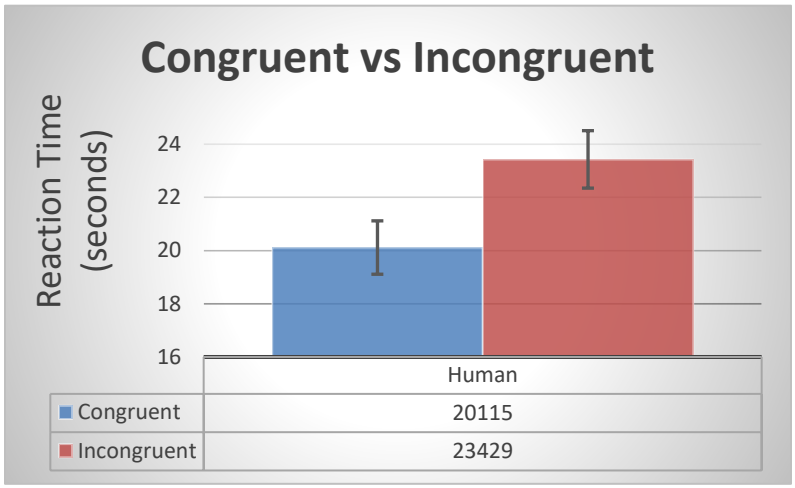


Figure 16

The final set of ANOVAs are defined by a finer breakdown of the congruency factor, which consists of a relationship between conclusion believability and its deductive validity. These 2x2x3 three-factor repeated measures ANOVAs consist of the within-subject IVs of believability (believable/unbelievable), validity (valid/invalid), and the same between-subject IV of lesion group. Unfortunately, separation of groups at such a fine grain resulted in a failure to satisfy the assumption of homogeneity of

variance for reaction time data. For the human data, the invalid-believable group had a significant result for Levene's test ($F(2, 69) = 3.48$, $p = .036$). After using the log10(√x) transformation, this unfortunate significant result disappeared ($F(2, 69) = 1.79$, $p = .17$).

The three-way ANOVA for human data concerning the DV of accuracy showed significant main effects for believability $F(1, 69) = 32.26$, $p < .001$, $\eta p^2 = .32$, and for validity $F(1, 69) = 8.53$, $p = .005$, $\eta p^2 = .11$. Humans were more accurate with unbelievable syllogisms ($M = .80$, $SE = .02$) than believable syllogisms ($M = .68$, $SE = .02$). Humans were also more accurate with valid problems ($M = .78$, $SE = .03$) than they were for invalid problems ($M = .69$, $SE = .02$). A strong interaction between believability and validity was present, $F(2, 69) = 64.88$, $p < .001$, $\eta p^2 = .49$, though as this essentially represents the congruency effect investigated previously, it warrants no further attention.

The identical format ANOVA investigating the DV of reaction time for human data showed a significant main effect of validity on reaction time, $F(1, 69) = 6.39$, $p = .014$, $\eta p^2 = .09$. Deductively valid problems are solved faster ($M = 20388$, $SE = 1072$) than invalid problems ($M = 22756$, $SE = 1072$). A significant interaction between believability and validity on reaction time is observed, $F(1, 69) = 19.42$, $p < .001$, $\eta p^2 = .22$. This known effect (Stupple et al., 2011) is the result of incongruent syllogisms which are believable but invalid taking significantly more time to process than any other type (see Figure 17).

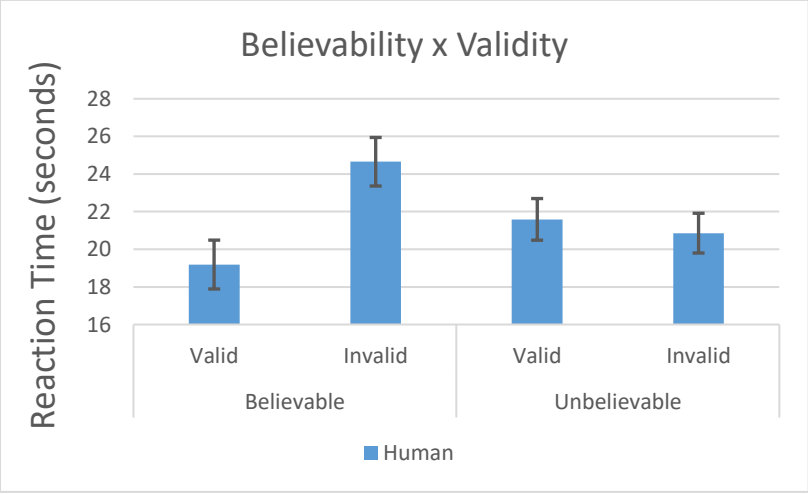


Figure 17

To investigate these reaction time differences further, the human data was probed in a manner similar to the study by Stupple et al. (2011), whom investigated the relationship between believability and validity on reaction time by dividing patients according to high and low logical ability. A similar index was formed by measuring the difference between an individual's acceptance of valid and invalid conclusions. Unlike in the previous study, two groups (high and low logic) were formed rather than three; this was done due to the fact that the present subject pool (71) is much smaller than the 130 used by Stupple et al. (2011). Ten subjects located in the exact center of the logic index were discarded, leaving 32 subjects in the high-logic condition, and 30 subjects in the low-logic condition.

A three-factor 2x2x2 mixed design ANOVA was conducted to examine the relationship between logic-group, believability, and validity. A comparative lack of power failed to produce significant interactions involving group, though the main effect of group was marginally significant ($F(1, 60) = 2.69$, $p = .11$, $\eta p^2 = .04$); the trend showed the high-logic group to take more time solving syllogisms ($M = 22040$, $SE = 1391$), than the low-logic group ($M = 18760$, $SE = 1437$). As particular interest is placed on believable-invalid syllogisms, a simple independent samples t-test was produced to probe this difference, which was found to be significant, $F(60) = 3.71$, $p = .020$, $d = .61$. The high-logic group spent significantly more time solving believable-invalid syllogisms ($M = 26427$, $SE = 1795$), than spent by the low-logic group ($M = 20278$, $SE = 1854$).

3.2 Human Results Discussion

The results of the human data is generally in line with previously discussed neuroimaging data and scientific literature on deductive reasoning. At the broad level of analysis, task by lesion group, the frontotemporal lesion group showed the largest increases in reaction time. This makes sense considering that, by impairing the network supporting belief-based responding, this makes early termination of problem-solving more difficult, and increases the demand for more time-consuming formal procedures. The frontotemporal lesion group also demonstrates some overlapping damage into more general left prefrontal cortex areas thought to support general reasoning abilities (Goel et al., 2006). All subject groups responded as predicted by literature supporting a mental models approach to formal reasoning (Bucciarelli & Johnson-Laird, 1999) such that single model problems were solved faster and with greater accuracy than multiple model problems.

The congruency effect on logical reasoning, well-known in reasoning literature (Evans, Barston, Pollard, 1983), was confirmed across all groups in a broad manner; syllogisms with conclusions whose deductive validity matches the beliefs of the subjects are solved with greater accuracy and speed than those demonstrating incongruence. Further, the results importantly justified the neuroanatomical distinctions drawn by Goel (2009) when accuracy effects were broken down by group. Belief-bias facilitates correct responding in the congruent condition, and the frontotemporal lesion group, employing a network thought to be important to these content effects, showed lower accuracies than control or bilateral parietal patient groups. Incongruent problems are thought to rely on formal reasoning processes, supported by the bilateral parietal network, intervening with belief-biased processes. Human data seems to support this neuroanatomical distinction as the bilateral parietal group displayed lower problem accuracies than control or frontotemporal lesion groups.

The interactions between believability and validity in the human data also conform to reasoning literature. Studies have found humans to be more successful at engaging logical reasoning for syllogisms with unbelievable conclusions than believable ones (Evans et al., 1983; Klauer, Musch, & Naumer, 2000; Stupple & Ball, 2011). An

implication of this is such that believable-invalid problems are particularly difficult, and consume the most time to complete. These theoretical suggestions agree with the results of the present human data. The final interesting result in the human data agrees with research by Stupple and Ball (2011) suggesting that the most logical responders (the best performers) take exceptionally longer time to solve believable-invalid problems than the lower performers who are much more apt to quickly solve the problem through incorrect belief-biased responding.

The Computational Model

4.1 Model design

Prior to discussing the results of the modelling data, it is necessary to explain the computational model in depth; including how it operates and what known cognitive biases it attempts to implement. Performance data from VHIS control subjects dealing with syllogisms that contained meaningful content and content-free forms were used for the initial design and calibration of the computational model. The content condition guided the implementation of the heuristic system's belief-bias effect, while the content-free condition played a greater role in the design and calibration of the effects of other prominent cognitive biases in categorical syllogism literature. Beyond this point, adjustments were made for the implementation of lesions to allow the model's performance to be measured against human controls and patients. This enables one potential framework of a fractionated deductive reasoning system to be empirically tested.

There are five main components in the computational model. The general pattern completer (left PFC) breaks down and distributes the propositions of the categorical syllogism to the heuristic and formal reasoning systems. It receives feedback from various centers and completes the patterns of information in these signals to arrive at a validity judgement. The formal system (bi-lateral parietal network) evaluates the syllogism in accordance with mental model theory. The heuristic system (left frontal/temporal) influences decision-making through the belief-bias effect. The conflict detector monitors these two processing networks for logical conflict and belief-logic conflict. The uncertainty maintenance system attempts to inhibit belief-bias during complicated logical representations, though its ability to do so is impaired under situations of strong belief.

The general pattern completer starts by identifying the syntactic role of the components of the premises. These include quantifiers (all/no/some) copular terms (are/is/have/can/etc.) and the negation 'not'. The original full sentences are broken down and a new WME has its variables filled with these relevant syntactic elements. At this point a check is performed to see if a premise conversion error is performed (based

upon elements to be discussed later) – involving a switch of the two terms in a particular premise (or conclusion). Following this, the pattern completer simultaneously passes the syllogism on to the formal and heuristic systems for evaluation. The pattern completer will receive feedback from the other parts of the deductive reasoning system as it attempts to reach a decision about the validity of the conclusion. To arrive at a decision, the activation level of two WMEs, representing a valid or invalid choice, will increase until it exceeds some critical firing rate threshold in a first-past-the-post decision-making paradigm.

