LEFT HEMISPHERE LESIONS DIFFERENTIALLY IMPACT CONDITIONAL REASONING WITH FAMILIAR AND EMOTIONAL CONTENT

MARY RUTH REBECCA MARLING

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

FEBRUARY 2015

Abstract

Conditional (*if-then*) reasoning has been widely studied in the cognitive literature, and in the past decade, neuroimaging studies have started to investigate brain networks recruited to solve these logical conditionals. A meta-analysis of these neuroimaging studies of healthy adults has shown that conditional arguments are primarily associated with left-lateralized activation in the parietal and frontal lobes (Goel, 2007). Beyond logical form, content factors such as *belief- logic congruency, familiarity, and emotion* have been shown to recruit networks different from the main effect of reasoning. To date, conditional connectives have not been investigated using traumatic brain injury patients, therefore, the goal of this thesis was to study the effect of brain lesions on conditional reasoning.

A whole brain analysis using voxel-based lesion symptom mapping (VLSM) was conducted on 72 neurological patients with unilateral lesions in order to explore the impact of brain lesions on reasoning accuracy scores. Results indicated that conditional reasoning with familiar content is highly dependent on left hemisphere intactness, whereas right hemisphere volume loss does not inhibit performance and in some conditions may even lead to improved performance. In particular, we found that familiar believable content failed to benefit patients with left hemisphere lesions. Additionally, VLSM analysis isolated a region in the left medial prefrontal cortex (MPFC) deemed necessary for reasoning with emotional content, the 10 patients with lesions in this cluster performed significantly worse than all other patients and controls on emotional conditionals.

Our findings provide additional evidence that reasoning processes involving familiar content are largely left lateralized and that the ventromedial prefrontal cortex is specifically engaged in reasoning about emotional content. This is the first study to use a lesion analysis to investigate conditionals, and thus contributes important new information to the existing neuroimaging literature.
Keywords: Lesions; Conditional reasoning; Emotion; Belief-logic congruency; Congruency; Familiar content; Deductive reasoning; VLSM; Traumatic brain injury

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Vinod Goel, for all his guidance, insight, and support throughout the duration of this thesis project. I have gained many valuable skills while working on this project, and am infinitely grateful for the knowledge and experience gained along the way.

I would also like to thank Dr. Maggie Toplak, part of my supervisory committee, for her thoughtful feedback and critical comments throughout the final stages of this project. Her input dramatically improved the quality of discussion and presentation of results.

I would like to give special thanks to Gorka Navarette for his contribution to the data analysis in developing a special MATLAB script for analyzing the data. And special thanks to Iryna Pyshonak, a research volunteer, for assisting in running additional ABLe analyses and collecting and double checking results along the way.

I would like to thank Dr. Walter Heinrichs for chairing my defence committee, and providing valuable feedback during the defence. Also, I would like to give special thanks to Dr. Judy Pelham for agreeing to act as the external examiner, and for her thoughtful input during the defence.

Lastly, many thanks to my partner for her moral support and encouragement throughout the completion of this thesis, it made a world of difference.

Introduction

The conditional form, 'if … then,' is central to human inference, allowing people to express causal relationships, e.g., *if you pet your cat, it will purr*; or laws, e.g., *if you get caught speeding, you will get a ticket*; or trade conditions, e.g., *if you give me your scarf, I'll give you an umbrella*. These are just a few examples of the many ways 'if-then' conditionals are used in everyday life. Conditional logic is also pivotal in a number of domains including computer programming, mathematics, and philosophical discourse. For more than 50 years, conditional reasoning has been the focus of rigorous investigation in cognitive psychology (Oaksford & Chater, 2010), and more recently neuroscience (Goel, 2007; Prado, Chadha, & Booth, 2011).

Over the past two decades, theories of deductive reasoning have been tested using neuroimaging techniques such as position emission tomography (PET) and functional magnetic resonance imaging (fMRI) to look at activation patterns during reasoning (for an overview of these neuroimaging studies see Goel, 2007; and Prado et al., 2011).

Two meta-analyses covering the first decade of deductive neuroimaging research indicate that the specific networks activated during reasoning depend on the type of deductive argument (e.g. syllogisms, conditionals) and content used to present the argument (Goel, 2007). Therefore, neuroimaging findings suggest that reasoning is not a unitary brain system as suggested by reasoning theories, but is "dynamically configured in response to certain task and environmental cues" (Goel, 2007). These environmental cues include differences in experimental stimuli such as content, determinacy, conflict, emotion, believability, and familiarity, or the type of deductive argument (i.e.

conditional, syllogism, or transitive form) (Goel, 2007; Prado et al., 2011). Thus, neuroimaging studies have shown that there are deductive brain networks differentially responsive to content and argument form.

The main objective of this study was to investigate the neural structures necessary for deductive reasoning by manipulating the effect of content as patients with brain lesions engage in conditional reasoning tasks. Three kinds of content manipulations were examined (see Table 1 for an example of all arguments presented): 1) *familiarity*, familiar (i.e. real-world) content versus abstract, unfamiliar content (i.e. A, B, C); 2) *congruency*, congruent arguments that support our beliefs (believable and valid or unbelievable and invalid) versus incongruent arguments with logical conclusions that contradict our beliefs (believable and invalid or unbelievable and valid); and 3) *emotion*, emotional content (e.g. *all Nazis were murderous*) versus non-emotional content (e.g. *some people are not children*). Since there are no studies of conditional reasoning involving lesion patients, this study seeks to address this gap in neuropsychological studies in addition to validating findings derived from brain activation studies with healthy, normal subjects.

The Role of Content in Reasoning

Neuroimaging studies have looked at various dimensions of logical reasoning by manipulating content aspects such as belief-congruency, familiarity (familiar versus unfamiliar), emotion ("hot" and "cold"), and determinacy (determinate conclusions that follow from the premises versus indeterminate conclusions) in order to observe how they impact the ability of a participant to arrive at logically correct conclusions. For instance, Goel (2007) identified neural systems responsible for familiar and unfamiliar content. Familiar content engages a left frontal and left temporal network, while unfamiliar

content engages a bilateral parietal and dorsal prefrontal cortex (PFC) system (Goel, 2007). Furthermore, Goel and Dolan (2003b) identified systems for dealing with information that conflict with beliefs; this conflict often leads to a phenomenon termed belief-bias. Belief-bias occurs when participants choose an answer consistent with their beliefs but inconsistent with the underlying logical structure of the arguments. According to Goel and Dolan (2003b), proper detection of this conflict requires the right PFC, whereas susceptibility to a bias is connected to activation in the right ventromedial PFC (vmPFC).

Regarding the effect of determinacy, lesion and neuroimaging studies have shown that there are two distinct neural systems for dealing with certain and uncertain information (Goel et al., 2007). Goel et al. (2007) found that patients with focal lesions in the left PFC were impaired on reasoning trials with complete information; in contrast, patients with focal lesions to the right PFC were impaired on trials with incomplete information (this was also inadvertently demonstrated in Parsons and Osherson (2001), see Goel (2007) for an explanation).

Lastly, while examining the effect of emotion on logical processes, Goel and Dolan (2003a), found that reasoning with emotional content was associated with activation in the bilateral vmPFC, whereas neutral syllogisms were associated with activation in the dorsolateral PFC. Goel and Dolan (2003a), also reports that patients with damage to the vmPFC experienced significant impairments emotionally and socially (see Anderson et al, 2000).

How Lesion Studies Compliment Neuroimaging Techniques

Neuroimaging studies have found that the neural substrates of reasoning consist of a fractionated system composed of various brain systems dependent on the particularities of the reasoning task (Goel, 2007). These studies provide evidence for the sufficiency of these systems. However, certain inferences cannot be made from fMRI studies. For instance, even though a reasoning task may be correlated to activation in a particular brain area, it is not clear whether that area is *necessary* for task performance because areas not pertinent to the task itself may become activated due to connections to regions *necessary* for a task (Rorden & Karnath, 2004). In addition, fMRI cannot detect the importance of regions that are always active irrespective of task—just because an area does not show a significant modulation of blood-flow during a task does not mean it does not support task performance ("absence of evidence is *not* evidence of absence!") (Rorden & Karnath, 2004). Furthermore, it is important to note that the way imaging data is processed (i.e. timing, contrasts used, kinds of responses analyzed—correct versus incorrect, baselines used etc.) can affect what areas appear in activations, thus accounting for some of the variation in results across neuroimaging studies (Goel, 2007). For instance, the temporal sequence of activation events in functional imaging studies can vary widely based on differences in stimuli timing, thus there is no exacting method for understanding the temporal order of events (see Rorden and Karnath, 2004 for a fuller treatment of the pros and cons of each both functional imaging and lesion methods).

With regards to lesion studies, damage is typically heterogeneous, making it very difficult obtain several patients with the exact same lesions, thus the power to find deficits specific to particular brain structures is limited by the lesion patterns in a given

sample. In addition to this, it is also difficult to know how much an individuals brain has reorganized or compensated for damage, thus even if there were two patients with identical damage, identical recovery or reorganization may not occur. Given the limitations of neuroimaging and lesion methods, it is important to use as many techniques as possible to fully understand and investigate the anatomy and temporality of brain functions (Rorden & Karnath, 2004).

Lesion studies help compliment fMRI studies by revealing the neural systems and structures *necessary* for reasoning. FMRI studies can only tell us what brain areas are involved in a process, whereas the lesion method can help us infer if an area is required. Furthermore, lesion studies circumvent technical limitations of using fMRI on brain injury patients. For instance, fMRI might fail to detect brain activity in lesion patients due to reduced brain metabolism to damaged areas or excessive blood flow to neighboring areas (referred to as "luxury perfusion") (Rorden & Karnath, 2004). Therefore, a functioning area may not show robust activation patterns due to poor blood flow patterns. Thus, the lesion method avoids this complication by not having to account for signal loss due to reduced blood flow.

In contrast, fMRI offers the advantage of offering more analysis opportunities due to the fact lesion patients are less common than the general population of non-brain damaged individuals. And despite limitations in the temporal sequence of activation events, fMRI gives neuroscientists the tools necessary to understand the approximate order of activation events. Therefore, both activation and lesion studies have strengths and weakness, and both can be seen as complimentary and vital for uncovering brain regions and networks responsible for reasoning processes. Unfortunately, very little

lesion data has been published in the domain of reasoning (see Goel et al. (2007) and Waechter, Goel, Raymont, Kruger, & Grafman (2013) for examples). Thus, even though we know much about the temporal sequence of brain activations during reasoning, it is not clear what brain systems are vital for logical reasoning. This lesion study seeks to add clarity to this issue in addition to complementing neuroimaging findings with patient data.

Purpose and Hypotheses

The aim of the present study is to help validate and clarify previous findings derived from neuroimaging studies of healthy, normal subjects, in addition to uncovering how brain damage affects reasoning abilities. Specifically, this study seeks to investigate how brain lesions affect conditional reasoning, and in particular, how it affects conditional reasoning as manipulated along the dimensions of familiarity (*familiar* and *unfamiliar*), belief-congruency (*congruent* and *incongruent*), and emotion (*emotional* versus *non-emotional*). The main objectives were to 1) establish what neural structures are *necessary* to successfully solve conditional problems, 2) delineate the brain systems required to carry out reasoning along the aforementioned dimensions and compare findings to the neuroimaging literature.

We tested three hypotheses: 1) that the left hemisphere patients would perform poorly with familiar conditionals relative to the right hemisphere group and controls; 2) that left hemisphere patients would not benefit from the effect of congruent conditionals, whereas controls and right hemisphere patients with intact left hemispheres would benefit from congruent content; and 3) that lesions to the LMPFC would impair reasoning with

emotional content, but that reasoning with familiar content and unfamiliar content should not be affected.

Method

Patient Selection

The present study is based on archival data collected between 2004-2006 as part of the Vietnam head injury study (VHIS) (see Raymont, Salazar, Krueger, & Grafman (2011) for more details about VHIS). All patients and controls were male and veterans of the Vietnam War. During service, patients suffered head trauma and controls remained intact.

This study is part of the research project "Vietnam Head Injury Study – Phase III: A 30-Year Post-Injury Follow-up Study." It received ethics approval from the Office of Research Ethics, and protocol approval from the Institutional Review Board of the National Institute of Neurological Disorders and Stroke. All controls and patients gave written informed consent prior to participation.

In order to aid with comparisons of findings to existing neuroimaging literature and avoid confounds due to handedness, only right-handed individuals and right-handed but forcibly left (due to injury) were included in this analysis (left-handed, ambidextrous, and left-handed but forcibly right were excluded). To control for extent of damage, patients with less than 1 cm^3 or greater than 100 cm^3 of volume loss were excluded. Furthermore, to examine differential contributions of each hemisphere, patients with bilateral lesions were excluded, leaving 39 unilateral left and 33 unilateral, right lesion patients (See Figures 1-2 for the lesion patterns of these two groups and Table 4 for average volume loss). Thirty-four, right-handed controls who were also Vietnam veterans

were used as a comparison group in all statistical analyses of the data.

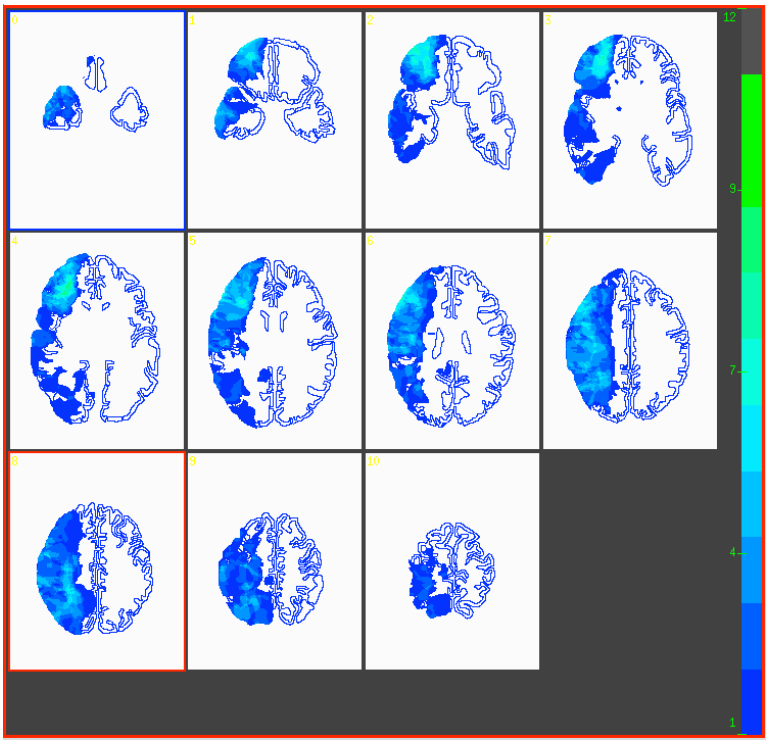


Figure 1. Overlay of damaged areas across all right hemisphere patients (n =39). Color indicates number of overlapping lesions. Images overlaid on a Damasio template, (R=L).

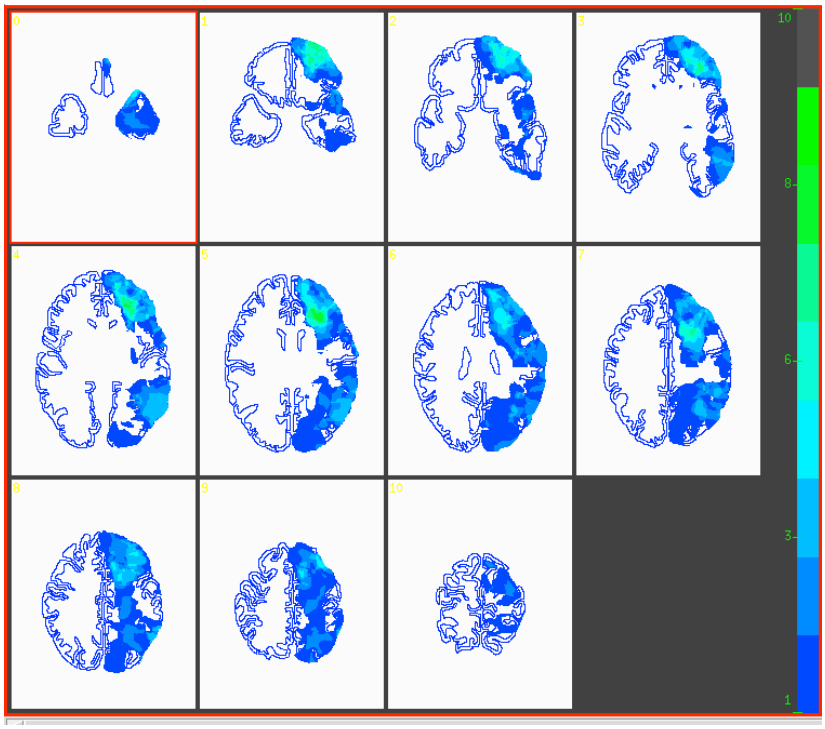


Figure 2. Overlay of damaged areas across all left hemisphere patients (n = 33, L=R).

Materials and Procedure

The task was administered through SuperLab Stimulus Presentation Software (Cedrus Corporation, San Pedro, CA). Before the experiment, participants were given an explanation of the nature of deductive reasoning and the task, and received several practice questions with the solutions. For the task, participants were instructed to determine the validity of each conclusion based on the premises by pressing 'C' if they thought it was valid or 'M' for not valid. For each trial, the premises and the conclusion were presented together on screen, and remained until a response was inputted. Participants were given unlimited time to complete each trial. Half the trials were valid, and the rest were either invalid or indeterminate. To avoid order effects, all trials were randomized, and the presentation of blocks counterbalanced.

Thus, a total of 12 conditional items (6 valid [determinate] and 6 invalid [indeterminate]) were analyzed. The first factor analyzed, *familiarity* (half valid), consisted of two levels, unfamiliar (4 non-emotional and 4 emotional combined) and unfamiliar (4 abstract). The second factor, *congruency* (half valid), had two levels, with 4 *congruent* items (when the validity of the conclusion and believability align; e.g., *valid and believable* or *invalid and unbelievable*) and 4 *incongruent* items (when the validity of the conclusion and believability do not align). The third factor analyzed, *emotion* (half valid), consisted of three levels with 4 items in each: *emotional*, *non-emotional*, and *no content* (see Table 1 for an example of these items and Table 3 for an item count in each factor).

Form	Non-emotional Content	No Content	Emotional Content
MP (determinate)	If a man is fat then he is immortal. Al Gore is fat. Al Gore is immortal. (Incongruent)	If there is A then there is B. There is A. There is B.	If a country has large reserves of oil then it may be invaded by America. Iraq has large reserves of oil. Iraq may be invaded by America. (Congruent)
MT (determinate)	If sheep eat shepherds then they will have golden fleece. There are no sheep with golden fleece. There are no sheep that eat shepherds. (Congruent)	If there is A then there is B. There is no B. There is no A.	If someone supports abortion then they are a murderers Democrats are not murderers Democrats do not support abortion. (Incongruent)
AC (indeterminate)	If there are Olympic Games then there are marathon runners. There are marathon runners. There are Olympic Games. (Incongruent)	If there is A then there is B. There is B. There is A.	If there are terrorist attacks against America then there will be a war. There is a war. There are terrorist attacks against America. (Incongruent)
DA (indeterminate)	If gazelles have hooves then they are fast runners. Gazelles do not have hooves. Gazelles are not fast runners. (Congruent)	If there is A then there is B. There is no A. There is no B.	If you are American then you abuse POWs. The Vietnamese are not American. The Vietnamese do not abuse POWs. (Congruent)

Table 1. All conditional items (12 trials) presented during task.

The four conditional forms included two valid forms: *modus ponens* (MP) and *modus tollens* (MT); and the two fallacies, *denying the antecedent* (DA), and *affirming the consequent* (AC). Each of these forms appeared three times in the set of twelve conditional items (see Table 2).

Conditional Form	Form of Premises	Form of Conclusion	Determinacy
Modus Ponens	If p then q. p.	q.	Determinate (valid)
Modus Tollens	If p then q. Not q.	Not p.	Determinate (valid)
Affirming the Consequent	If p then q. q.	p.	Indeterminate
Denying the Antecedent	If p then q. Not p.	Not q.	Indeterminate

Table 2. Example of the form and determinacy of the four conditional forms presented during the experiment

Main Category	Congruency (n = 8)	Familiarity (n = 12)	Emotion (n = 8)
Subcategories	Congruent (n = 4)	Familiar (n = 8)	Emotional (n = 4)
	Incongruent (n = 4)	Unfamiliar (n = 4)	Non-emotional (n = 8)

Table 3. Number of items in each cell for ANOVA analyses

Neuropsychological Assessment *

Patients and controls received a battery of neuropsychological assessments, reported in Table 4. The scores indicate that the groups' working memory and IQ are within the normal range (the normal range has a mean of 100 with a standard deviation of 15). Pair-wise comparisons† were used to examine differences in demographic and neuropsychological scores between the three groups. There were no significant differences between the groups in mean age, number of years of education, or total volume loss (between the left and right hemisphere groups). There were no significant differences between the left and right hemisphere groups on any of the Wechsler Adult Intelligence Scale III (WAIS-III) measures ($p > .22$). However, there were significant differences between the right hemisphere group and controls on all WAIS-III measures with the right hemisphere group performing significantly worse on all measures. But the left hemisphere group did not perform significantly different ($p > .40$) from controls except on the Working Memory Index Score (WMIS) ($p = .031$).

Pre-injury general intelligence was measured via the Armed Forces Qualification Test (AFQT-7A), which is administered to individuals admitted to the U.S. military. The

* Adapted from Waechter et al. (2013), which reported a similar assessment from another subset of patients in the VHIS study.

† All comparisons between groups in this section were controlled for false discovery rate (FDR) using the adjusted p-value method introduced by Benjamini and Hochberg (1995) and implemented in the function mafdr in the MATLAB Bioinformatics Toolbox Release 2014b (2014).

AFQT-7A is standardized and is highly correlated with WAIS IQ scores (Grafman et al., 1988). Comparisons between the groups revealed no significant differences in pre-injury AFQT-7A percentile scores between controls, left hemisphere and right hemisphere patient groups ($p > .22$, Table 4).‡

Measure	Normal controls ($n = 34$)	Left hemisphere patients ($n = 33$)	Right hemisphere patients ($n = 39$)
Age	59.36 (3.99)	58.24 (2.91)	57.97 (2.90)
Education	15.08 (2.60)	14.89 (2.32)	14.51 (2.46)
Pre-injury AFQT-7A	67.58 (28.17)	66.22 (23.72)	57.02 (28.19)
Total volume loss cm^3	-	25.94 (19.86)	24.73 (18.51)
WAIS Full IQ	109.3 (11.47)	106.6 (15.20)	101.7 (11.81)
WAIS Performance IQ	107.3 (12.16)	104.3 (16.22)	98.88 (13.37)
WAIS Verbal IQ	109.7 (12.57)	107.3 (14.44)	102.7 (12.15)
WAIS WMIS	105.1 (12.97)	96.93 (13.84)	97.23 (12.07)

Table 4. Demographic and neuropsychological mean (standard deviation) scores across the three participant groups. AFQT-7A: Armed Forces Qualification Test percentile rank; WAIS represents the Wechsler Adult Intelligence Scale-III.

CT Data Acquisition and Characterization of Lesion Location and Extent §

All brain scans were acquired with a General Electric Medical Systems Light Speed Plus CT scanner in helical mode (150 slices per image). Images were reconstructed with an in-plane voxel size of 0.4 x 0.4 mm, overlapping slice thickness of 2.5 mm, and a 1 mm slice interval. Skull and scalp components were removed using the Brain Extraction Tool (BET) algorithm incorporated in MEDx (Medical Numerics Inc.) (Smith, 2002). Patient CT volumes were imported into ABLe (Analysis of Brain Lesions; Makale et al., 2002; Solomon et al., 2007) and displayed as a series of slices in a Lightbox viewer. A trained neuropsychiatrist (V. R.) manually traced the lesions on all

‡ All comparisons between groups in this section were controlled for false discovery rate (FDR) using the adjusted *p*-value method introduced by Benjamini and Hochberg (1995) and implemented in the function mafdr in the MATLAB Bioinformatics Toolbox Release 2014b (2014).

§ Adapted from Raymont, Greathouse, Reding, Lipsky, Salazar, and Grafman (2008); and Waechter et al. (2013), which reports how the images were originally processed.

relevant slices. The tracings were then reviewed by neuropsychologist J. G., enabling a consensus regarding the limits of each lesion. Lesion volume was calculated and the brain images automatically registered to a template brain in Talairach space (Talairach and Tournoux, 1988). The template image was derived from a CT scan of a 27-year-old male, and a MRI of a 27-year-old male, transformed to Talairach space using a12-parameter affine linear transformation using the Automated Image Registration (AIR) algorithm in MEDx (Woods, Cherry, and Mazziotta, 1992; Makale et al., 2002). The volume was resliced at 17 degrees relative to the inferior orbitomeatal line, and 11 transverse slices that best match the Damasio (1989) templates were selected by a neuroradiologist and interactively labeled with Brodmann areas (BA) by reference to the Damasio templates. Although the locations of BAs in these templates are approximate, they are widely accepted in the neuropsychology and neurology communities. Lesion location and volume were determined using the previously described template volume as a reference in ABLe (MEDx version 3.44, Medical Numerics Inc.). Labels for lesion location were derived from the Automated Anatomical Labeling atlas in ABLe (Makale et al., 2002; Tzourio-Mazoyera et al., 2002; Solomon et al., 2007).

Voxel-Based Lesion Symptom Mapping Analysis

Reasoning scores and CT lesion tracings were entered into ABLe (version 2.8b) for a voxel-based lesion symptom mapping (VLSM) analysis using FDR ($q = .05$, which keeps the false positive rate at 5%) correction for multiple comparisons (see Solomon et al. (2002) for a description of this technique in section 3.3). *T*-tests were computed for clusters with a minimum of 10 contiguous voxels ($k = 10$) containing at least three patients with overlapping damage. Each *t*-test compared performance of patients with

damage affecting a given cluster (k = 10) to patients with damage outside that cluster, or if normal controls were selected as the comparison group, each cluster was compared to controls.

Results

VLSM Analysis

A whole brain VLSM analysis conducted in ABLe isolated a lesion cluster with ten patients exhibiting decreased performance on emotional conditionals in comparison to controls and other patients in the left medial PFC (LMPFC, see Figure 7 for details). This cluster appeared in two separate contrasts: 1) all emotional conditionals (uncorrected) where selected patients (cluster size k = 358 voxels) performed worse than controls (two-tailed p = .045) and non-selected patients (two-tailed p = .019); and 2) emotional determinate items (FDR corrected, k = 254 voxels) where selected patients performed worse than controls (p = .0096) and non-selected patients (p = .00034). A total of 10 unique patients appeared in these clusters, and their neuropsychological and demographic characteristics can be seen in Table 5. No other brain areas were selected by ABLe for emotional conditionals, thus this damage to this brain area was specifically related to impairments on emotional conditionals.

Beyond emotional conditionals, ABLe isolated a total 18 different clusters associated deficits in emotion, familiarity, and congruency in the conditional reasoning task. The majority of these clusters appeared in the left hemisphere, and in particular the frontal lobes and left temporal lobes. The few clusters that appeared in the right hemisphere did not exhibit deficits for the same contrasts as the left hemisphere clusters

(in fact, there were a few right hemisphere clusters that showed improved performance in comparison to the rest of the lesion patients). Thus, since there was a pattern of laterality consistent with the literature (Goel, 2007), it was decided to examine the hemispheres as a whole by comparing groups based on hemisphere of damage with ANOVA tests in order to examine interactions between hemisphere of damage and levels of each main factor since VLSM is limited to pair-wise comparisons and cannot perform factorial ANOVAs.