The heuristic system primarily provides a mechanism for belief bias, which has shown a highly robust effect on the processing of syllogisms (Evans, Barston, & Pollard, 1983). For this effect, performance on syllogisms is improved when the conclusion is consistent with the beliefs of the subject compared to when it is inconsistent. The conclusion ‘all men are smokers’ is an example of a conclusion incongruent with human beliefs, while the conclusion ‘some mushrooms are not poisonous’ would be congruent with human beliefs. The computational model introduces this effect by increasing the pattern completer’s validity judgement (valid/invalid) WME activation depending upon the belief held and the strength to which it is held. As previous research indicates believable conclusions provoke stronger levels of belief-bias than for unbelievable conclusions (Evans et al., 1983), believable conclusions have a stronger influence over the pattern completer’s valid signal than unbelievable conclusions do over the invalid signal.

The rates with which the model believes a particular conclusion to be true or false are set by conclusion endorsement rates gathered from the controls in the Vietnam Head Injury study. The strength and direction of the belief is randomly generated from the proportion of responses indicating a particular belief rating; if 10% of responses indicated strong disbelief in the conclusion (rating of 1), then there is a 10% chance the belief generated will be strong disbelief. A valid belief increases the pattern completer’s decision-valid WME activation level, and an invalid one increases the decision-invalid activation level, where a stronger belief (or disbelief) will introduce a stronger activation boost. This change also depends upon the health of the heuristic

system, for if it is lesioned it will have less of an influence over the pattern completer's decision signals.

The formal system is the most computationally intense aspect of the program. It generates evaluations of categorical syllogisms in accordance with mental model theory. Mental model theory (Johnson-Laird, 1983) suggests the relationships between the terms of categorical syllogisms are represented by a finite set of mental tokens. These tokens are arranged to represent the two premises and integrated to form an initial model of the situation. Reasoners will attempt to derive conclusions from this model, following this, they may or may not perform a number of manipulations to the model to search for counterexamples and refute these conclusions. The data used for this thesis involves syllogisms where the subject is provided conclusions, rather than asked to generate their own from the premises as is done in the computational model explained by Bucciarelli and Johnson-Laird (1999). While this required some departures in the evaluation of models and the search for alternate models, the overall process is highly similar. The original mental model program would not generate negative conclusion for premises lacking negations; therefore if a negative conclusion is provided for such a set of premises, the program is forced to proceed with operations it would use as if negative tokens were present in order to falsify such a conclusion.

The four possible syllogistic moods are represented by four mental models (see Figure 18). Square brackets are used to show that the token has been exhaustively represented, as is the case with the two universal moods, where we have a full account of this token so that no more instances of it will be added to the model in the search for alternatives. Before we can combine premises the models must be arranged so that the middle terms line up. As mentioned previously, there are four different figures for categorical syllogisms which represent the different orders in which the terms may occur. For the first figure (see Figure 19) the middle term (B) is already in the middle, so no change is necessary. For the second figure the order of the premises is switched. The third and fourth figure require that the terms are swapped in the second premise and in the first premise, respectively before integration can occur. An inspection-time analysis (Espino, Santamaria, & Garcia-Madruga, 2000) indicates that this additional

effort, for figures other than the first, slightly increases the time it takes to integrate the premises, but has no effect on the accuracy of subjects in solving syllogisms.

Figure 18

[X]	Y	[X]	-Y	X	Y	X	-Y
[X]	Y	[X]	-Y	X		X	-Y
	Y		Y		Y		Y
			Y				Y
All X are Y		No X are Y		Some X are Y		Some X are not Y	
universal		*universal*		*particular*		*particular*	
affirmative		*negative*		*affirmative*		*negative*	

Figure 19

First Figure		Second Figure		Third Figure		Fourth Figure	
A	B	B	A	A	B	B	A
B	C	C	B	C	B	B	C

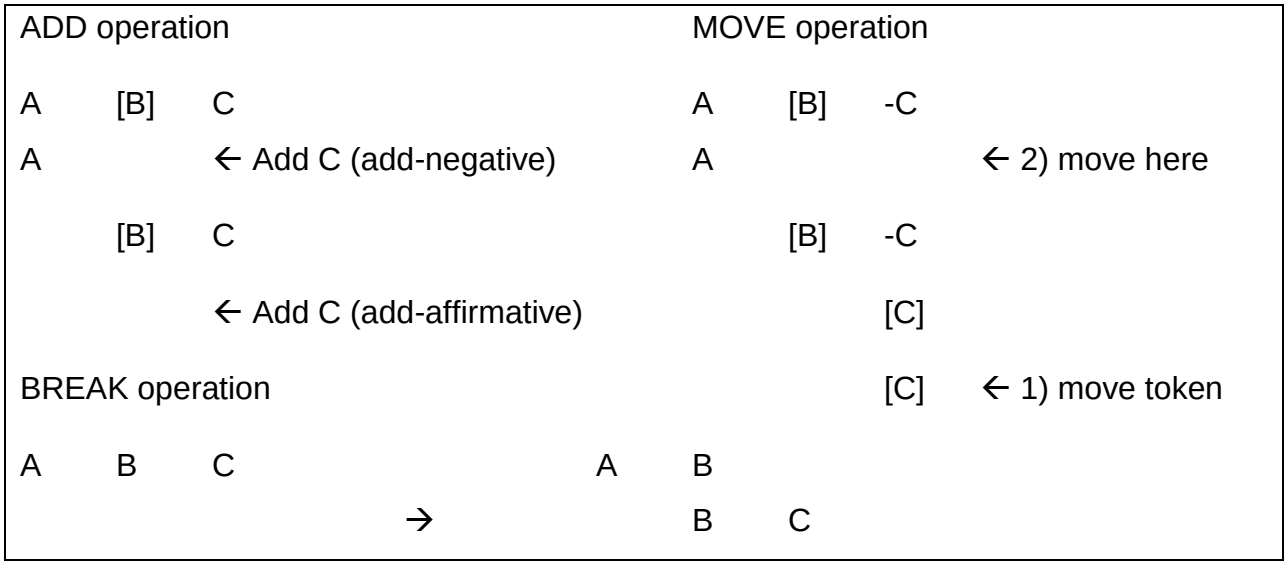
To unite the two models, we find the middle terms for both models and move the end-term attached to the middle term in the second model so that it is now beside the middle term in the first model. If during this move the middle term is exhausted in either model it becomes exhausted in the integrated model. Once all of these moves are completed any remaining free tokens are appended to the end of the integrated model (see Figure 20).

Figure 20

Premise 1		Premise 2					
Some A are B		All B are C			Integrated model		
A	B	[B]	C		A	[B]	C
A		[B]	C		A		
	B					[B]	C

To evaluate this integrated model the end tokens are read bi-directionally, from left to right and from right to left, to see if a particular conclusion holds true across all lines. If the conclusion to be tested here was that ‘Some A are not C’ this would appear to be true as in the second line we have an A but no C. To see if that can be falsified additional models are created. For positive models (ones lacking negatives no/not in the premises or conclusion) model creation functions add-affirmative, move, and break may be executed. For add-affirmative, new lines can be created containing end tokens that are not exhaustively represented. Move will move ‘free end-tokens’ into empty spaces. Break will divides a line with unexhausted middle terms into two lines. For negative models (which have a negation in a premise or conclusion) also utilize the move and break operations, though their method for adding tokens is different, here non-exhausted tokens are added to empty spaces to falsify conclusions (see Figure 21). These new models will be tested and the process repeated until an invalid case is found, or the general pattern completer arrives at a validity judgement. Figure 22 provides a flow diagram of the mental model procedure, where dashed-line arrows to output represent stages of tentative output to the general pattern completer regarding the likely validity of the tested conclusion.

Figure 21



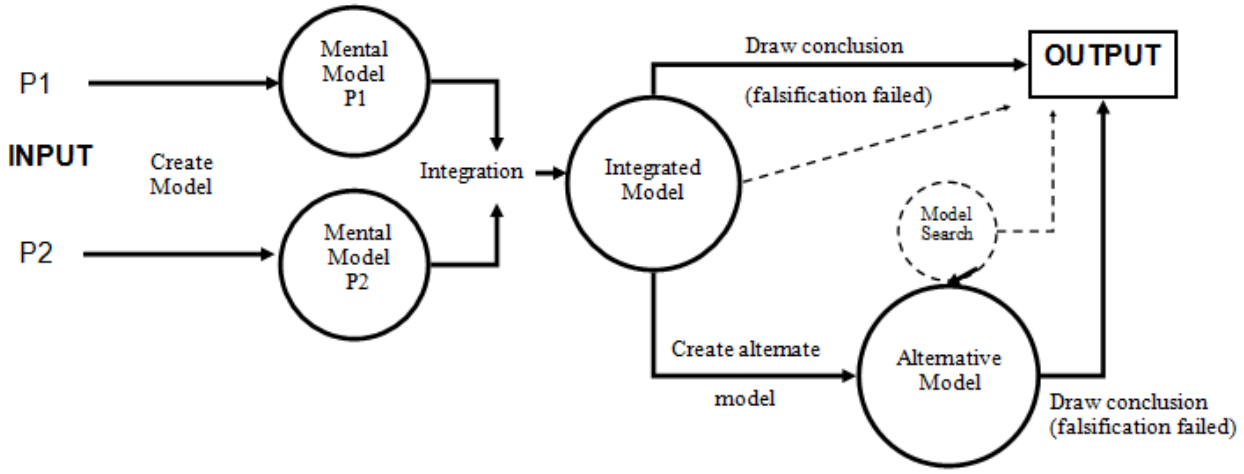


Figure 22

The deductive validity of a single-model problem will be immediately obvious after the integration of the information of both premises. For multiple model problems, the appropriate deductive decision is ambiguous at this step, and requires the formulation of additional potential scenarios to determine the validity of the conclusion. If we use a previous example syllogism,

		Premise 1		Premise 2	
Premise 1 All birds can fly	All B are A				
Premise 2 All pigeons are birds	All C are B	[B]	A	[C]	B
Conclusion: All pigeons can fly	All C are A	[B]	A	[C]	B

upon integrating the models of these two premises we arrive the following model:

[C] [B] A
[C] [B] A

When we evaluate the conclusion “All C are A” we see C to be exhaustively represented, and it always has an A on the other side of its line, confirming this conclusion to be deductively valid. If we instead turn to a multiple-model syllogism,

		Premise 1		Premise 2	
Premise 1 No coffee contains nicotine	No A are B	[A]	-B	[B]	-C
Premise 2 No nicotine contains tea	No B are C	[A]	-B	[B]	-C
Conclusion: No tea contains coffee	No C are A		B		C
			B		C

then this is the resultant integrated model:

[A] -B

[A] -B

[B] -C

[B] -C

C

C

If we try to observe whether “No C are A” is valid, we are unsure, as the only instances of C have no token at the other end after integration. It is a possible case that no C come with A due to this blank space, but we are not sure if there are other possible cases where this is false. To be deductively valid, there must be no cases where C can be followed by A. To confirm or disconfirm this conclusion we must tweak this model. By moving the C token up into the blank spaces above it (see below), we create an alternate scenario where “No C are A” is false, demonstrating this conclusion to be deductively invalid. For a multiple model problem, the initial integrated model may suggest a conclusion to be valid, but there remains uncertainty regarding other possibilities. This is why these problems should take longer, and are more difficult to accurately solve.