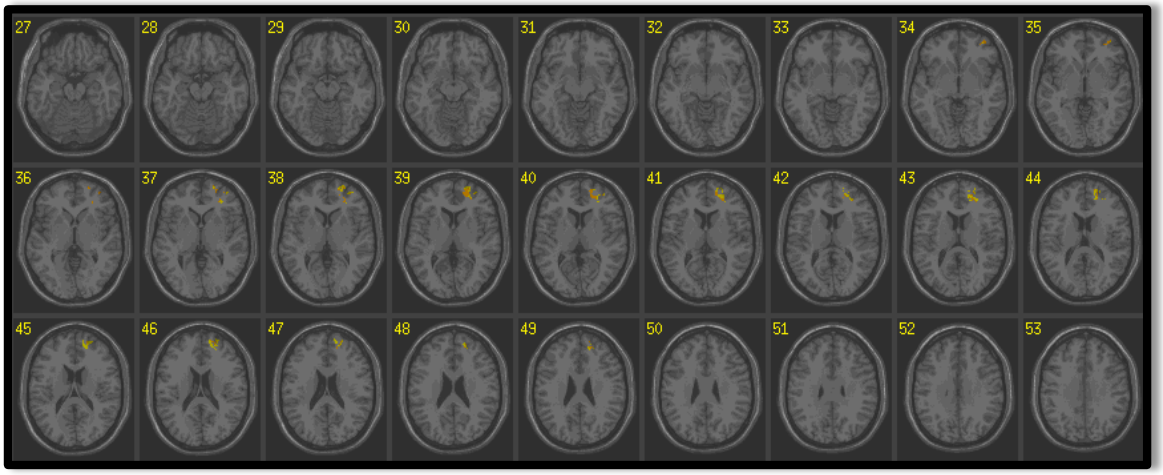


Figure 3. Lesion map produced in VLSM analysis for impairment associated specifically with emotional conditionals. This cluster spans the left superior and medial PFC (Talairach coordinates: -32, 44, 2 visualized on an MNI brain template).

Neuropsychological Characterization of the LMPFC Group

Pairwise comparisons between the LMPFC group selected by ABLe and the right hemisphere group, and the LMFPC group and controls[**] and revealed no significant

[**] The results for the pairwise comparisons between the right hemisphere group and controls are reported in the first section covering the neuropsychological characteristics of the groups. As noted in that section, there were significant differences between the right hemisphere group and controls on the WAIS-III IQ measures presented there.

differences in pre-injury AFQT-7A percentile scores, age, education, total brain volume loss, or any WAIS-III measure presented in Table 5 ($p > .36$).[††]

Measure	Left PFC patients (n =10)	Right hemisphere patients (n = 39)	Normal controls (n = 34)
Age	58.9 (1.36)	57.97 (2.90)	59.36 (3.99)
Education	14.1 (0.52)	14.51 (2.46)	15.08 (2.60)
Pre-injury AFQT-7A	68.44 (6.73)	57.02 (28.19)	67.58 (28.17)
Total brain volume loss cm^3	2.865 (0.48)	24.73 (18.51)	-
WAIS Full IQ	111.6 (5.01)	101.7 (11.81)	109.3 (11.47)
WAIS Performance IQ	110.9 (5.26)	98.88 (13.37)	107.3 (12.16)
WAIS Verbal IQ	109.6 (4.33)	102.7 (12.15)	109.7 (12.57)
WAIS WMIS	99 (4.22)	97.23 (12.07)	105.1 (12.97)

Table 5. Demographic and neuropsychological mean (standard deviation) scores for the LMPFC group and comparison groups.

Behavioural Performance

Analyses of parametric assumptions.

Given the small number of reasoning items in each category (4-8), we performed an exploratory data analysis looking at the normality and variance of each s distribution of scores in order to assess whether score distributions violated parametric test assumptions. Since there were 21 distributions to consider, the results of that analysis are not reported here, but can be viewed in *Appendix A*. To summarize the results here, there were no violations of normality as indicated by non-significant Levene's test for the equality of variances (see *Appendix A* for details of each result between each group for each ANOVA analysis). There were significant departures from normality as indicated

[††] All comparisons between groups in this section were controlled for false discovery rate (FDR) using the adjusted p-value method introduced by Benjamini and Hochberg (1995) and implemented in the function mafdr in the MATLAB Bioinformatics Toolbox Release 2014b (2014).

by significant Shapiro-Wilk tests of normality, but diagnostic tests such as kurtosis and skewness recommended by Field (2013) showed that none of the departures were extreme. Even though there were departures from normality, there is strong evidence via high quality Monte Carlo simulations using sample sizes of 25 (with varying effect sizes) that ANOVA tests are robust to violations of normality with distributions displaying mild skewness and mild to moderate kurtosis (Schmider, Ziegler, Danay, Beyer, and Bühner, 2010; a more recent Monte Carlo simulation study by Ferreira, Rocha & Mequelino (2012) confirms that ANOVA is equally as robust to violations of normality in comparison to the non-parametric Kruskal-Wallis test, which is considered far less powerful). Therefore, we used ANOVA tests rather than nonparametric tests to carry out the analyses in the proceeding sections.

Group by familiarity (familiar, unfamiliar).

Based on neuroimaging studies, we hypothesized that the left hemisphere patients would be more likely to perform poorly with familiar conditionals relative to the right hemisphere group and controls. A mixed measures ANOVA with Group (Left Hemisphere, Right Hemisphere, Controls) as the independent variable and Familiarity (All Familiar, All Unfamiliar) as the repeated measures variable revealed a significant interaction between Group x Familiarity, $F(2, 103) = 4.65$, $p = .012$, $\eta^2 = .083$ (see Figure 4). This interaction indicates that the reasoning accuracy across the two levels of Familiarity differed in the groups. Specifically, a follow-up one-way ANOVA showed a significant difference between the groups on conditionals with familiar content, $F(2, 103)$

= 6.39, p = .002, η^2 = .11.[‡‡] Planned contrasts revealed that the left hemisphere lesion group scored significantly lower than controls (p = .047, r = .19) and right hemisphere patients (p = .002, r = .35) on problems with familiarity (see Figure 4 for mean comparisons and *Appendix B* for a description of the effect sizes reported here). A follow-up one-way ANOVA showed no group differences on unfamiliar items, $F(2, 103)$ = .079, p = . 924. Furthermore, there was no significant main effect for Familiarity, $F(1,103)$ = .874, p = .352, r = .09, or Group, $F(2, 103)$ = 1.97, p = .145, η^2 = .037.

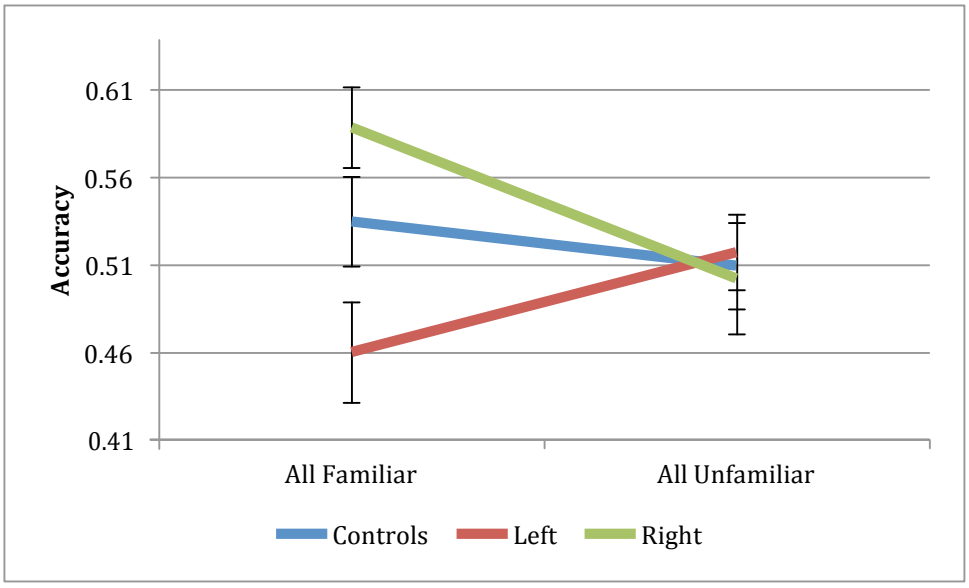


Figure 4. Significant group by familiarity interaction. Error bars represent standard errors.

Group by congruency (all congruent vs. all incongruent).

Our second hypothesis was that left hemisphere patients would probably not benefit from the effect of congruent conditionals, whereas controls and right hemisphere patients with intact left hemispheres would benefit from congruent content. To test this, a

---

[‡‡] The effect size reported here is eta squared which was calculated as the sum of squares of the effect divided by the total sum of squares of the model ($SS^2_{effect}/SS^2_{total}$), and not partial eta squared (sum of squares of the effect divided by $SS^2_{effect} + SS^2_{error}$, this reduces the denominator creating an inflated effect size that cannot be accurately compared across studies) that is normally seen in SPSS output (see Field (2013) and Cohen (1973) for an explanation).

mixed measures ANOVA with Group (Left Hemisphere, Right Hemisphere, Controls) as the independent variable and Congruency (All Congruent, All Incongruent) as the repeated measures variable revealed a significant interaction between Group x Congruency, $F(2, 103) = 4.11$, $p = .019$, $\eta^2 = .062$ (see Figure 5). Specifically, there was a significant difference between the groups on congruent problems, $F(2, 103) = 12.75$, $p < .001$, $\eta^2 = .20$. Since group sizes were not equal, Hochberg's GT2 pairwise test procedure designed for unequal group sizes was used for post hoc comparisons (Field, 2013). This analysis indicated that the left hemisphere lesion group performed worse on congruent items in comparison to controls (mean difference = -.14, $p = .006$). The left hemisphere group also performed worse than right hemisphere patients (mean difference = -.21, $p < .001$). There were no significant difference between the right hemisphere patients and controls ($p = .23$). Lastly, there were no significant differences between the groups on incongruent items $F(2, 103) = .28$, $p = .76$.

There was significant main effect for Congruency, $F(1,103) = 20.69$, $p < .001$, $r = .41$, with congruent conditionals being easier to solve ($M_{congruent} = 59\%$) than incongruent conditionals ($M_{incongruent} = 47\%$). Finally, there was a significant main effect of Group, $F(2, 103) = 6.16$, $p = .003$, $\eta^2 = .107$. Hotchberg GT2 post hoc comparisons showed that overall, left hemisphere patients performed significantly worse than right hemisphere patients (mean difference = - .13, $p = .002$). But there was no difference between left patients and controls (mean difference = -.07, $p = .13$) or right patients and controls (mean difference -.05, $p = .39$).

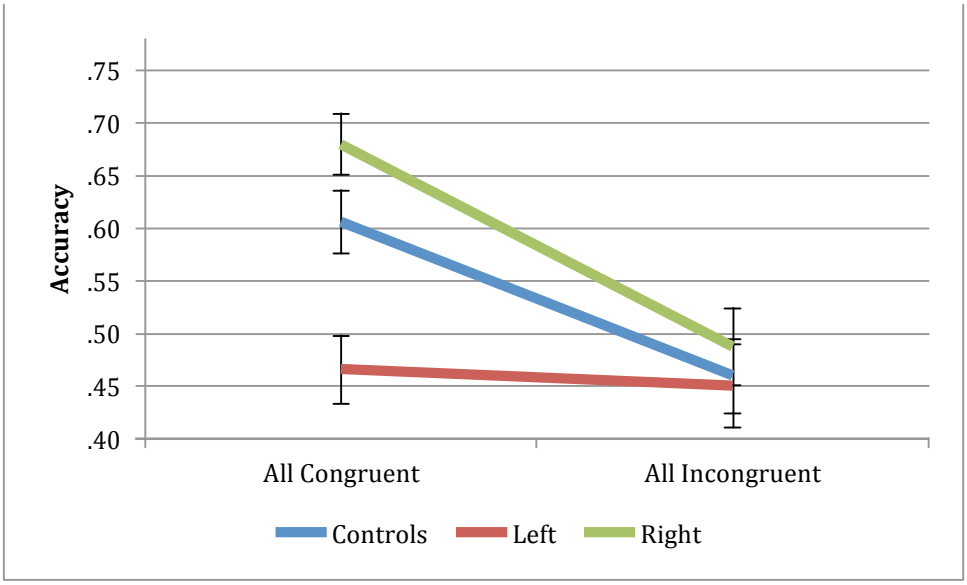


Figure 5. Significant group by congruency interaction, and main effect of group and congruency.

Group by emotion.

Since there was only one brain cluster in the Left MPFC identified by ABLe as being pertinent to conditional reasoning with emotional items, the following behavioural analysis only included left hemisphere patients with LMPFC lesions and compared them to right hemisphere patients and controls. Our hypothesis was that lesions to the LMPFC would specifically impair reasoning with emotional content, but that reasoning with familiar content and unfamiliar content should not be affected.

A mixed measures ANOVA with Group (Left MPFC, Right Hemisphere, Controls) as the independent variable and Emotion (Emotional, Familiar, Unfamiliar) as the repeated measures variable revealed a significant interaction between Group x Emotion, $F(3.69, 147.69) = 2.54$, $p = .046$, $\eta^2 = .059$ (see Figure 10). Follow-up one-way ANOVAs showed no significant group differences on any of the levels ($p > .4$) except for Emotional conditionals, $F(2, 80) = 3.78$, $p = .027$, $\eta^2 = .086$. This suggests that the groups

performed similarly on the other two levels without emotional conditionals, thus the effect of emotion drove this interaction. Hochberg GT post hoc tests indicated that LMPFC patients performed significantly worse on emotional conditionals than right hemisphere patients (mean difference = -.21, 1-tailed [since this hypothesis was directional] $p = .011$), and worse at a trend level in comparison to controls (mean difference = -.17, 1-tailed $p = .053$). There was no difference between controls and right hemisphere patients on emotional conditionals (mean difference = -.04, 1-tailed $p = .75$). Lastly, there was no significant main effect for Emotion, $F(1.85, 147.69) = .636$, $p = .519$, $\eta^2 = .007$, or Group, $F(2, 80) = 1.22$, $p = .30$, $\eta^2 = .029$.

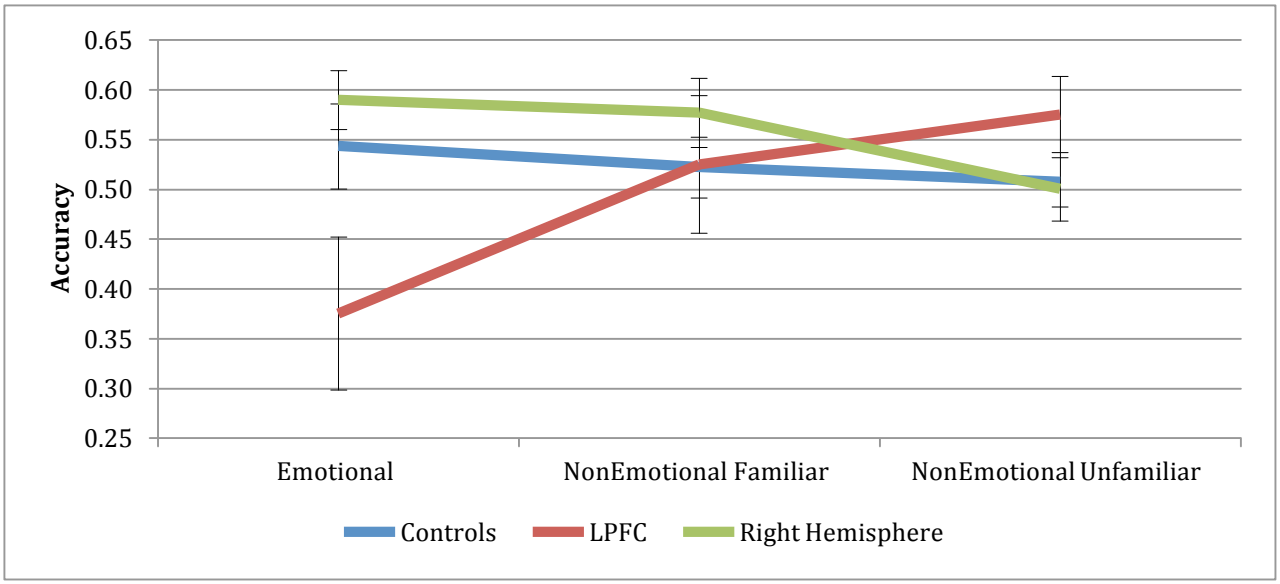


Figure 6. Group performance on conditionals with different levels of emotion. Significant group by emotion interaction.

Discussion

The purpose of this study was to examine the impact of lesions in the left and right hemispheres in resolving conditional reasoning problems while varying different

aspects of content such as familiarity, belief congruency, and emotion. As predicted, there were significant interactions between group membership and content for each of the content manipulations investigated. Before considering the particulars of these results, it is important to address the question of whether or not participants were performing at chance since several of the groups performed between 35% - 60%.

To understand the scores that look like chance performance, it is important to delve into the logical forms that lead to these accuracy rates. It is well established in the conditional literature that performance on the indeterminate forms is quite often below that of chance with endorsement rates vary from approximately 56% for AC and 49.5% for DA; whereas endorsement rates vary more frequently above chance on the determinate forms from close to 100% for MP to 65% for MT (see the meta-analysis by Schroyens, Schaeken, & d'Ydewalle, 2001; also see Noveck and Politzer, 1998). For example, in a neuroimaging study by Noveck et al. (2004), endorsement for the determinate form MP equaled 94% and MT = 77%, whereas for the indeterminate forms, endorsement for AC = 29% and DA = 31%, thus there is a clear division in difficulty between the determinate and indeterminate forms. Part of the explanation for this pattern is that AC and DA are considered logical fallacies, and many intelligent participants endorse them rather than reject them as being inconclusive due to problems thinking of counterfactuals (i.e. examples that would falsify the conclusion) (Oaksford and Chater, 2010). Consider the following AC argument used in this study:

> *If there are terrorist attacks against America then there will be a war.*
> *There is a war.*
> *There are terrorist attacks against America.*

In addition, inductive arguments have this particular form, therefore, it could be that the participants in this study treated the indeterminate forms like inductive arguments where they infer general rules from examples easily generated from everyday experiences (Angluin & Smith, 1983). Thus, the strategy employed to solve these problems would be appropriate in the context of an inductive reasoning problem.

The performance across each group for each familiar conditional form is depicted in Figure 7. It is clear the endorsement rates for each form are in line with the literature with lower endorsement rating for the indeterminate forms than the determinate forms. For AC, group endorsement ranges from 19% to 22%, meaning each group in this study failed to recognize the logical fallacy and answer correctly (i.e. false). Endorsement responses for DA range from 23% to 47%. Figures 7 depicts performance on the four conditional forms and Figure 8 depicts endorsement rates for both familiar and unfamiliar conditionals, the discrepancy in performance rates between the determinate and indeterminate forms is clear. For instance, determinate accuracy ranging from 59% to 83% is much higher than indeterminate accuracy, which ranges from 23% to 44%. Thus it is clear when the two forms (indeterminate and determinate) are combined that accuracy rates suffer and come close to chance (50%) performance. The reason why the main analyses were not broken into determinate and indeterminate forms is because we did not have enough items for each contrast (emotion, congruency, familiarity) to break them into these component forms. In closing, performance patterns across the groups in this study are in line with the cognitive and neuroimaging literature, therefore, participants were not performing at chance on each level of determinacy in the emotional, familiar, and congruent factors. Thus performance can be explained by the differential

impact of determinate and indeterminate logical forms rather than in the inability to do the task. Beyond this explanation, see *Appendix C* for a demonstration of how the binomial theorem accounts for the "chance" level performance seen in these groups.

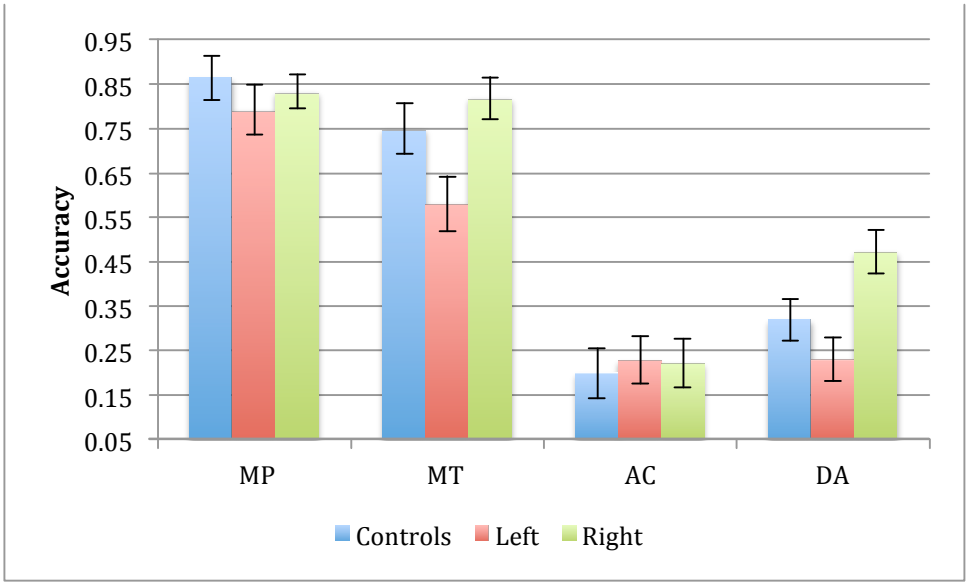


Figure 7. Performance on the different unfamiliar conditional forms (two items per form). Error bars represent the standard error of the mean.

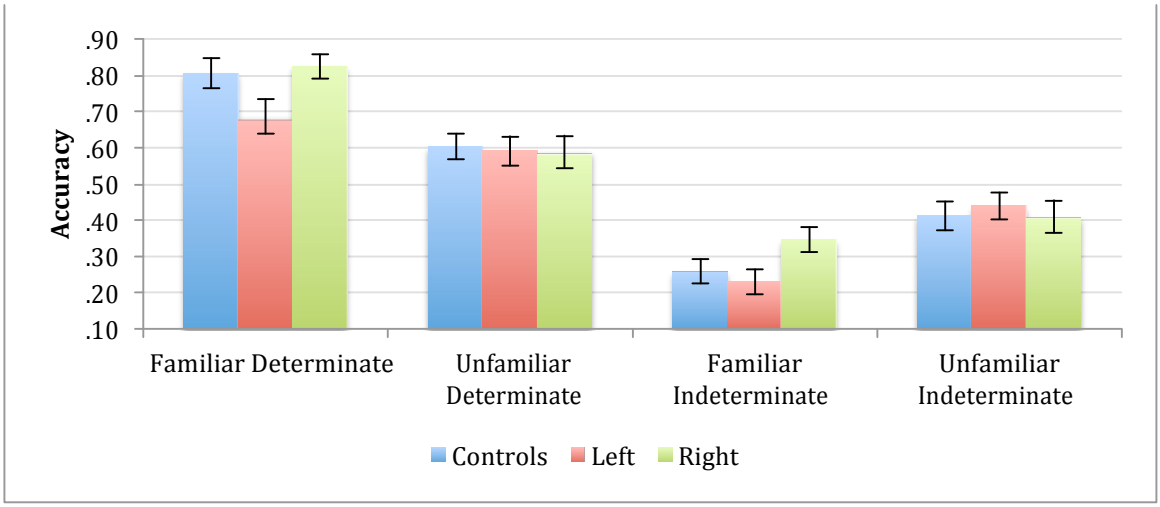


Figure 8. Performance on all familiar and unfamiliar determinate (MP and MT) and indeterminate (AC and DA) conditionals. Error bars represent the standard error of the mean.

Also to be noted, in the results concerning congruency and familiarity, the left hemisphere group performed worse than the controls and right hemisphere group, and since the left hemisphere group performed significantly worse than controls (there was no difference between the two patient groups) on the working memory measure, ANCOVAs were run to see if the effect would hold when it was added as a covariate. We retained the same pattern of significance when it was added as a covariate, showing that working memory could not account for the effect. Therefore, lesions in the left hemisphere drove this effect rather than working memory deficits. It is important to note that the right hemisphere group also had a significantly lower working memory than the controls, yet their performance was greater than the controls in these conditions (see Figures 4 and 5).

The Effect of Familiar Content

The result regarding the effect of content on reasoning replicates previous findings (to name a few, Goel & Dolan, 2003a; Goel & Dolan, 2003b; Goel, 2007; Houde et al., 2000; Knauff et al., 2002; Noveck, Goel, & Smith, 2004) that the left hemisphere is primarily responsible for reasoning with familiar content. As hypothesized, this study observed a significant interaction whereby left hemisphere lesion patients performed worse on familiar conditionals. This finding is in line with patient studies implicating the left hemisphere as being critical to logical reasoning and inferential processes (Caramazza, Gordon, Zurif, & Deluca, 1976; Gazzaniga and Smylie, 1984; Langdon & Warrington, 2000; Read, 1981; Whitaker, Markovits, Savary, Grou, & Braun, 1991).

For instance, in a study by Landgon and Warrington (2000), it was found that left hemisphere patients failed the verbal category (recognize an item that belongs to a different category from other items) and analogy (choose a word that has the same

relation to the test word as the sample pair) tasks whereas both left and right hemisphere patients failed the spatial reasoning task. They summarized this finding as evidence for the essential role of the left hemisphere in both verbal and spatial reasoning. In addition to this, Goel, Buchel, Frith, and Dolan (2000) point out that the largely left-lateralized network in these imaging and lesion studies seems to indicate that the left hemisphere is "necessary and often sufficient for logical reasoning, [while the] right hemisphere is sometimes necessary but never sufficient."

With regards to familiar versus unfamiliar content, in a review by Goel (2007), it was noted that the network for familiar, content-based reasoning consists of a left-lateralized frontal-temporal network, whereas a *bilateral* parietal visuospatial system (BA 7, 40) is responsible for reasoning with unfamiliar or content-free material. Upon closer inspection of the lesion patterns in our right hemisphere group, we only had 12 patients with damage to either BA 7 or 40, with only 9 patients with damage to both and only 5 having more than 2% of both BAs damaged. Thus, our sample may not have adequate power to detect the effect of damage to these structures in addition to the fact that only half the network was damaged. In addition to this, a couple of neuroimaging studies have shown that in the case of content-free conditionals presented in the same format as this study (in particular, modus ponens and modus tollens), that an exclusively left-lateralized network is recruited to successfully solve content-free conditionals (Noveck, Goel, & Smith, 2004; Prado, Van Der Henst, Noveck, 2010). Therefore, one might also expect to see impairments in the left hemisphere group on unfamiliar conditionals. But as with the right hemisphere group, there were only 6 patients with damage to BA 7, and 10 patients with damage to BA 40, and only 5 patients with damage to both, therefore, there may not

have been enough power to detect if damage to these areas affected reasoning with unfamiliar content. Therefore, having part of the bilateral network intact on either side seems to preserve the ability to reason with unfamiliar material, leading to the conclusion that perhaps bilateral damage is necessary to impair content-free reasoning. Thus it can be concluded that the network for reasoning with unfamiliar content has a built in redundancy in the sense that several parts need to breakdown in order for it to malfunction.