[A] -B C

[A] -B C

[B] -C

[B] -C

The formal system will provide its first indications to the pattern completer concerning potential judgments at the conclusion evaluation stage. A seemingly valid conclusion will result in a mild boost of decision-valid activation levels. An invalid conclusion will impose a strong change in activation, though it is often not enough on its own to completely and immediately determine the pattern completer’s decision. If the conflict detection system is not heavily taxed this invalid signal will be more powerful, and it will be repeated until the pattern completer arrives at a final decision and signals other systems to cease operations. A valid conclusion will necessitate the construction

of additional models (if possible) and a continuation of this process until a final validity decision is reached.

The conflict detection system consumes activation when the formal reasoning system is checking for logical inconsistencies. This system 'works harder' and consumes more activation depending upon the number of lines the current model takes to represent. Universal affirmative moods (All X are Y) are the easiest to represent, while models containing negative premises are more complicated – they typically require more lines in their representation. Thus, different combinations of premises induce different levels of strain on conflict detection; and having to monitor for consistency with belief adds even more burden. The greater the burden on the system, the weaker the effect the formal system has in identifying an invalid conclusion to the general pattern completer. When an invalid conclusion is not picked up, the formal system may continue to attempt to derive additional models after an invalid case has been generated. These new derivations may either change the logical interpretation of the syllogism, or they may simply delay finding the correct solution long enough for belief-bias to induce the wrong decision for incongruent problems. Higher levels of strain on the conflict detection system also decreases the inhibition this system is able to apply to the heuristic system to reduce belief-bias effects when there is a conflict between logic and belief.

The uncertainty maintenance system alerts the general pattern completer of indeterminate or ambiguous situations. It activates or refreshes its activation during the creation of alternate models, which are taken to represent more ambiguous situations. In these cases, the uncertainty maintenance system will also inhibit the belief bias response. When belief levels are at their strongest (5 or 1) it is hypothesized in the current model that there is less uncertainty in the situation due to these particularly potent beliefs. To represent this, the heuristic system will also have inhibitory connections onto the uncertainty maintenance system. Its inhibition will be stronger when the beliefs are stronger, and weaker if the heuristic system is lesioned.

5.2 Cognitive Biases

Beyond the functional division (and interaction) of processing among the 5 brain-based modules, the computational model also incorporates functions for introducing a number of well-known cognitive biases in the literature concerning the categorical syllogism. These biases include the atmosphere effect, matching bias, and premise conversion errors. The effects of these biases in the system were primarily configured using the content-free data, as without the belief-bias effect from content these additional biases remain as the primary sources of error. The biases take their effect by influencing the general pattern completer's judgement decision, or by transforming the implications of the premises prior to being passed on from the pattern completer.

The atmosphere effect (Woodworth & Sells, 1935) is an example of an inductive or probabilistic reasoning process where the presence of negative items – as found in the propositions '*No* X are Y' or 'Some X are *not* Y' – or the presence of the particular quantifier *some* in the premises of a syllogism have implications for the likelihood of a conclusion being valid. Associative heuristics may be employed where if at least one premise is negative the conclusion is more likely to be negative, and if at least one premise is particular the conclusion is more likely to be particular.

If one of these matches exist between the moods of the premises and conclusion, the model acts on this by having the general pattern completer raise the activation level of the conclusion-valid WME. If there is a negative or particular premise (or both) and the conclusion does not match, the activation of the conclusion-invalid WME increases. The increase is even greater for valid (or invalid) WMEs if both conditions are (un-)satisfied. Another logical extension of this relationship is that if neither premise is negative the conclusion should be affirmative ('All X are Y' or 'Some X are Y'), and if neither premise is particular the conclusion should be universal ('All X are Y' or 'No X are Y'). Bucciarelli and Laird (1999) have an alternative explanation of the atmosphere effect using mental model theory. It is stated that conclusions derived from an initial model of the premises will also match the mood of at least one premise, similar to a superficial matching of verbal forms. In an effort to appeal to both

interpretations, judgements concerning the validity of a conclusion as derived from the initial model will have stronger weight than those of subsequent models.

The matching bias is similar to the atmosphere effect as it involves heuristics operating on the quantifiers of the syllogism. Wetherick (1989) suggests that when the validity of a situation is not immediately obvious (suggesting its deployment in multiple model problems) additional heuristics may provide estimates of validity. When all premises are the same it suggests conclusions identical to the atmosphere effect. When they are different, the matching bias is said to prefer conclusions that match the more conservative premise (Wetherick & Golhooly, 1990). When one premise contains the quantifier ‘All’, matching bias selects the form of the other premise. When one premise is of the form ‘Some X are Y’, and the other is of the form ‘No X are Y’, matching prefers a ‘No X are Y’ conclusion. If a match is found, the general pattern completer’s conclusion-valid WME activation increases.

The remaining cognitive biases involve a misrepresentation of the information provided by the premises prior to creating a formal model of the situation. Faulty logical implications (Rips, 1994) arise from a common-sense (mis-)understanding of the premises based upon communicative norms. The premise ‘Some X are Y’ is taken to imply that there are also some X which are not Y, because if this was not the case it is assumed the stronger premise ‘All X are Y’ would be provided instead. Similarly, ‘Some X are not Y’ is taken to imply there are also some X which are Y, as otherwise the premise ‘No X are Y’ would be present. Inferring more information from premises than they actually present, in accordance with logical norms, leads reasoners to build faulty initial models of situations that lend themselves to faulty conclusions.

Faulty implications are said to be, in part, based upon communicative norms because a speaker generally would not provide the weaker relationship between two terms if a stronger one were true. However, the premises of syllogisms present relationship-knowledge assumed to be true rather than what might be actually true. In certain cases where a faulty implication makes strong intuitive sense, the likelihood of these conversions increase. For example, if one premise states that ‘some Olympic athletes are smokers’, the likelihood of this also implying ‘some Olympic athletes are not

smokers' is even greater than when presented in an abstract token form. Information on the believability of premises was not collected, let alone information concerning the believability of alternate forms of premises, so this is an assumption of the model estimated to improve its fit with data by increasing the potential for error. Implementing this change within mental models is relatively easy, as it basically involves adding positive or negative tokens to the premise models depending on the additional implication applied.

The final bias to be implemented are conversion errors (Chapman & Chapman, 1959). In this type of error, the terms of a premise are switched, causing the reasoner to incorrectly infer the inverse of a proposition. All X are Y is taken to imply All Y are X, and Some X are not Y is taken to mean Some Y are not X. In the former situation, we do not have any information about what logically follows given Y; we only know what is true when given X. In the latter situation, while Some X are not Y, it is possible that in all the cases where X leads to Y that these are all the instances where Y occurs; this means that the inverse (Some Y are not X) is incorrect as instead 'All Y are X' is true. Conversion errors are implemented relatively easily by the pattern completer sometimes accidentally switching the terms before passing the information elsewhere. As with the faulty implications bias, the rate of this occurring increases if the inverse of a proposition is more intuitively appealing in content conditions, which again is an assumption of the model.

To generate data for statistical comparison of model performance against human performance, sets of thirty simulated subjects were generated for the various lesion groups. This number was chosen so that it would not be so high as to make every result significant by artificially inflating the degrees of freedom. The groups were large enough so that these equal samples would be more robust against violations of heterogeneity of variance. Lesions were applied to the formal and heuristic systems by approximately halving the activation capacity of these centers. Accuracy and reaction time data was organized for simulated subjects under a number of different categories. These categories included distinctions between single and multiple models, congruent and incongruent problems, and all combinations of believability and validity: believable-valid, believable-invalid, unbelievable-valid, and unbelievable-invalid.

5.3 Model Data Results

Two separate one-way ANOVAs were conducted to examine the basic relationship between lesion group (IV) and the dependent variables of reaction time and accuracy on model performance. There was no significant main effect of lesion group on accuracy, $F(2, 87) = 2.10$, $p = .129$. The effect of lesion group on reaction time for the computational model was significant, $F(2, 87) = 29.59$, $p < .001$, $\eta p^2 = .41$. Post-hoc analysis using the Sidak correction showed the simulated control group to perform faster than other lesion groups (all $p < .001$).