Belief-Logic Congruency

Regarding the effect of congruency on reasoning, our results confirm previous research regarding reasoning with belief congruent and incongruent content (Goel et. al., 2000; Goel & Dolan, 2003a; Rotello & Heit, 2014; Tsujii et. al., 2010) and provide additional evidence for the role of the left hemisphere in belief-laden reasoning. The results indicated that the left hemisphere patients did not benefit from the effect of belief-congruent content, they performed significantly worse than right hemisphere patients but there was no significant difference between left hemisphere patients and controls. And all three groups performed similarly on incongruent (belief-logic contradictory) trials. This finding is consistent with Tsujii et al.'s (2010) finding that TMS disruption to the right inferior PFC increased the accuracy on congruent trials while disruption to the left IPFC eliminated accuracy advantage for congruent trials, meaning accuracy across congruent and incongruent trials was about the same. Interestingly, in our findings, controls did not perform significantly better than left hemisphere patients, so the effect seems to be driven by the increase in accuracy rates seen in right hemisphere patients because they have an

intact left hemisphere with less inhibition from the right hemisphere on the left, which might help facilitate left hemisphere reasoning with congruent items.

Upon closer inspection of performance rates for congruent and incongruent items, it becomes clear that the right hemisphere group is primarily responding based on believability since accuracy rates go down for incongruent items, meaning they are biased towards their beliefs and answer inaccurately in the incongruent condition with the bias being more clear in the congruent condition where they go with their beliefs more often than left hemisphere patients and controls, while the left hemisphere group, analogous to left TMS inhibition/disruption, do not gain the accuracy advantage that congruent trials should facilitate (see Figure 5). As noted by Goel (2007), the role of the right hemisphere is belief-biased conditions is to suppress this bias mechanism, but because of damage to these suppression areas, they respond more strongly to the content of the arguments rather than the logical form.

The Effect of Emotional Content

Using VLSM whole brain analysis, emotionally salient content resulted in the significant correlation of reasoning deficits associated with lesions in the LMPFC. The results indicated that the LMPFC patients were impaired on emotionally salient conditionals compared to right hemisphere patients and controls. There were no significant differences between groups on non-emotional familiar or unfamiliar conditionals, thus the deficit was specific to emotional conditionals. This finding is consistent with previous neuroimaging studies that point to the importance of the Left Ventromedial PFC (VMPFC) in emotional reasoning (Goel & Dolan, 2003a; Smith, Vartanian, Goel, 2014). The finding that the LMPFC patients were impaired on reasoning

with emotionally salient conditionals is important because this is the first time conditional forms have been analyzed with emotionally salient material in both imaging and lesion studies. Thus this result seems to indicate that the LMPFC is critical to reasoning with emotional content above and beyond logical form (i.e. syllogisms, conjunctions, etc.)

Limitations

There were a few limitations in this study. First, our results mainly demonstrated that left hemisphere lesion patients were selectively impaired reasoning with content and congruent material. We were not able to find any group differences with respect to unfamiliar content, part of the reason for this may be that the areas necessary for reasoning with unfamiliar, abstract content depends on a bilateral network rather than the unilateral network examined in this study. Furthermore, as demonstrated in the discussion for familiar and unfamiliar content, there may not have been enough power to find clusters using VLSM pertaining to networks associated familiar, unfamiliar, congruent, and incongruent contrasts due to the fact we did not have enough patients with overlapping damage to areas associated with these contrasts. Lastly, there may be concerns regarding the low number of test items in each category (4-8), and the strength of inferences based on such low number. One statistical model that has been shown to be robust with for use on a limited range of data is the proportional odds logistic regression model (Parsons, Costa, Achten, and Stallard, 2009). With this model, a low number of test items can be used to draw reliable inferences with our group sizes (approximately 30 in this case). It works well with data that is not normally distributed or continuous as in the case of our data. Therefore, for more accurate parameter estimates given present data,

this regression model will be used to re-test all results, thus some of the *p*-values may change, but the pattern of significance and group differences should remain stable.

Conclusion

In summary, our findings point to the critical role of the left hemisphere for conditionals with content in general, and more specifically content with belief-congruent conclusions. Whereas content and belief-congruent conclusions was generally impacted by left hemisphere damage, the finding that emotional content is specific to the LMPFC supports research showing that this region is important for processing conditional arguments with emotional content, including during conditional reasoning. We provide additional evidence for the lateralization of reasoning processes, and in particular support the observation made by Goel et al. (2000) that the left hemisphere is "necessary and often sufficient for logical reasoning, [while the] right hemisphere is sometimes necessary but never sufficient," and the finding that the LMPFC plays a special role in the processing of emotional content (Goel & Dolan, 2003a).

References

Anderson, S. W., Damasio, H., Tranel, D., Damasio, A. R. (2000). Long-term sequelae of prefrontal cortex damage acquired in early childhood. *Developmental Neuropsychology, 18*(3) ,281-296.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289-300.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85(*1*),* 1-21.

Braine, M, & O'Brien, D. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Caramazza, A., Capitani, E., Arnaud, R., Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language, 76*, 158-184. Doi:10.1006/brln.1999.2275.

Caramazza, A., Gordon, J., Zurif, E. B., and Deluca, D. (1976). Right-hemispheric damage and verbal problem solving behavior. *Brain and Language*, *3*, 41-46.

Cohen, J. (1973). Eta squared and partial eta squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*, 107-112.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New Jersey: Lawrence Erlbaum Associates Publishers.

Ferreira, E. B, Rocha, M. C., Mequelino, D. B. (2012). Monte Carlo evaluation of the ANOVA's F and Kruskal-Wallis tests under binomial distribution. *SimgaE, 1*(1), 126-139, Retrieved from http://publicacoes.unifalmg.edu.br/revistas/index.php/sigmae/article/view/99.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics (4th Ed.)*. London: Sage Publications Ltd.

Gazzaniga M. S., and Smylie C. S. (1984). Dissociation of language and cognition. *Brain*, *107,* 145-153.

Grafman, J., Jonas, B. S., Martin, A., Salazar, A. M., Weingartner, H., Ludlow, C., Smutok, M. A, & Vance, S. C. (1988). Intellectual function following penetrating head injury in Vietnam veterans. *Brain, 111,* 169-184.

Goel, V. and Dolan, R. J. (2003a). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *NeuroImage*, *20*, 2314-2321.

Goel, V., and Dolan, R. J. (2003b). Explaining modulation of reasoning by belief. *Cognition, 87*(1)*,* B11-B22.

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

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences, 11*(10), 435-441.

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*, 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*, 2245–2250.

Hopkins, W. G. (1997). New view of statistics. Retrieved November 9, 2014 from http://www.sportsci.org/resource/stats/effectmag.html

Houde, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N., 2000. Shifting from the perceptual brain to the logical brain: the neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience*. *12*(5), 721–728.

Johnson-Laird, P.N., 2001. Mental models and deduction. *Trends in Cognitive Science*. *5*(10), 434–442.

Knauff, M., Mulack, T., Kassubek, J., Salih, H.R., & Greenlee, M.W., 2002. Spatial imagery in deductive reasoning: a functional MRI study. *Brain Research Cognitive Brain Research*. *13*(2), 203–212.

Langdon, D., & Warrington, E .K. (2000). The role of the left hemisphere in verbal and spatial reasoning tasks. *Cortex, 36*, 691–702

Lieberman, M. D., Cunningham, W. A. (2009). Type I and type II error concerns in fMRI research: Re-balancing the scale. *Scan, 4*, 423-428.

MATLAB Bioinformatics Toolbox Release 2014b. (2014). *Bioinformatics Toolbox: User's Guide*. Natick, Massachusetts: The MathWorks, Inc.

Makale, M., Solomon, J., Patronas, N. J., Danek, A., Butman, J. A., & Grafman, J. (2002). Quantification of brain lesions using interactive automated software. *Behavior Research Methods, Instruments, & Computers*, *34*(1), 6–18.

Monti, M. M., Osherson, D. N., Martinez, M. J., Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage, 37,* 1005-1016.

Noveck, I. A., Goel, V., Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex, 40*, 613-622.

Noveck I. A., & Politzer G. Leveling the playing field: Investigating competing claims concerning relative inference difficulty. In M. D. S. Braine, D. P. O'Brien (Ed), *Mental Logic*. Mahwah: Lawrence Erlbaum, 1998.

Oaksford, M., & Chater, N. (2010). Cognition and conditionals: An introduction. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 3-36). Cambridge: Cambridge University Press.

Parsons, N. R., Costa, M. L., Achten, J., Stallard, N. (2009). Repeated measures proportional odds logistic regression analysis of ordinal data in the statistical software package R. *Computational Statistics and Data Analysis, 53,* 632-641.

Parsons, L.M., and Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex*, *11*, 954-965.

Prado, J., Chadha, A., & Booth, J. (2011). The brain works for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *Journal of Cognitive Neuroscience*, *23*(11), 3483-3497.

Prado, J., & Noveck, I. A. (2007). Overcoming perceptual features in logical reasoning: A parametric functional magnetic resonance imaging study.

Prado, J., Van Der Henst, J., Noveck, I. A. (2010). Recomposing a fragmented literature: How conditional and relational arguments engage different neural systems for deductive reasoning. *NeuroImage, 51*, 1213-1221.

Raymont, V., Greathouse, A., Reding, K., Lipsky, R., Salazar, A., & Grafman, J. (2008). Demographic, structural and genetic predictors of late cognitive decline after penetrating head injury. *Brain*, *131*, 543-558.

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.

Read, D. E. (1981). Solving deductive-reasoning problems after unilateral temporal lobectomy. *Brain and Language*, *12*, 116-127.

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*, MA: MIT Press.

Rorden, C., & Karnath, H. (2004). Using human brain lesions to infer function: a relic from the past era in the fMRI age? *Nature Reviews Neuroscience, 5*, 813-819.

Rotello, C. M., & Heit, E. (2014). The neural correlates of belief bias: activation in infereior frontal coretex reflects response rate differences. *Frontiers in Human Neuroscience, 8,* 1-4.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is It Really Robust? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*(4), 147-151. doi:10.1027/1614-2241/a000016

Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: a meta-analytic case study in mental model and/or mental logic theory. *Thinking and Reasoning, 7*, 121-172.

Smith, S.M. (2002) Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3):143-155.

Smith, K. W., Vartanian, O., Goel, V. (2014). Dissociable neural systems underwrite logical reasoning in the context of induced emotions with positive and negative valence. *Frontiers in human neuroscience* doi:10.3389/fnhum.2014.00736.

Solomon, J., Raymont, V., Braun, A., Butman, J. A., & Grafman, J. (2007). User-friendly software for the analysis of brain lesions (ABLe). *Computer Methods and Programs in Biomedicine, 86*, 245-254.

Tzourio-Mazoyera, N., Landeaub, B., Papathanassioua, D., Crivelloa, F., Etarda, O., Delcroixa, N., Mazoyerc, B., & Joliota, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273-289.

Waechter, R. L., Goel, V., Raymont, V., Kruger, F., & Grafman, J. (2013). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia*, *51*, 464-471.

Whitaker, H. A., Markovits, H., Savary, F., Grou, C., & Braun, C. J. (1991). Inference deficits after brain damage. *Journal of Clinical and Experimental Neuropsychology, 13*, 38.

Woods, R.P., Cherry, S.R., Mazziotta, J.C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*,*16*, 620-633.

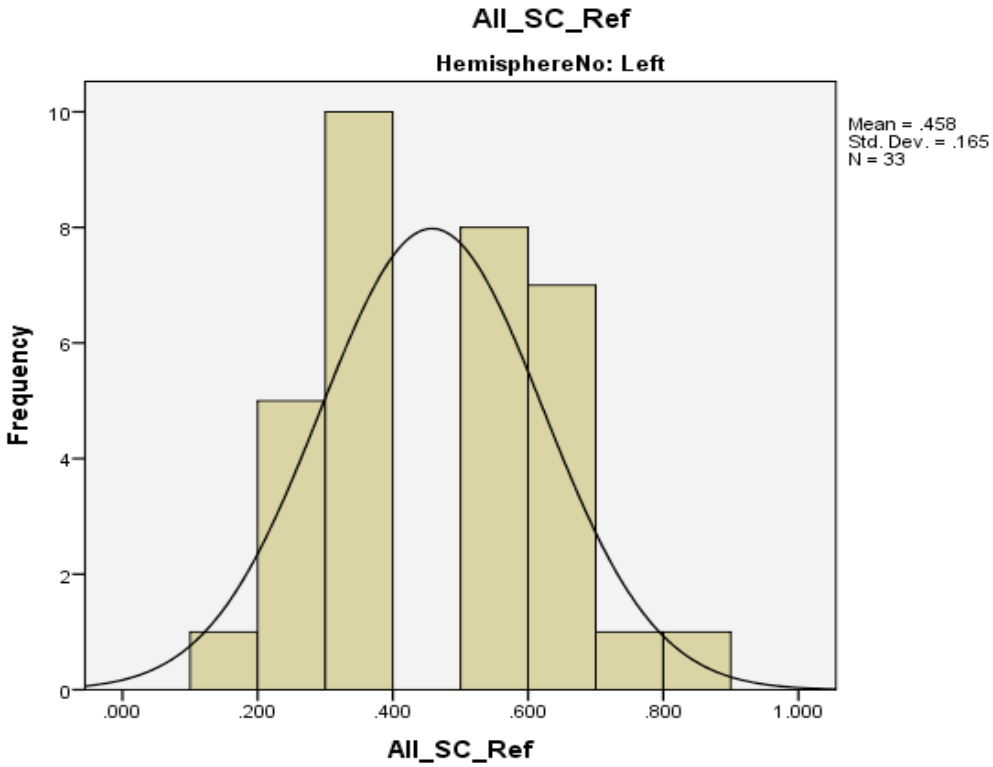
Tsujii, T., Masuda, S., Akiyama,T., and Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. *Neuropsychologia*, *48*, 2005–2008.

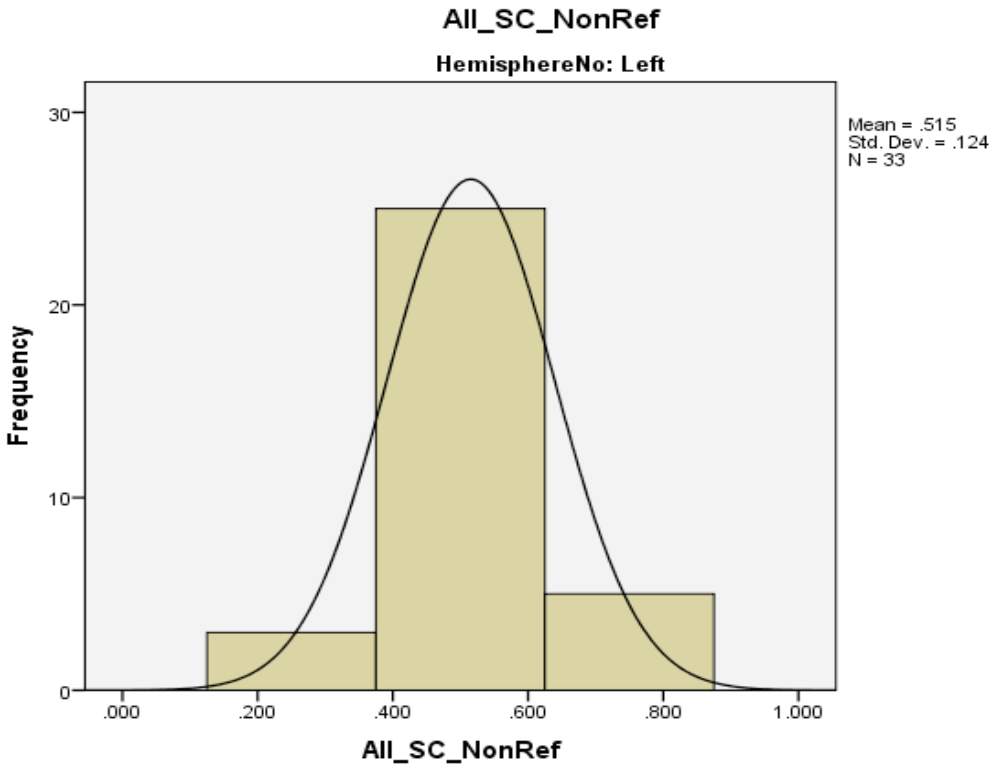
Appendix A. Assessment of Parametric Assumptions and Normality Plots

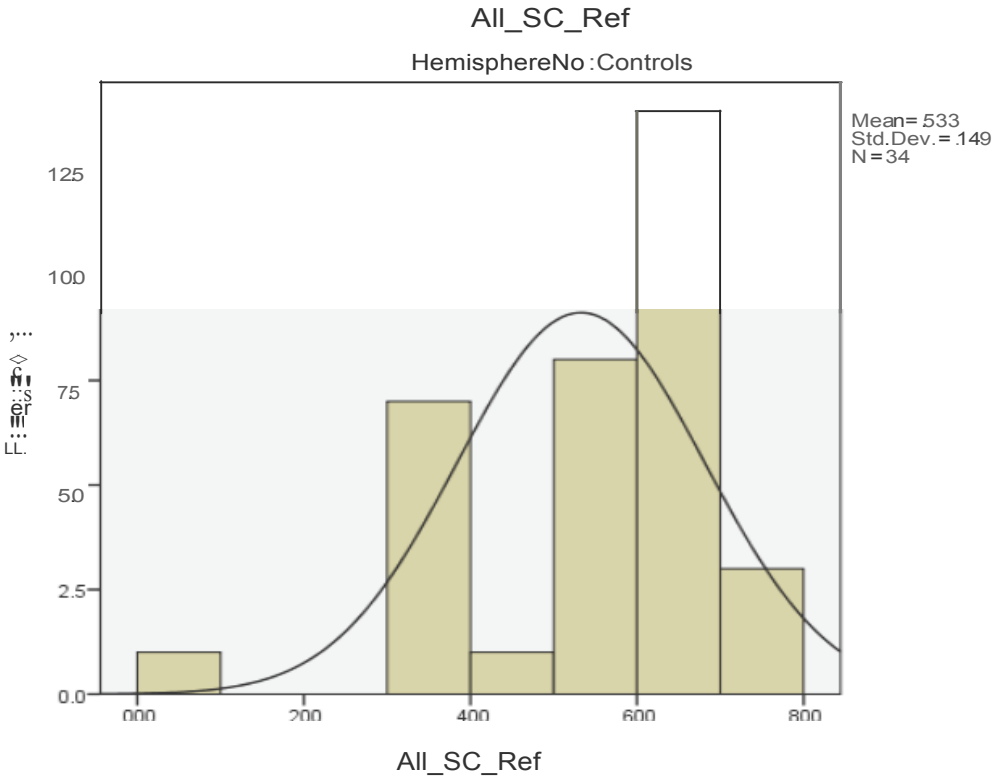
Group by familiarity (familiar, unfamiliar)

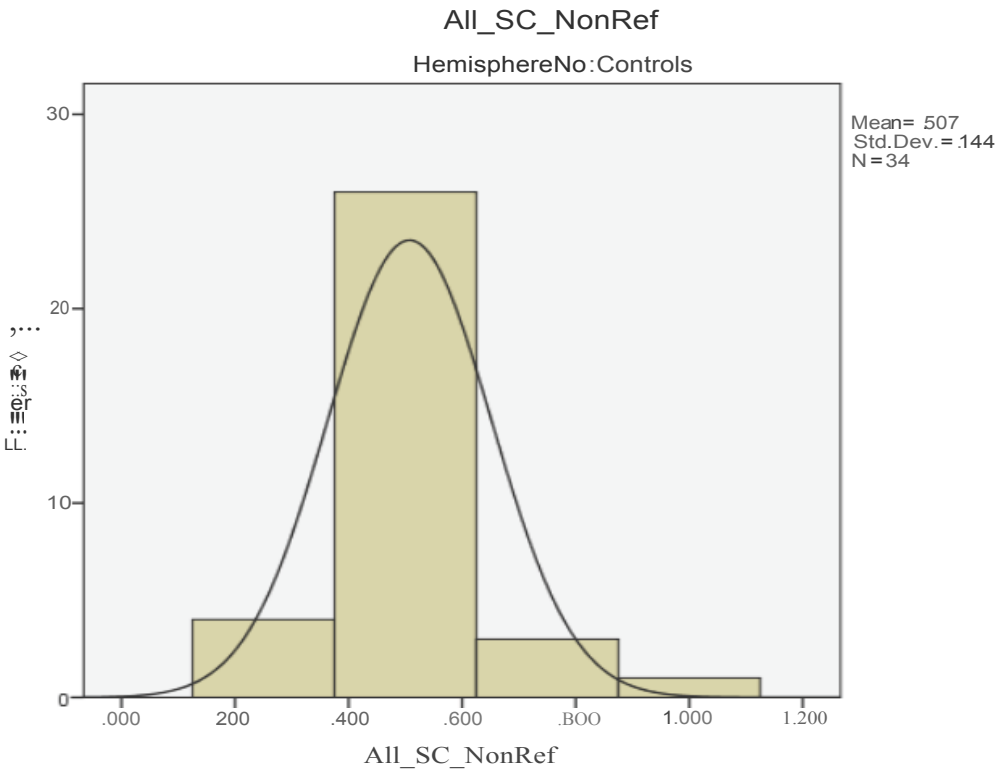
Normality was assessed by looking at the skewness and kurtosis of familiar and unfamiliar trials in each group, the scores of both patient groups was normally distributed as indicated by non-significant skewness ($p > .01$ or $z \leq \pm 2.58$); but the controls had significantly skewed ($z = -3.33$) and significant kurtosis ($z = 4.5$) on familiar scores, and significant kurtosis ($z = 5.10$) on unfamiliar scores, hence the control scores violated assumptions of normality. The only problem with kurtosis in the patient groups occurred with the right hemisphere scores on unfamiliar items ($z = 3.12$). Furthermore, normality was also assessed with Shapiro-Wilk tests of normality, which indicated that all score distributions were not normally distributed ($p < .01$) except for the distribution of left hemisphere familiar scores (see below for normality plots). Equality of variances was tested with Levene's test for homogeneity of variances, which showed equal variances across all groups in both conditions ($p > .5$).

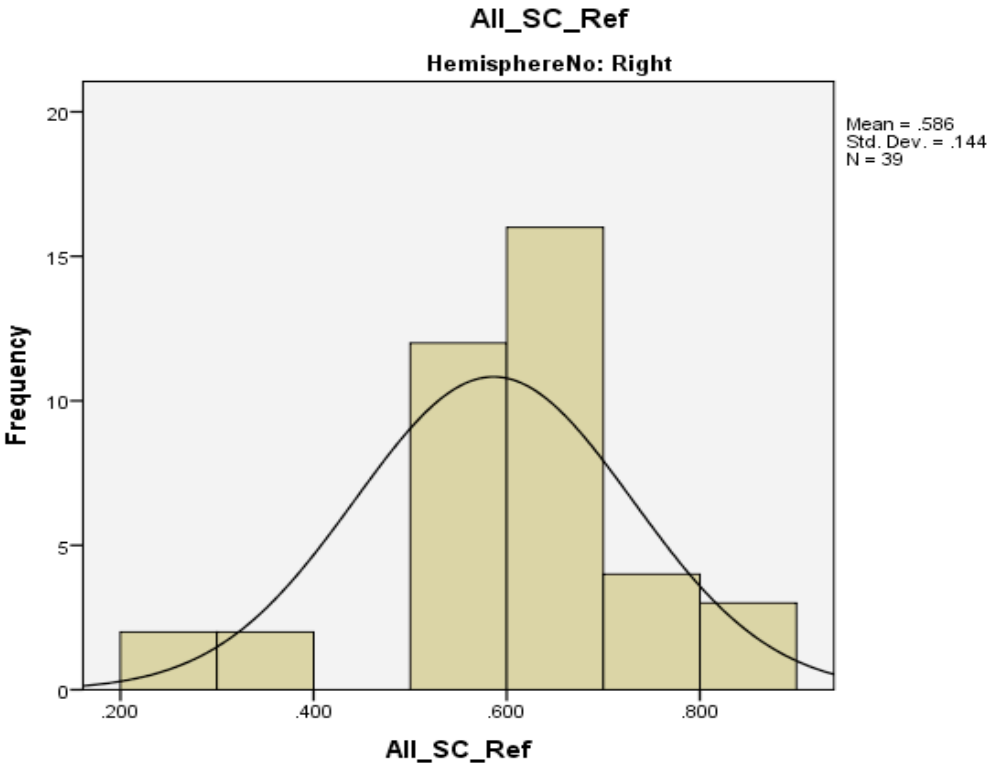
Content by Group Charts

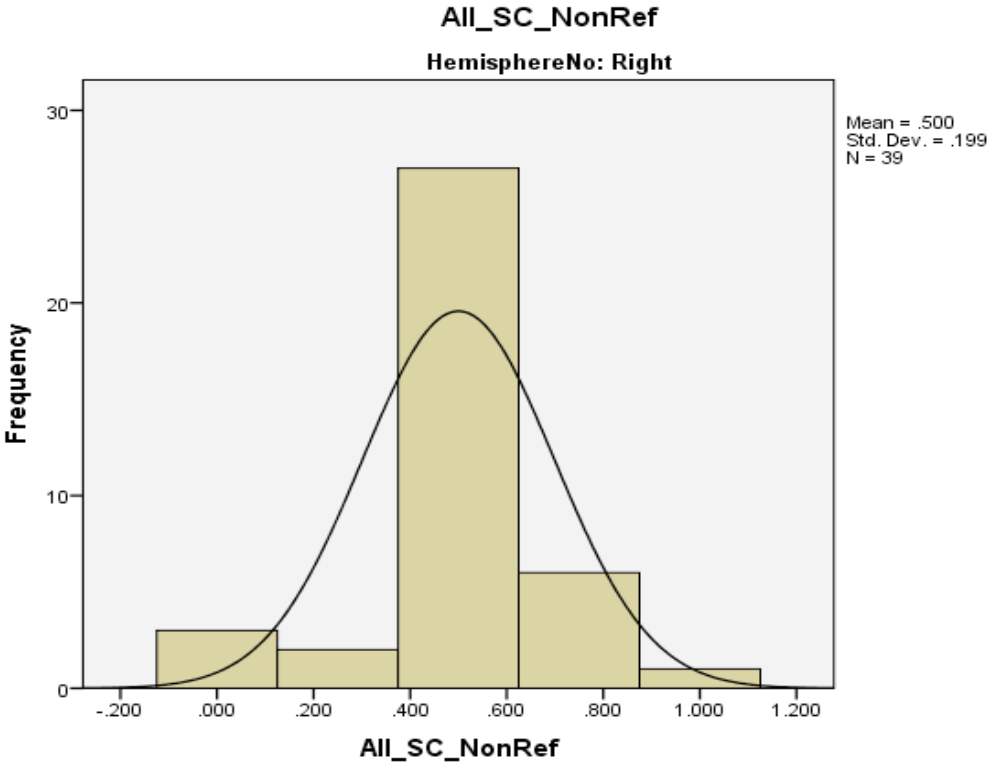





All_SC_Ref
HemisphereNo:Controls
Mean=.533
Std.Dev.=.149
N=34
125
100
75
50
2.5
0.0
000
200
400
600
800
All_SC_Ref


All_SC_NonRef
HemisphereNo:Controls
Mean=.507
Std.Dev.=.144
N=34
30
20
10
0
.000
200
.400
.600
.800
1.000
1.200
All_SC_NonRef