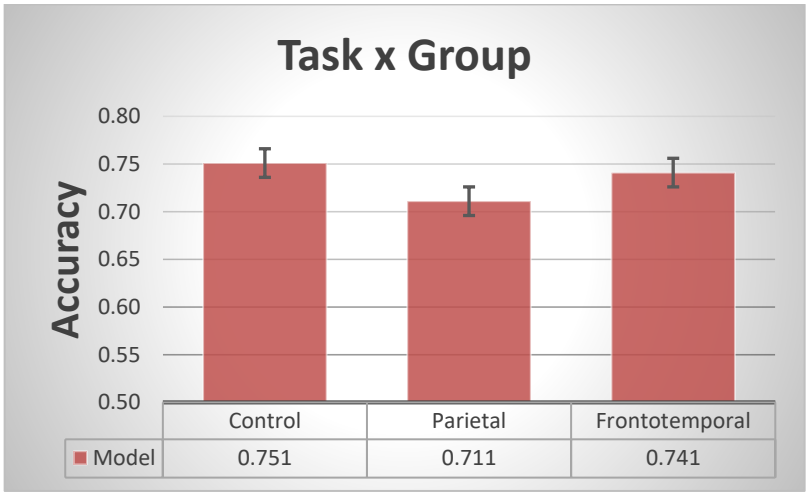


Figure 23

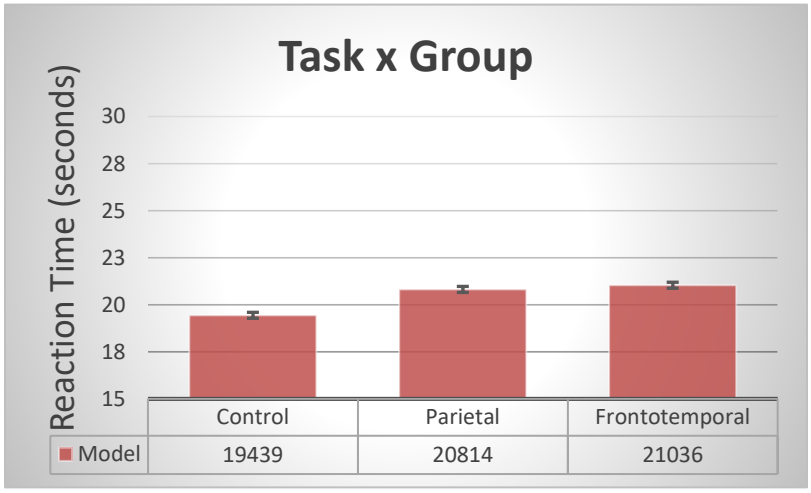


Figure 24

Two-factor 2x3 repeated measures ANOVAs were conducted on model data for accuracy and reaction time DVs; this consisted of a between-subjects IV of lesion group with the same three levels, and a within-subjects factor of model complexity (single or multiple models). Accuracy significantly differed depending upon model complexity, $F(1, 87) = 196.84$, $p < .001$, $\eta p^2 = .69$; single-model problems generated higher accuracy

scores (*M* = .83, *SE* = .009), than multiple-model problems (*M* = .56, *SE* = .016). Reaction time significantly differed depending upon model complexity, $F(1, 87) = 67.81$, $p < .001$, $\eta p^2 = .44$; single-model problems took less time to solve, (*M* = 19455, *SE* = 67.96) than multiple-model problems (*M* = 21349, *SE* = 1150.49).

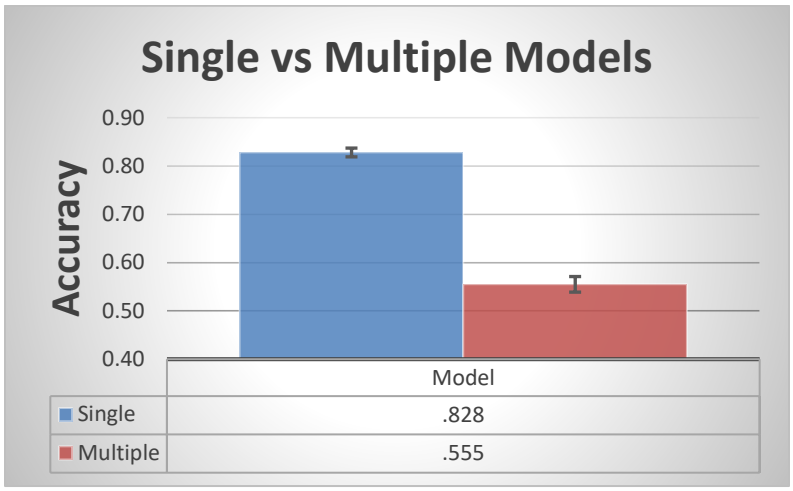


Figure 25

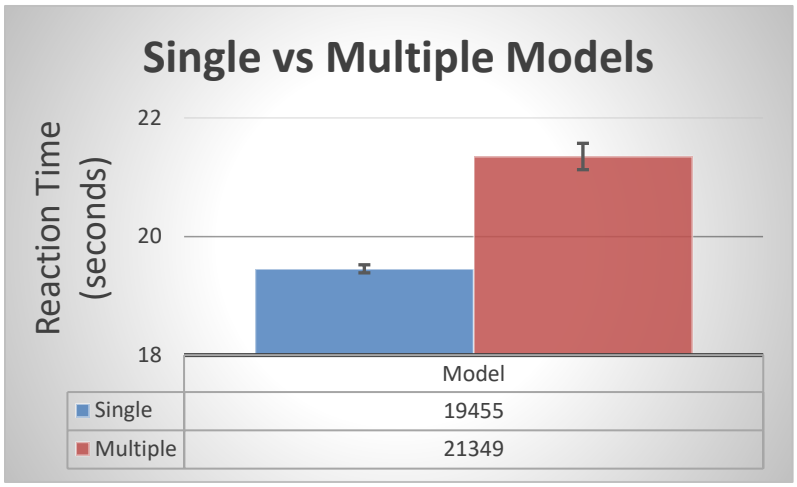


Figure 26

The next set of analyses investigated the relationship between congruency (IV) and lesion group (IV), on accuracy and reaction time (DVs) in two sets of 2x3 two-factor repeated measures ANOVAs. The model data investigation of accuracy demonstrated a significant effect of congruency on problem accuracy, $F(1, 87) = 343.08$, $p < .001$, $\eta p^2 = .80$. Higher accuracy levels are generated for congruent syllogisms (*M* = .90, *SE* = .01), than for incongruent syllogisms (*M* = .55, *SE* = .01). A significant interaction between congruency and group on problem accuracy is again observed, ($F(2, 87) = 10.94$, $p < .001$, $\eta p^2 = .20$). A stronger interaction effect and increased power allowed for tighter control over error rate than the human data when performing a post-hoc

analysis. The conservative Bonferroni correction showed the frontotemporal group to be significantly impaired (p = .01) on congruent syllogisms (M = .87, SE = .02), compared to the parietal lesion group (M = .93, SE = .02). The same correction applied to the incongruent condition showed the parietal group (M = .46, SE = .03) to be impaired compared to the frontotemporal group (M = .60, SE = .03, p = .002), and the control group (M = .60, SE = .03, p = .001). Figure 27 provides a visual representation of this interaction for the computational model.

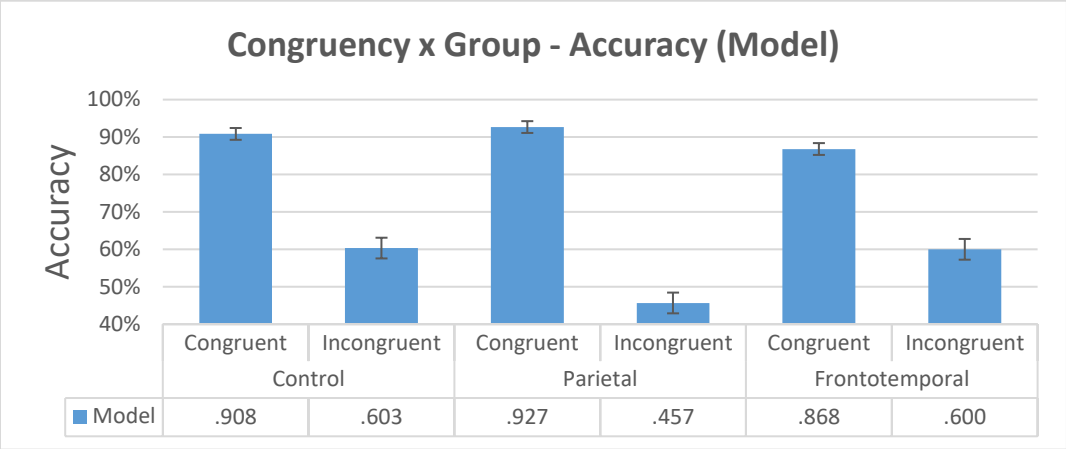


Figure 27

A two-factor repeated measures ANOVA investigating the DV of reaction time shows a significant main effect of congruency on reaction time, $F(1, 87) = 60.46$, $p < .001$, $\eta p^2 = .41$. Congruent syllogisms are solved significantly faster (M = 19484, SE = 132.67), than incongruent syllogisms (M =20955, SE = 125.98). A significant interaction between congruency and group on reaction time ($F(2, 87) = 4.64$, p = .012, ηp^2 = .10) is found only for the computational model. This result reflects the reaction time increase, moving from the congruent to the incongruent condition, being sharper for the parietal lesion group than other groups. Figures 28 and 29 summaries the general effects of congruency for this section of the data analysis.

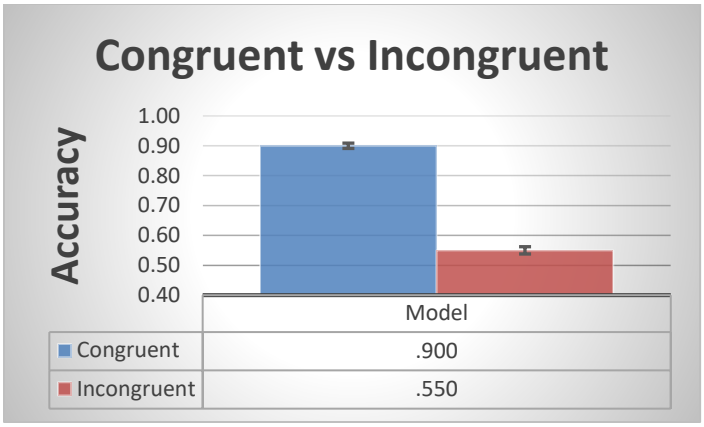


Figure 28

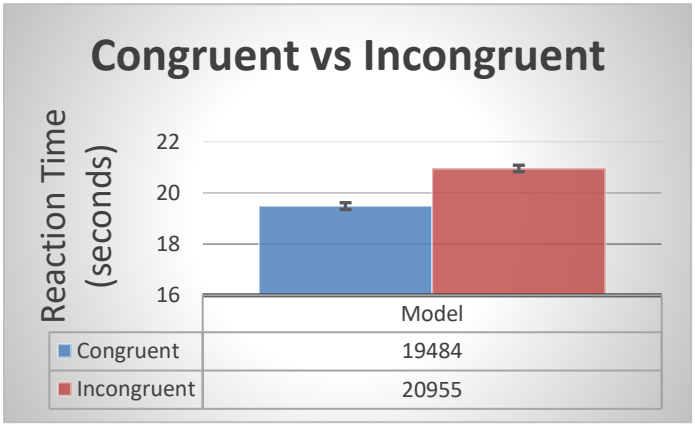


Figure 29

The three-way ANOVA investigating the relationship between believability, validity, and group on the DV of accuracy for the model showed a significant main effect of believability on accuracy, $F(1, 87) = 94.66$, $p < .001$, $\eta p^2 = .52$. Simulated reasoners are more accurate with unbelievable problems ($M = 0.84$, $SE = 0.01$) than they are with believable problems ($M = 0.67$, $SE = 0.01$). Unlike for the human data, the main effect of validity was not significant: $F(1, 87) = 0.40$, $p = .53$, $\eta p^2 = .01$. A significant interaction between believability and validity on problem accuracy, again representing the congruency effect, was found: $F(1, 87) = 239.82$, $p < .001$, $\eta p^2 = .73$.

Prior to analyzing the reaction time results, the RT values for the model data were transformed using the log10(√x) transformation. This is in part motivated by the fact that the model violated Levene's test for homogeneity of variance for the invalid-unbelievable grouping, $F(2, 87) = 6.19$, $p = .003$. While this transformation may not typically be necessary as the model data is more robust to violations of homogeneity, due to having a larger sample with equal numbers of subjects in each group, because

the human data for this variable was transformed it seemed appropriate to do this for the model data as well.

The three-way ANOVA analysing the DV of reaction time for the model data shows differences in its effects compared to the significant effects of the human data. While no significant reaction time difference was found between problems with believable and unbelievable conclusions in the human data (p = .97), for the computational model unbelievable problems take slightly longer (M = 21116, SE = 124.27) than believable ones (M = 19836, SE = 130.30), demonstrating an F ratio of $F(1, 87) = 45.54$, $p < .001$, $\eta p^2 = .34$. Furthermore, while invalid problems took significantly longer for the human data (p = .003), valid problems showed slightly increased reaction times (M = 20924, SE = 138.41) compared to invalid problems (M =20028, SE = 100.65). The difference in means is extremely small, and only finds significance ($F(1, 87) = 18.94$, $p < .001$, $\eta p^2 = .179$) in the fact that standard errors for model reaction time are exceptionally low. The differences between human and model data on reaction time, for these factors, are largely captured by the significant interaction between believability and validity on reaction time. While the significant difference for the human data ($p < .001$) was accounted for by believable-invalid problems taking the longest amount of time, the significant interaction for model data ($F(1, 87) = 120.15$, $p < .001$, $\eta p^2 = .58$) is the result of believable-invalid incongruent syllogisms taking the most time. This systemic difference in reaction time is of interest for future model building efforts.

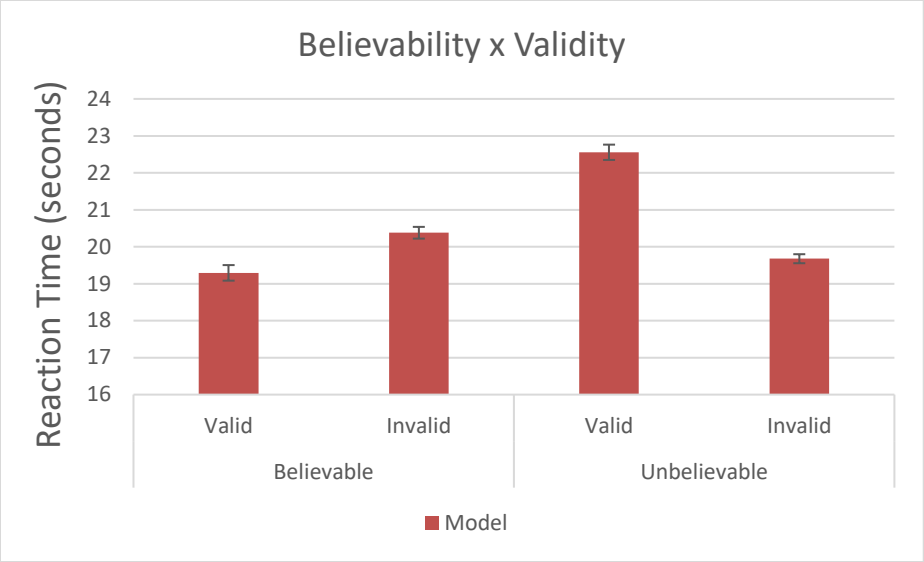


Figure 30

To directly compare the performance of the computational model to that of the human subjects, the Pearson correlation is used to compare accuracy and reaction time values across the nineteen syllogisms utilized (see Appendix C for a list of the syllogisms). Highly significant values were obtained for all correlations of accuracy comparing human and model data: the control groups ($r(19) = .88$, $p < .001$), the frontotemporal group ($r(19) = .72$, $p < .001$), and bilateral parietal group ($r(19) = .75$, $p < .001$) were all significantly correlated. For reaction time, due to the wide variance present in the human data not being reflected in the model data the correlations were much weaker. This is particularly true for the control group $r(19) = .01$, $p = .98$, even though their overall means were highly similar (for humans $M = 19516$, $SD = 3927$; for the model M = 19433, $SD = 1751$). The comparisons of the parietal groups were more promising ($r(19) = .31$, $p = .19$), and the correlations between the frontotemporal groups was significant: $r(19) = .41$, $p = .08$. Graphs depicting the fit for accuracy and reaction time for each syllogism used are provided in Appendix D.

5.4 Model Results Discussion

Results of the modelling data largely agree with human subject data, where the departures that are observed have reasonable explanations. The coarse task by group analysis showed no significant accuracy differences for either the human or the model data depending upon lesion group. In terms of reaction time, the human data displayed a trend of increase (p = .077) in reaction time for the frontotemporal group compared to the control group. For the model data, the control group performed significantly faster than both of the lesion groups. The frontotemporal group for the human data may be impacted slightly more in its processing time as it is noted the left prefrontal cortex, correlated with general reasoning ability, possesses some overlapping lesions for which no equivalent general impairment is employed in the computational model. On top of this, the computational model employs lesions by halving the processing capacity of brain areas, which is possibly more dramatic than the human impairments, though it is necessary to provide meaningful differences in a more deterministic and less variable computational system.

Significant reduction in accuracy and reaction time for the computational model when dealing with multiple models as opposed to single models is replicates the effects found in human data. The main effects of congruency on these DVs are also the same; where congruent problems are easier and solved faster than incongruent problems. The interaction between lesion group and congruency for the dependent variable of accuracy also agrees with the human data; frontotemporal lesions impair the model on congruent syllogisms and bilateral parietal lesions impair the model on incongruent syllogisms.

Aside from the main effects, congruency by lesion group interactions were insignificant for the human reaction time data. The model displayed one additional result where the bilateral parietal lesion group was more sharply impaired in terms of reaction time than was seen in the human data. This is explainable for much the same reasons as the differences in reaction time between humans and the model in the coarse task by group analysis; namely, a more strongly impaired deterministic system with less variability (low standard error) compared to human performance.

For the last set of analyses, breaking down congruency into an examination of interactions between believability and validity, results are again highly similar (with some systematic differences). Both the model and human data showed unbelievable problems to be solved more successfully than believable problems. The key difference between human and model data lies in the interaction between believability and validity for reaction time data. For human data, believable-invalid problems consumed the most time, while for the model data unbelievable-valid problems took the longest to solve. Stupple et al. (2011) highlight the fact that the most logical reasoners spend large amounts of time on believable-valid problems. The computational model did not employ any differences in cognitive style as it utilized a single operation profile that was adjusted by introducing lesions. Future modelling efforts could attempt to include simulated high-logic individuals who respond differently to believable-invalid problems, and therefore bring results more in line with human data.

Discussion

6.1 Results Discussion

We now summarize the most pertinent results of the project. The most direct evidence supporting a fractionated deductive reasoning system, and the only piece which ties a kind of functioning (heuristic or formal) to a specific brain area, comes from the significant interaction (p = .046) of congruency and group in terms of accuracy in solving categorical syllogisms for human data. Those with frontotemporal lesions, an area associated with belief-biased responding, were impaired on congruent syllogisms where belief-bias would have more easily led reasoners to the correct conclusion. In congruent syllogisms, belief-bias improves subject accuracy as past knowledge provides an intuitive bias leading them towards a judgement that coincidentally agrees with the formal validity of the syllogism. On more challenging multiple-model problems, or problems more prone to errors of conversion (misrepresentation of the premises), where the formal system would have more difficulty arriving at the correct conclusion, the heuristic system can influence the judgment process towards the correct solution. With this benefit removed, frontotemporal patients showed impairment in accuracy on congruent syllogisms.

Patients with bilateral parietal lesions, associated with formal reasoning through manipulation of spatial representations (Goel et al., 2000), were impaired on incongruent syllogisms where more deliberate and formal reasoning processes could have been used to oppose – pre-potent and incorrect - belief-biased responding patterns. It is implied that those with damage to the bilateral parietal areas are less capable of applying formal reasoning through methods, like mental models theory, to solve categorical syllogisms. This congruency effect on reasoning was successfully replicated in the computational model using these separate but interacting formal and heuristic mechanisms; this replication was done in a way such that the accuracy of responses generated was highly correlated to the human data.

Incongruent syllogisms were more difficult and took longer to solve than congruent syllogisms for both model and human data (all ps < .001) demonstrating that syllogistic processing is not entirely driven by beliefs. Multiple-model syllogisms were

also more difficult and took longer to solve than single-model syllogisms. These results provides necessary, though not sufficient, evidence for formal representational reasoning processes in human brains. Syllogisms with unbelievable conclusions also more frequently evoked these formal reasoning processes by demonstrating significantly increased mean accuracies in model and human data than is seen for syllogisms with believable conclusions.

The interacting effects of belief and validity on accuracy seen in the human and model data is in compliance with the frequently cited results of Evans, Barston, and Pollard (1983) and other studies replicating these results (Klauer, Musch, & Naumer, 2000; Stupple & Ball, 2008). Evans et al. (1983) observed the lowest accuracy (29%) for invalid-believable incongruent syllogisms, moderate accuracy (56%) for valid-unbelievable incongruent problems, high accuracy (89%) for valid-believable congruent problems, and the highest accuracy (90%) for invalid-unbelievable congruent problems. This pattern is reflected in the human and model data for this set of syllogisms, and the values found for each are highly similar (see Figure 35).