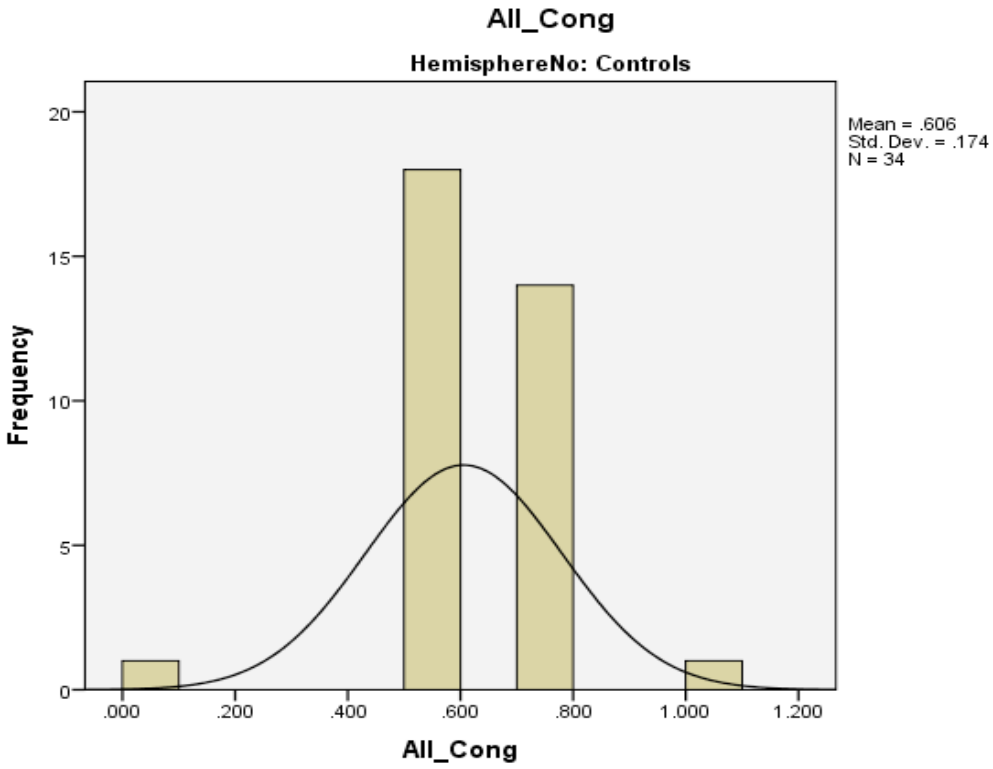

All_SC_Ref
HemisphereNo: Right
Mean = .586
Std. Dev. = .144
N = 39
Frequency
20
15
10
5
0
.200
.400
.600
.800
All_SC_Ref

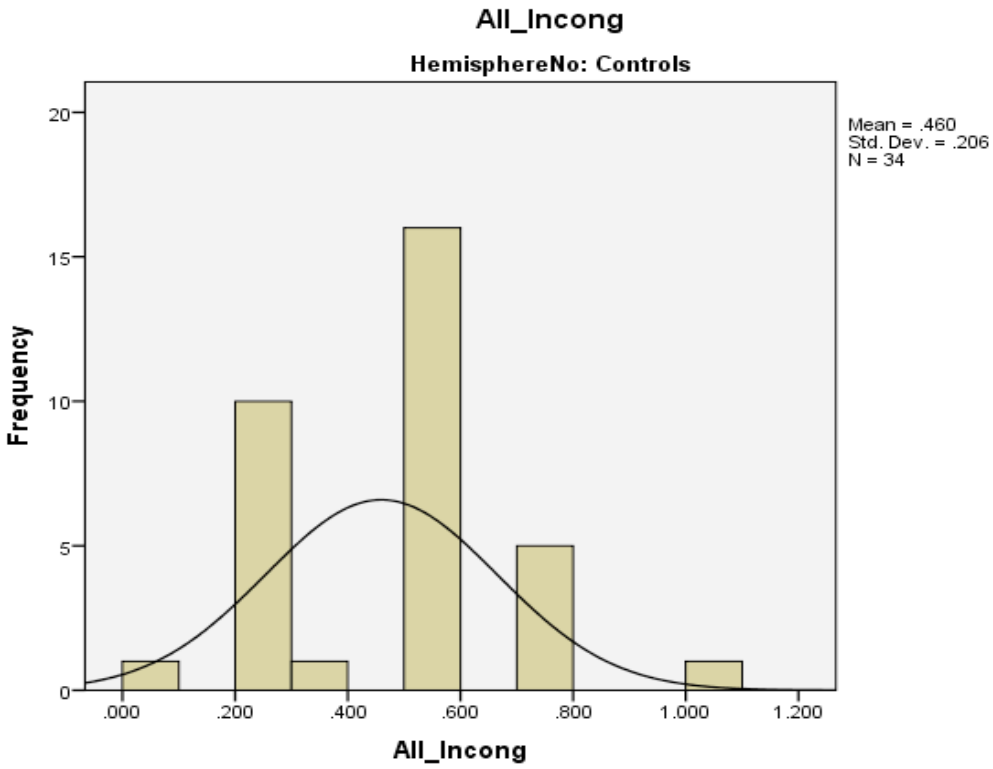

All_SC_NonRef
HemisphereNo: Right
Mean = .500
Std. Dev. = .199
N = 39
Frequency
30
20
10
0
-.200
.000
.200
.400
.600
.800
1.000
1.200
All_SC_NonRef

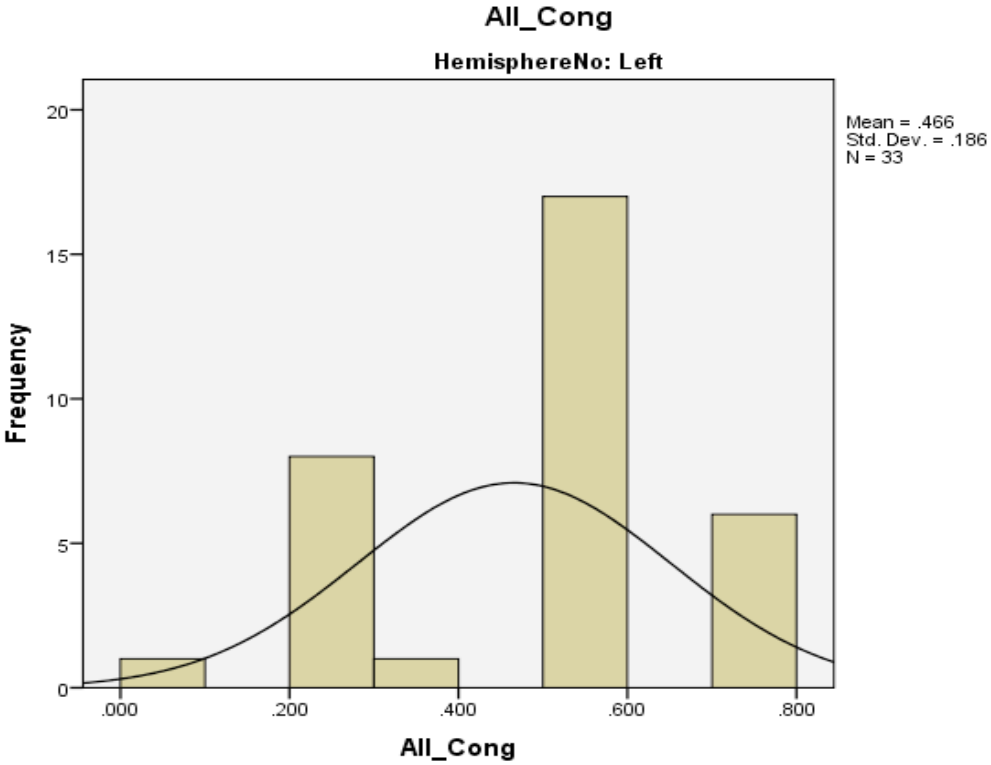
Group by congruency (all congruent vs. all incongruent)

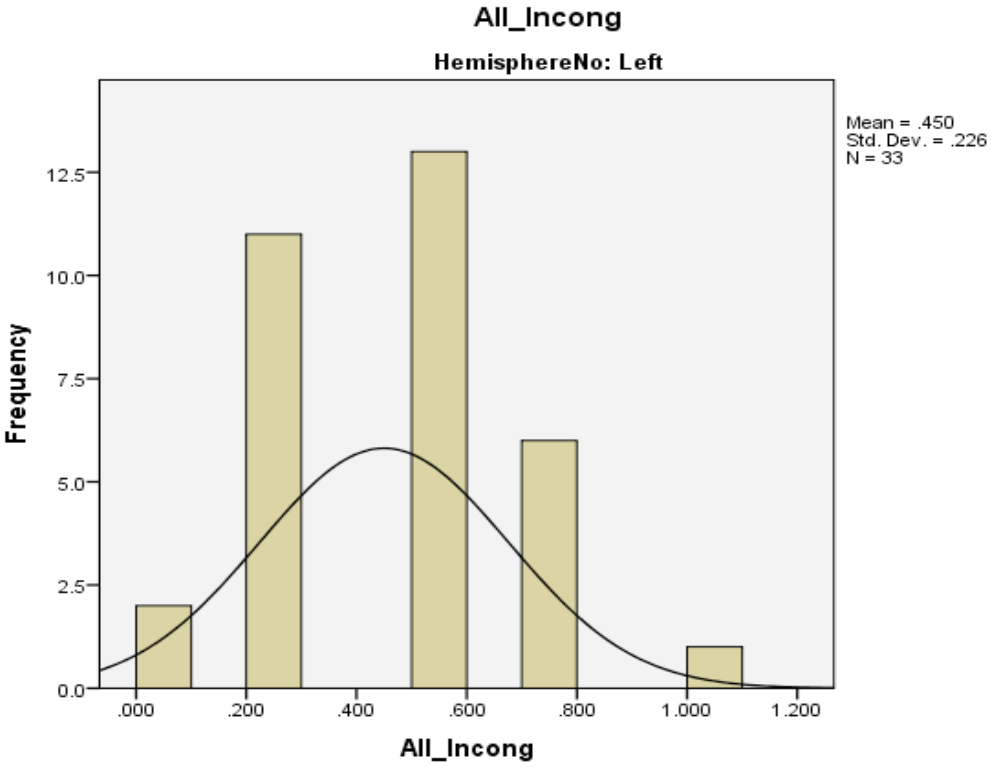
Normality was assessed by looking at the skewness and kurtosis in the distribution of all congruent and all incongruent trials for each group. The distribution of scores in each condition for both patient groups did not show any significant skewness or kurtosis ($p > .01$ or $z \leq \pm 2.58$). The only distribution of scores showing problems with normality was the controls scores in the congruent condition, with significant kurtosis ($z = 4.14$); the rest of control scores exhibited normal skewness and kurtosis ($p > .01$ or $z \leq \pm 2.58$). Furthermore, normality was also assessed with the Shapiro-Wilk tests of normality, which indicated that all score distributions (for each level and each group) were not normally distributed ($p < .01$). Equality of variances was tested with Levene's test for homogeneity of variances, which showed equal variances across all groups in both conditions ($p > .7$).

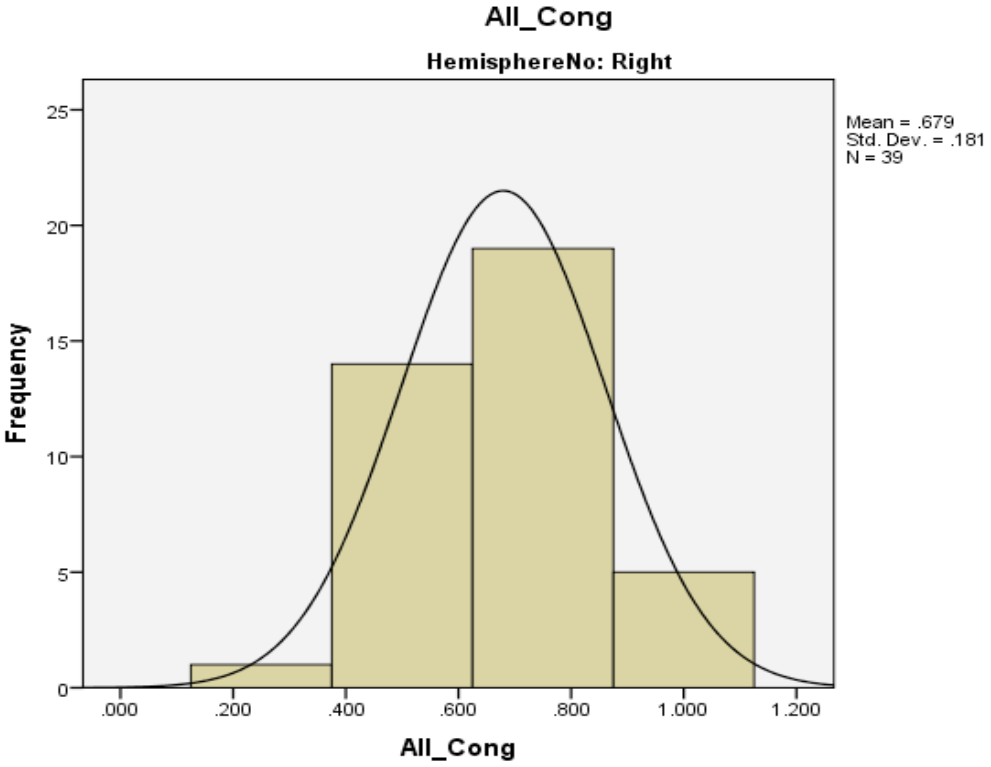
Congruency by Group Charts

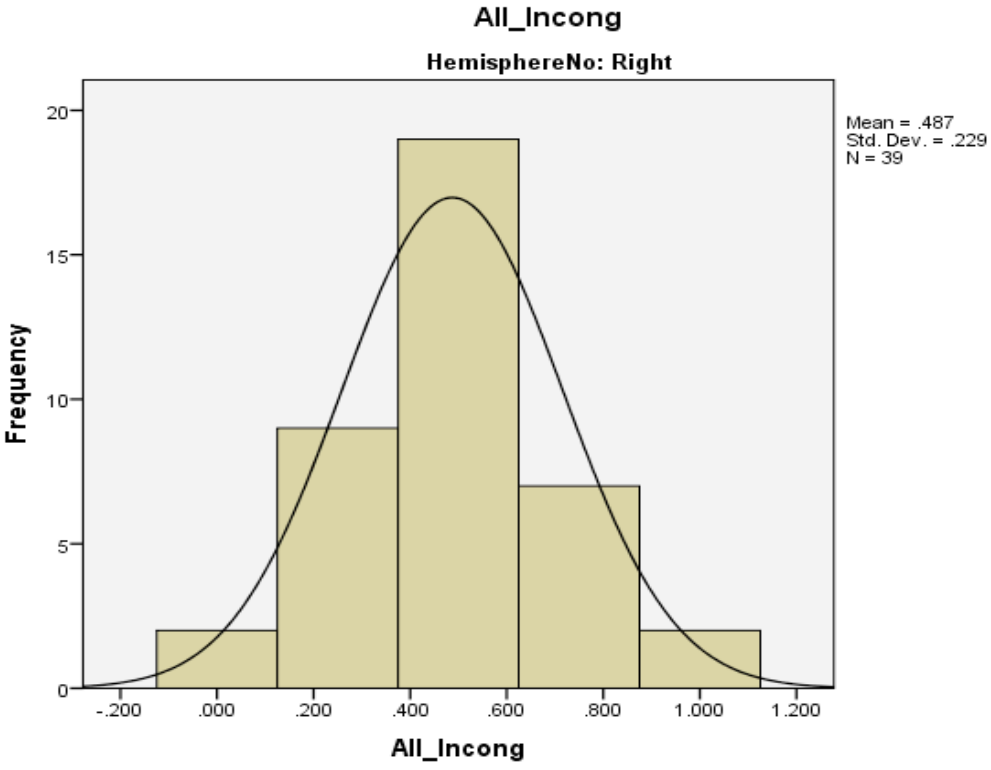





All_Cong
HemisphereNo: Left
Mean = .466
Std. Dev. = .186
N = 33
Frequency
All_Cong
0
5
10
15
20
.000
.200
.400
.600
.800