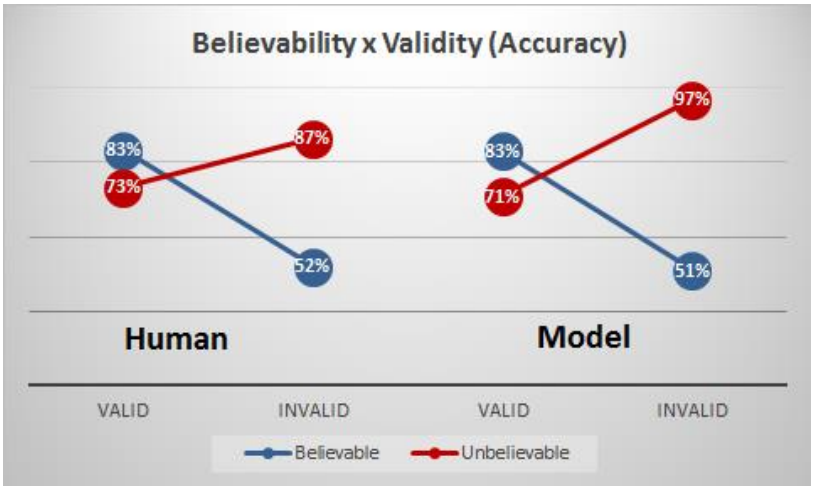


Figure 31

Model data which is in agreement with human data and past research is very appealing, though one must speak to the divergences in performance as well. If these divergences cannot be explained, it brings into question the viability of a fractionated deductive reasoning system, arranged in this way, for being a potential model of true human performance. The main divergence between the model and human data lies in the different interpretations of the significant believability by validity interaction for

reaction time. For the model data, unbelievable-valid incongruent problems were the slowest, while believable-invalid incongruent problems were slowest for human data, which agrees with typical findings in past research (Stupple et al., 2011). Stupple et al. suggest a sub-group of individuals more frequently understand the underlying logic of these problems, and attempt to resist fallacious conclusions through formal reasoning. This requires processing effort above and beyond that of other problems in order to resist the belief-bias effect – found to be stronger for syllogisms with believable conclusions. These high-logic individuals typically take longer to solve syllogisms overall, but this is especially true for believable-invalid problems.

Attempting to repeat the results of Stupple et al. (2011) showed only marginally success at reproducing a reaction time difference across all syllogisms between the high and low logic groups (p = .110) in the human data. However, the human high-logic group was significantly slower on believable-invalid problems (p = .020) as predicted by past research. For the computational model, the initial parameters of subjects mostly vary in the application of lesion damage. There is no distinction made for those with a more logical cognitive style which is more apt to resist belief-bias. If such a distinction were added, it would become possible for an increase in reaction time to emerge for believable-invalid problems. This would also be one part of the remedy for increasing the low variability, in terms of standard errors, found for the model data compared to human data for reaction time results.

The final part of the remedy for this problem of low variability in reaction time for model data comes from the deterministic nature of the formal mental model system in the construction of alternate models. Experiments by Bucciarelli and Johnson-Laird (1999) had participants construct alternate models of premises to refute conclusions to categorical syllogisms using cut-out shapes or pen and paper methods. While participants did search for counter-examples in ways that utilized the major operations of the mental model program – adding or moving tokens and breaking entities in two – they varied considerably from one another in what they did, and even the same participant could vary when encountering a similar problem twice. Construction of a grammar with alternate rules allowing for alternate ways to modify or represent

problems would add more variability to the model generation process. It could add a greater potential for occasionally largely increased reaction times that is presently lacking. For example, the current computational model will add all tokens it can possibly add at once during the add-token operation; the number of tokens added and the order in which this is done could vary.

The high versus low logic distinction could also generate differences in model formulation, and adjust the degree to which finding a model that agrees with a belief influences decision-making. Low-logic individuals could apply more “satisficing” (Evans, 2007) searches which look for a single model that supports a belief, and if it is found, decide that their belief is correct. High-logic individuals in contrast could be more apt to employ more exhaustive and analytical approaches which require a conclusion to be true in all models to be valid. One major concern highlighted by Bucciarelli and Johnson-Laird (1999) in their computational model, that reasoners do not adopt fixed interpretations for each kind of premise, is already employed in this model through errors of conversion and faulty implications. One final difference was noted in this paper that was not implemented in the current model. Reasoners demonstrated a marked difference in understanding what constituted a proper refutation of an O-type conclusion (‘Some X are not Y’) often using a model showing ‘Some X are Y’ as a refutation; they were far less successful at refuting syllogisms in this mood (35% accuracy) compared to other moods: A-type (‘All X are Y’) 72%, I-type (‘Some X are Y’) 66%, E-type (‘No X are Y’) 82%. The computational model as-is can recognize correct refutations with relatively equal ability. A more complete discussion of possible modifications to the falsification process are beyond the scope of this topic, though their investigation would prove useful for ensuring stronger correlations between simulated and human performance measures.

The computational model displays highly significant accuracy correlations to human data across all lesion groups. The significant congruency by group interaction for accuracy provides a link for distinct analytical and heuristic reasoning strategies to different areas of the brain. Frontotemporal areas support belief-driven responses, while bilateral parietal areas support spatial manipulation of models for formal analytical

procedures. The performance measures for human and model data show a high degree of coherence between each other, and with previous research. The model itself is able to incorporate cognitive biases like the atmosphere and congruency effect into one system, and provide an account for systematic errors in misrepresenting premises. Other explanations of human reasoning over syllogisms can provide piece-wise theories or models of aspects of reasoning, but few can demonstrate such a wide variety of coherence all at once. Adding distinctions in performance due to cognitive style, a differential preference for heuristic or analytical search, may improve the fit to human data; though what has been demonstrated presently provides compelling evidence for a fractionated deductive reasoning system which involves different but interactive reasoning strategies.

6.2 Theoretical Discussion

Returning to the discussion of theories of the structure of the reasoning mind, a fractionated system of deductive reasoning appears to be in the strongest position to account for our intuitions of rational behaviour, neuropsychological evidence, and behavioural evidence of deductive reasoning. Massive modularity suggests a diverse and isolated network of evolutionarily specified modules which quickly and reactively respond to environmental stimuli, similar to that of a reflex arc. Similar to reflexes, these modules would tend to exhibit a trait of *cognitive impenetrability* – meaning we are unable to be consciously aware of nor influence the activity of these reasoning modules – though this contrasts with our intuitive understanding of what rational behaviour is. Rational choices are thought to be selected for a reason: to provide reasoned means for satisfying the goals we choose to pursue. Rational actions should demonstrate a gap between stimulus and response for some degree of decision-making or weighing of alternatives to occur. Environmental conditions should not be sufficient for rational action as it would be for reflexive action. Cognitively impenetrable modules supporting reasoning would ultimately deny that deliberate reasoning even occurs. Furthermore, rigid and evolutionarily specified modules supporting reasoning would be unable to exhibit the vast variability in performance between participants, or even for a single participant on similar syllogisms as noted by Bucciarelli and Johnson-Laird (1999).

A simple heuristics account which suggests multiple modules supporting only an inductive-heuristic reasoning system also fails to account for neuropsychological evidence. Reasoning with familiar material often recruits a frontotemporal network, and unfamiliar or content-free material recruits a bi-lateral parietal network. If a fractionated heuristics system explained all reasoning, it is difficult to explain how linguistic or spatially relevant networks could be preferentially recruited depending upon content. The probabilistic heuristics model (PHM) proposed by Chater and Oaksford (1999) does make an attempt to explain activation of the conflict detection system as a result of conflicts among heuristics suggesting different conclusions. However, it is unclear how they would justify this distinction between linguistic and spatial network recruitment, as

their decision heuristics largely appear to draw upon linguistic inferences concerning the likely meanings of specific terms.

If we suggest that heuristics, or even rigid modules, are the primary driver of reasoning behaviour, we are also left with the difficult task of explaining how deliberate reasoning can take place at all. Bucciarelli and Johnson-Laird (1999) had participants manipulate cut-out shapes or use pen and paper to create their own models of situations as they tried to reason, with little instruction beyond to try and construct a picture of the premises to see which of the provided conclusions held. Use of operations similar to the mental models program were observed by subjects in these experiments. Bucciarelli and Johnson-Liard (1999) suggest these participants are not merely generating conclusions in accordance with atmosphere, nor are they selecting conclusions that match least informative premises like with the matching bias of PHM. Without a system supporting styles of reasoning other than intuitive assumptions from probable linguistic inferences, it is difficult to imagine how this task is accomplished by untrained individuals. Unless one is prepared to suggest that construction of models for reasoning is epiphenomenal and bears no impact on judgements.

Furthermore, the participants of the Vietnam Head Injury study, as well as the study by Stupple et al. (2011), showed a distinction between high and low logic individuals. High logic individuals took more time and showed a greater resistance to belief-bias heuristics to ensure greater performance – particularly for believable-invalid problems. Differences in the cognitive style of reasoning, and the brain networks recruited for different methods of approach, pose problems for many theories of logical reasoning. Simple heuristics or massive modularity have limited ability to explain this wide variation. A pure mental models or mental logic approach fails to account for belief-bias effects of reasoning, and why networks not related to their approach may be engaged; such as how mental logic explains visuospatial engagement, or how mental models explain linguistic network engagement.

A pure mental logic approach is constrained by the formal rules of logic. The inferential roles of logical terms completely determines a course of action which is unable to support the congruency effect. It lacks an explainable method for introducing

differences due to belief into the logical calculus, or why syllogisms with unbelievable conclusions would tend to be subject to a more rigorous degree of analysis than those with believable conclusions. Mental logic would have little methods for interaction with belief, and would have to be completely disregarded and overridden by a belief-heuristic system.

A pure mental models approach is also similarly constrained by logical rules, though a belief-bias system can be introduced to interfere with the search for alternate models. In a less pure version of mental model theory, belief-bias may cause an early termination in logical procedure if the initial model agrees with held beliefs. When attempting to construct alternate models, individuals may have increased difficulty in constructing models that are implausible to held beliefs. The application of formal rules through mental logic does not have the benefit of this affordance. As is explained in more detail by Johnson-Laird (2010) logic is monotonic, and as more premises are added the number of potentially valid conclusions increases, including a large number of silly but logically valid assertions. Humans are instead agents exhibiting rationality bounded by the constraints of time and limited cognitive resources. Creation of a problem space with incremental alterations in the form of alternate mental models provides a more time-optimized solution to deductive reasoning problems. Constraining this problem space further through a belief-bias system conserves costly cognitive resources, and allows effects like the congruency effect to surface.

As an example of what would happen to our results with a more pure logic system, the computational model was run on the syllogisms while excluding the heuristic system from participating. While this represents a pure mental models approach, its adherence to formal rules extends its implications to a pure mental logic approach as well. Occasional errors through mispresenting or converting premises was maintained in the general pattern completer in this simulation. As the results in Figure 36 demonstrate – with a pure mental model system employed on the left – the lines representing the accuracies of syllogisms with believable or unbelievable conclusions across valid and invalid problems becomes much more parallel, which effectively eliminates the congruency effect due to belief-bias. This demonstrates the necessity of

providing some plausible method for including a belief-driven heuristic system in the explanation of human deductive reasoning.

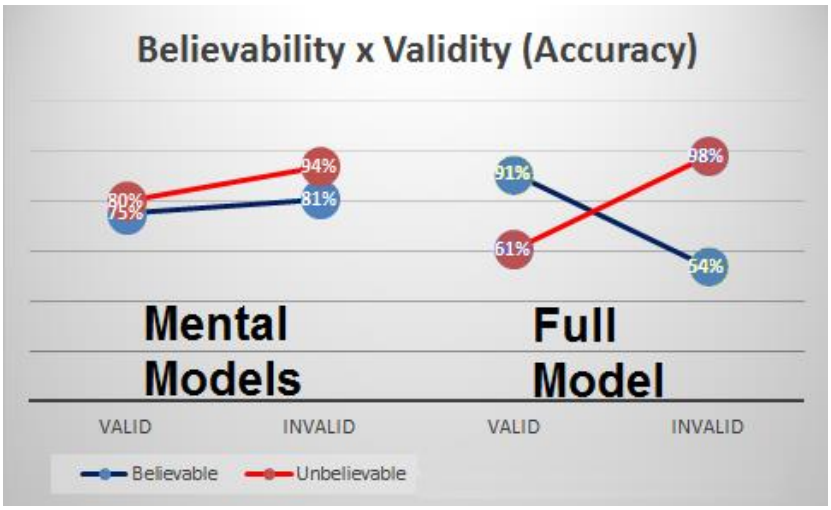


Figure 32

As the current investigation was contained largely to the formal and heuristic systems, the architecture demonstrated here does not deeply diverge from less assumptive (and more conservative) dual-mechanism theories. It bears the closest resemblance to *parallel-cooperative* (Barbey & Sloman, 2007) dual-mechanism theory, which proposes formal and heuristic systems operate in parallel and provide input to some conflict resolution process. Further research into the operation of the uncertainty maintenance system, or other future proposed systems, may help distinguish a fractionated deductive reasoning system of reasoning from conservative dual-mechanism theories.

Behavioural evidence can suggest particular modes of reasoning, and neurological investigations can tie these modes or functions to particular brain areas. Computational modelling, however, provides a means for testing different methods of organizing these system components to create better approximations of how these components interact. The computational model suggested here has successfully provided significant correlations with the accuracies of a number of human lesion groups. It has done so by incorporating various published biases of belief and atmosphere into a number of interacting systems: the formal system supported by a bi-lateral parietal network, the heuristic system through the frontotemporal network, the

prefrontal cortex general pattern completer, and uncertainty and conflict detection mechanisms. It has done so even without a highly flexible and variable alternate model generation process. Limitations with the subject pool prevented a deeper examination of these last two systems, though working within these limitations has still provided evidence for a fractionated system supporting deductive reasoning.

References

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.

Barbey, A. K., & Barsalou, L. W. (2009). Reasoning and problem solving: models. *Encyclopedia of neuroscience*, *8*, 35-43.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1995). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.

Bermudez. J. L. (2002). Rationality and psychological explanation without language. In J. L. Bermudez & A. Millar (Eds.), *Reason and nature: essays in the theory of rationality,* 233 – 264. New York, NY: Oxford University Press.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*(3), 247-303.

Carruthers, P. (2006). Simple heuristics meet massive modularity. *The innate mind*, 2, 181-98.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adapted mind*, 163- 228.

De Neys, W. (2006). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dualtask considerations. *Quarterly Journal of Experimental Psychology, 59*, 1070–1100.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations?. *Cortex*, *39*(1), 1-7.

Espino, O., Santamaría, C., & García-Madruga, J. A. (2000). Figure and difficulty in syllogistic reasoning. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.

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

Evans, J. S. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, *61*, 263-285.

Evans, J. S. B., Over D. E. (1996). *Rationality and reasoning. Psychology*, New York

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454-459.

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378-395.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321-339.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on psychological science*, *8*(3), 223-241.

Fodor, J. (1983). *The Modularity of Mind*. MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart*, 3-34. Oxford University Press.

Goel V., Buchel C., Frith C., & Dolan R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage 12*(5):504–514.

Goel V., Shuren J., Sheesley L., & Grafman J. (2004). Asymmetrical involvement of frontal lobes in social reasoning. *Brain 127*(4):783–790.

Goel V., Tierney M., Sheesley L., Bartolo A., Vartanian O., & Grafman J. (2006). Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. *Cereb Cortex* 17, 2245 – 2250.

Goel, V., Tierney, M., Sheesley, L., Bartolo, A., Vartanian, O., & Grafman, J. (2007). Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. *Cerebral cortex*, *17*(10), 2245-2250.

Goel, V. (2009). Fractionating the system of deductive reasoning. In *Neural correlates of thinking* (pp. 203-218). Springer Berlin Heidelberg.

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

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243-18250.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 153-191.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. American Psychologist, 80, 237–251

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

Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia*, *41*(12), 1668-1682.

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

Pinker, S., (1997). *How the Mind Works*, New York: W. W. Norton & Company.

Raymont, V., Salazar, A. M., Krueger, F., & Grafman, J. (2011). "Studying injured minds" – the Vietnam Head Injury study and 40 years of brain injury research. *Frontiers in Neurology, 2*(15), 1 - 13.

Rips L.J. (1994). *The psychology of proof: deductive reasoning in human thinking*. MIT Press, Cambridge.

Sperber, D., (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld and S. A. Gelman (eds.), *Mapping the Mind*, Cambridge: Cambridge University Press, 39–67.

Stanovich, K. (2004). *The robot's rebellion: finding meaning in the age of Darwin*. University of Chicago Press, Chicago

Stanovich, K. E. (2011). Rationality and the reflective mind. New York, NY: Oxford University Press.

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99– 118.

Simon H.A. (1983). *Reason in human affairs.* Stanford University Press, Stanford

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization science*, *2*(1), 125- 134.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*(3), 619-632.

Stupple, E. J., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*(2), 168-181.

Stupple, E. J., Ball, L. J., Evans, J. S. B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*(8), 931-941.

Varma, S. (2006). *A computational model of Tower of Hanoi problem solving* (Doctoral dissertation, Vanderbilt University).

Varma, S. (2014). The CAPS family of cognitive architectures. In S. E. F. Chipman (Ed.), *The Oxford Handbook of Cognitive Science*. Oxford University Press.

Wetherick, N. E. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology*, *2*(1), 111-124.

Wetherick, N. E., & Gilhooly, K. J. (1990). Syllogistic reasoning: Effects of premise order. *Lines of thinking*, *1*, 99-108.

Appendix A: Patient Group Lesions

Bilateral Parietal Patients

Patient ID	BA 7 (right)	BA 40 (right)	BA 7 (left)	BA 40 (left)	Total
230	0.24	0.00	42.18	4.36	46.78
408	15.11	8.02	0.00	0.00	23.13
439	0.65	19.06	0.00	0.00	19.71
1061	0.02	15.44	0.00	0.00	15.46
1298	0.00	22.95	0.00	0.00	22.95
1324	7.73	56.65	0.00	0.00	64.38
1341	11.17	0.00	51.42	1.73	64.32
1366	0.00	20.86	0.00	0.00	20.86
1434	30.50	0.00	0.00	0.00	30.50
1443	0.00	33.61	0.00	0.00	33.61
1461	17.01	12.47	0.00	0.00	29.48
1510	0.00	0.00	1.09	27.18	28.27
1621	0.00	0.00	20.82	6.84	27.66
2005	16.48	0.19	3.38	0.00	20.05
2028	10.33	27.27	0.42	0.04	38.06
2116	29.54	33.48	22.20	32.17	117.39
2341	0.00	0.04	32.75	17.09	49.88
3081	57.57	0.21	0.00	0.00	57.78

* damage in a BA is represented as a % proportion of that area

Frontotemporal Patients (left hemisphere)

Patient ID	BA 21 (left)	BA 22 (left)	BA 47 (left)	Total
5	10.44	2.49	11.61	24.54
103	7.83	24.05	0.00	31.88
181	12.30	23.06	0.20	35.56
182	0.00	1.24	26.02	27.26
318	0.00	0.00	10.64	10.64
473	25.06	11.73	9.80	46.59
495	19.63	18.05	0.00	37.68
528	0.00	5.90	6.89	12.79
1003	17.74	17.42	3.62	38.78
1127	1.02	9.88	10.13	21.03
1433	9.76	30.29	0.00	40.05
1561	0.04	0.00	9.96	10.00
1715	19.71	19.42	0.00	39.13
2135	0.00	0.10	23.08	23.18
2146	1.44	0.25	10.85	12.54
2288	3.11	24.86	0.00	27.97
2386	0.00	4.36	25.24	29.60

Appendix B: Human Demographics

Measure	Controls (N = 37)	Parietal Patients (N = 18)	Frontotemporal Patients (N = 17)
Age (years)	59.27 (3.78)	58.11 (3.32)	58.89 (3.53)
Education (years)	14.49 (4.29)	14.68 (4.36)	14.42 (4.00)
CT Total Volume Loss (cm^3)	0	29.05 (22.43)	30.24 (22.37)
WAIS Verbal IQ	110.41 (12.28)	103.47 (13.00)	108.47 (16.17)
WAIS Performance IQ	111.86 (12.25)	99.28 (14.85)	103.16 (16.75)
WAIS Full IQ	111.92 (11.65)	102.67 (12.06)	106.63 (17.06)
WAIS Working Memory	106.19 (12.74)	92.83 (12.28)	100.63 (19.15)
WMS Working Memory	106.84 (13.36)	98.00 (12.94)	101.32 (16.18)
WMS General Memory	107.62 (12.96)	102.05 (16.92)	94.32 (15.40)

Appendix C: Categorical Syllogisms

LEGEND

Bel = Believable conclusion
Unbel = Unbelievable conclusion
Con. = Conclusion congruent with belief
Incon. = Conclusion incongruent with belief
V = Conclusion Logically Valid
NV = Conclusion Logically Invalid
SM = Single-model problem
MM = Multiple-model problem

[1] all fruit are pears
all bananas are fruit
no bananas are pears
Bel-NV Incon. SM

[2] all cars have four wheels
no scooters have four wheels
no scooters are cars
Bel-V Con. SM

[3] all gods are immortals
no immortals are men
no men are gods
Bel-V Con. SM

[4] all bikes are red
some bikes are broken
no broken bikes are red
Unbel-NV Con. SM

[5] all airplanes can fly
some boats can not fly
some boats are not airplanes
Bel-V Con. MM

[6] no liquids are red
all paints are liquids
some paints are red
Bel-NV Incon. SM

[7] no cuban cigars are dogs
no cuban cigars are cats
no cats are dogs
Bel-NV Incon. MM

[8] no coffee contains nicotine
no nicotine contains tea
no tea contains coffee
Bel-NV Incon. MM

[9] no skiers are smokers
some men are not skiers
all men are smokers
Unbel-NV Con. SM

[10] no cats have stripes
some tigers are cats
Some tigers do not have stripes
Unbel-V Incon. MM

[11] no poisons are sold at the grocers
some mushrooms are sold at the grocers
some mushrooms are not poisons
Bel-V Con. MM

[12] no men are children
some men are girls
all girls are children
Unbel-NV Con. SM

[13] no olympic runners are smokers
some smokers are not men
some men are olympic runners
Bel-NV Incon. SM

[14] some felines have gills
all felines are cats
some cats have gills
Unbel-V Incon. SM

[15] some apples are sweet fruit
all sweet fruit are grapes
some grapes are apples
Unbel-V Incon. SM

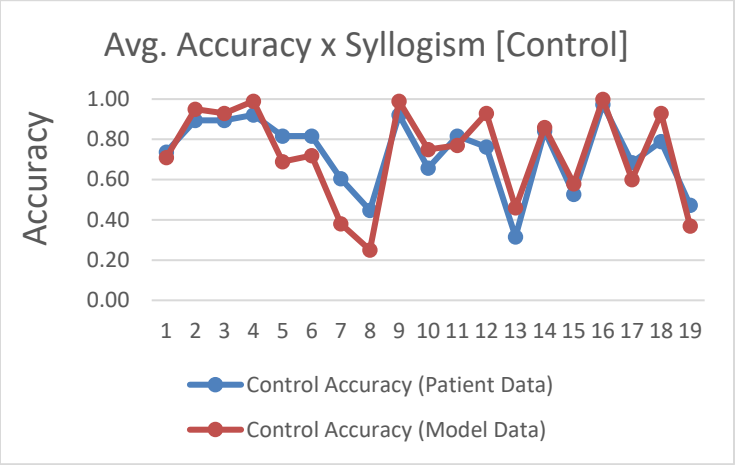
[16] no fruit are blue
some apples are fruit
all apples are blue
Unbel-NV Con. SM

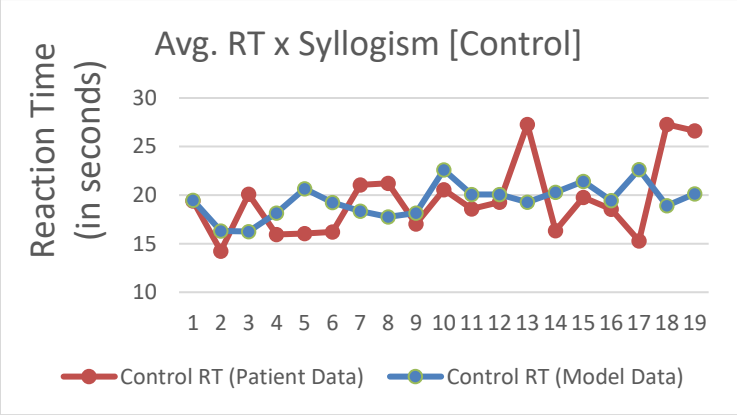
[17] some dogs do not have ears
all dogs are german sheperds
some german sheperds do not have ears
Unbel-V Incon. MM

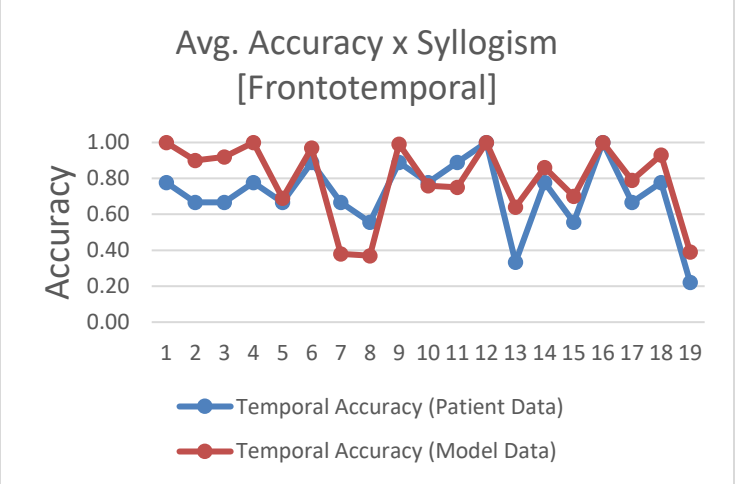
[18] some italians are not martians
no french are Italians
some french are martians
Unbel-NV Con. SM

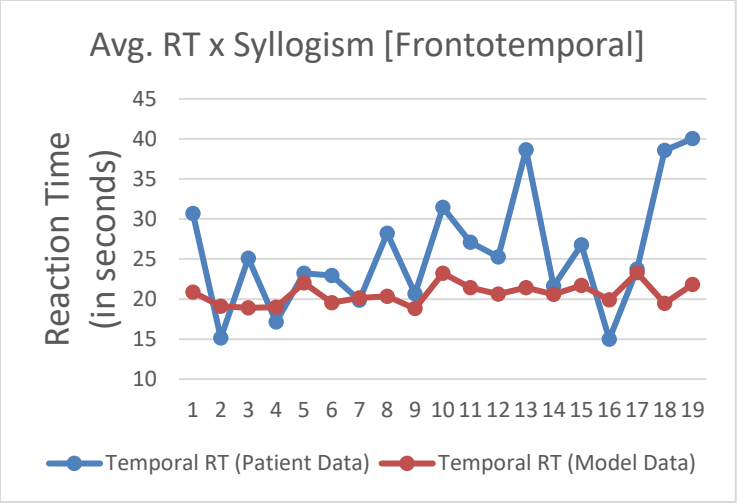
[19] some mice are not rabbits
some cats are not mice
some cats are not rabbits
Bel-NV Incon. MM

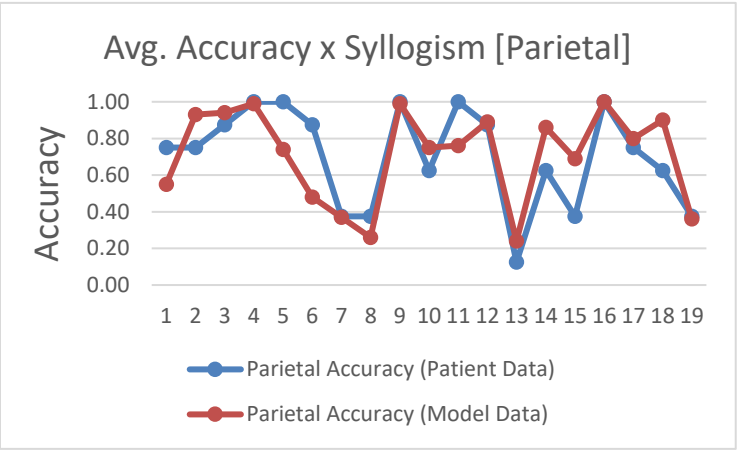
Appendix D: Correlation Graphs

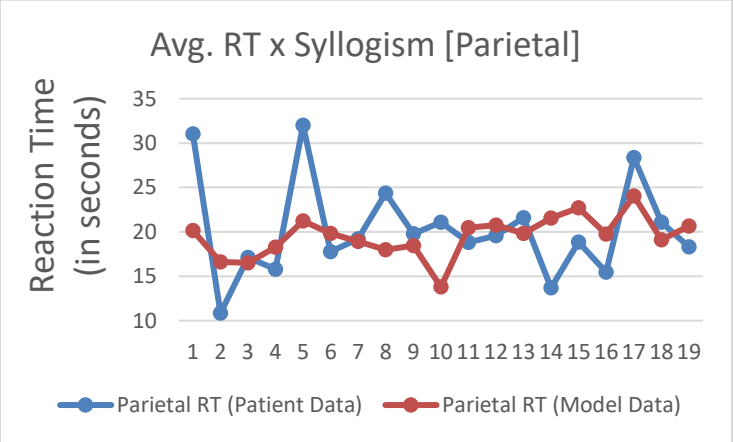







Avg. RT x Syllogism [Frontotemporal]
Reaction Time (in seconds)
45
40
35
30
25
20
15
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Temporal RT (Patient Data)
Temporal RT (Model Data)


Avg. Accuracy x Syllogism [Parietal]
Accuracy
1.00
0.80
0.60
0.40
0.20
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Parietal Accuracy (Patient Data)
Parietal Accuracy (Model Data)


Avg. RT x Syllogism [Parietal]
Reaction Time (in seconds)
35
30
25
20
15
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Parietal RT (Patient Data)
Parietal RT (Model Data)