All_Incong
HemisphereNo: Left
Mean = .450
Std. Dev. = .226
N = 33
Frequency
All_Incong
0.0
2.5
5.0
7.5
10.0
12.5
.000
.200
.400
.600
.800
1.000
1.200


All_Cong
HemisphereNo: Right
Mean = .679
Std. Dev. = .181
N = 39
Frequency
25
20
15
10
5
0
.000
.200
.400
.600
.800
1.000
1.200
All_Cong

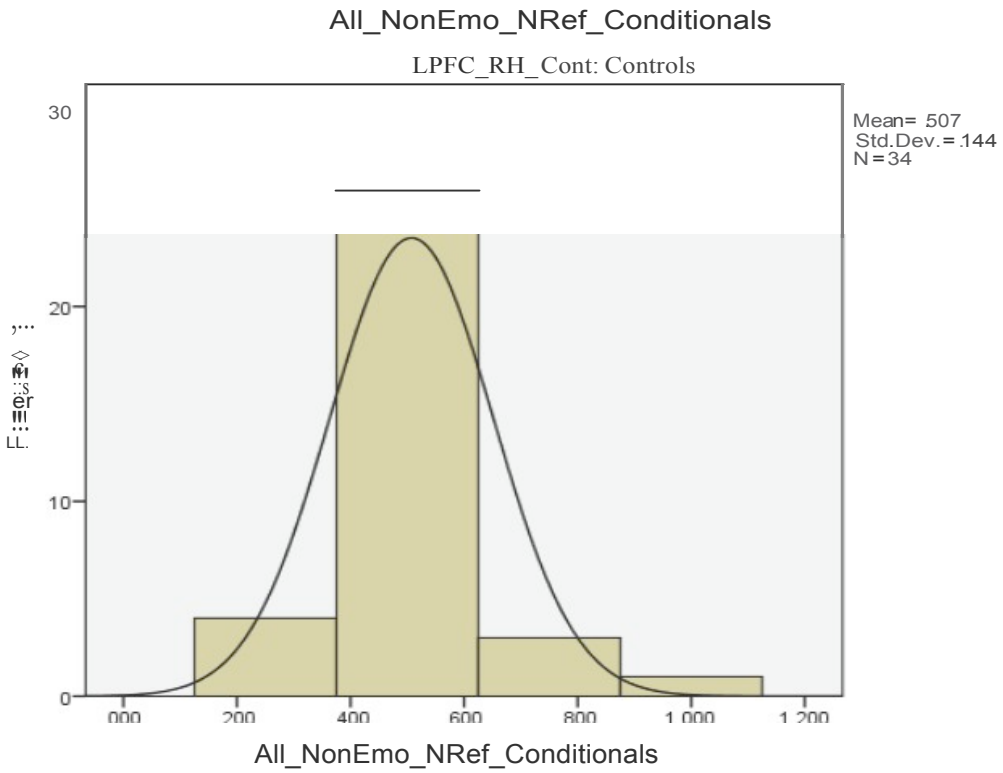

All_Incong
HemisphereNo: Right
Mean = .487
Std. Dev. = .229
N = 39
Frequency
20
15
10
5
0
-.200
.000
.200
.400
.600
.800
1.000
1.200
All_Incong

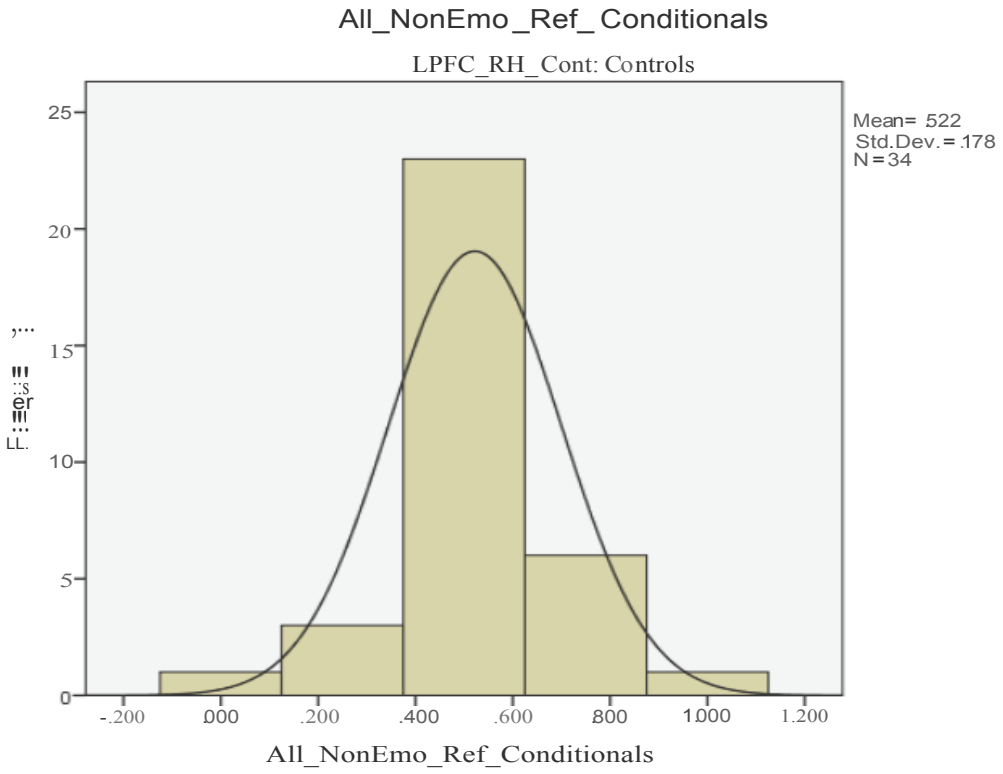
Group by emotion (emotional, non-emotional familiar, non-emotional unfamiliar)

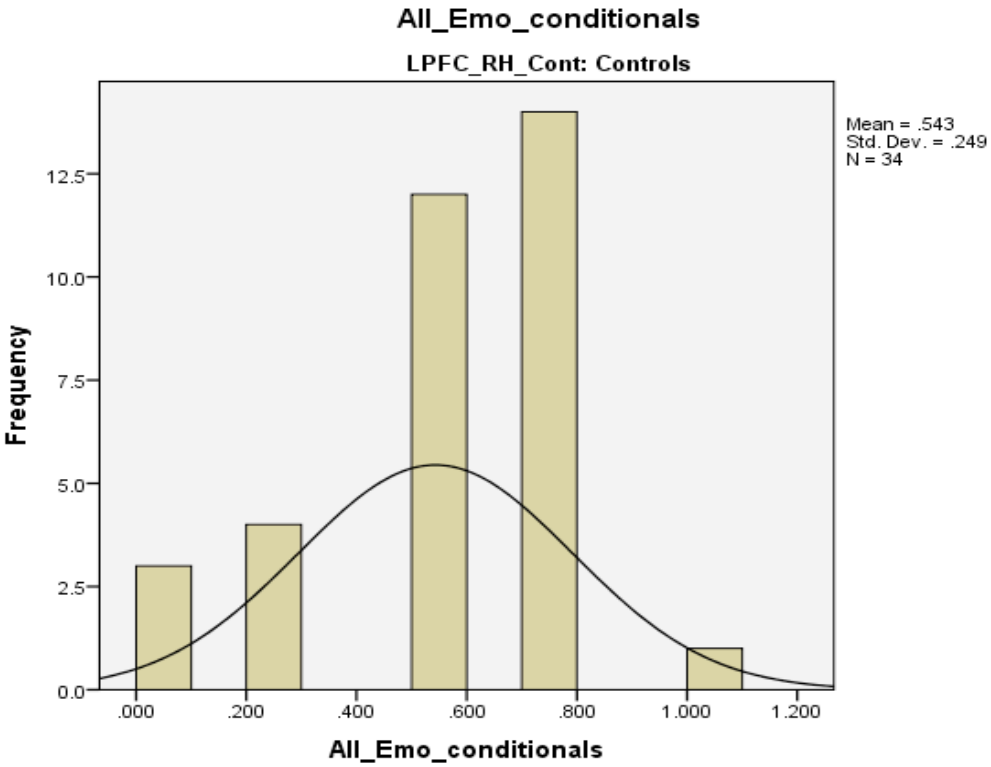
Normality was assessed by looking at the skewness and kurtosis in the distribution of all emotional, all non-emotional familiar, and all non-emotional unfamiliar scores for each group. The distribution of scores for emotional conditionals for all three groups did not show any significant skewness or kurtosis ($p > .01$ or $z \leq \pm 2.58$). For the unfamiliar scores, the only significant deviation from normality occurred in the kurtosis of controls scores ($z = 5.10$) and the kurtosis of right hemisphere scores ($z = 3.12$). Lastly, there was significant kurtosis in the distribution of familiar scores for controls ($z = 3.13$) and LPFC hemisphere patients ($z = 2.78$). Furthermore, normality was also assessed with the Shapiro-Wilk tests of normality, which indicated that all score distributions (for each level and each group) were not normally distributed ($p < .01$) except for the LPFC scores in emotional conditionals.

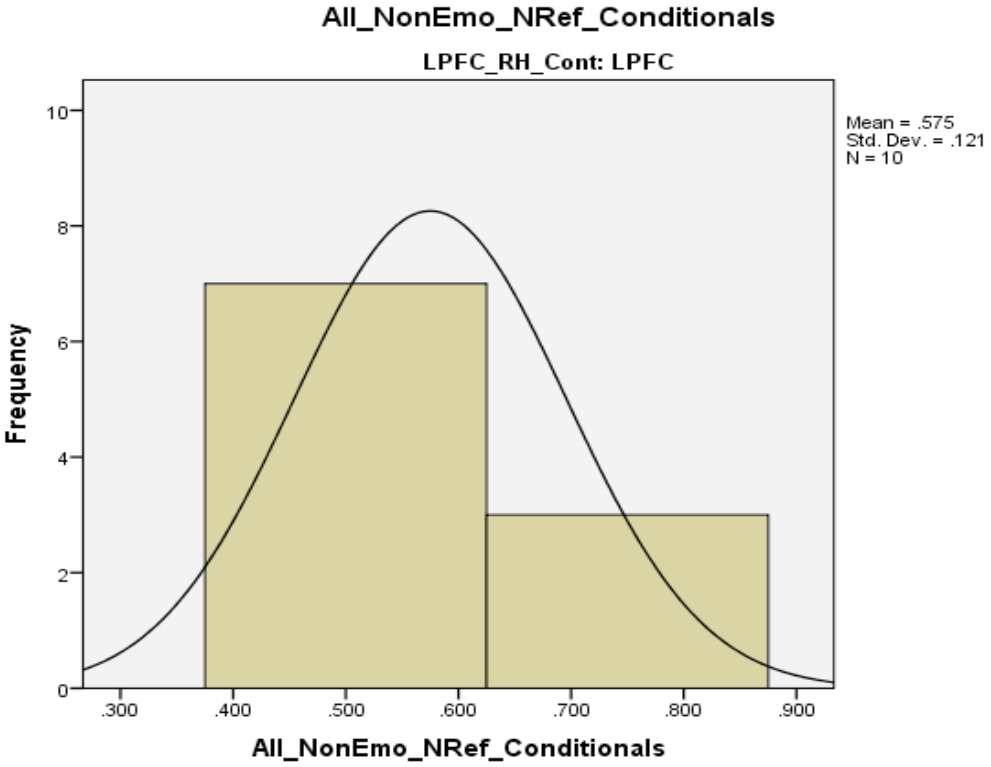
Equality of variances was tested with Levene's test for homogeneity of variances, which showed equal variances across all groups in all three conditions ($p > .15$). Also, since there were three levels of emotion, Mauchly's test for the assumption of sphericity was consulted, and was significant $\chi^2(2) = .917$, $p = .032$. Therefore degrees of freedom were corrected with the Greenhouse-Geisser (GG) estimates of sphericity ($\epsilon = .923$).

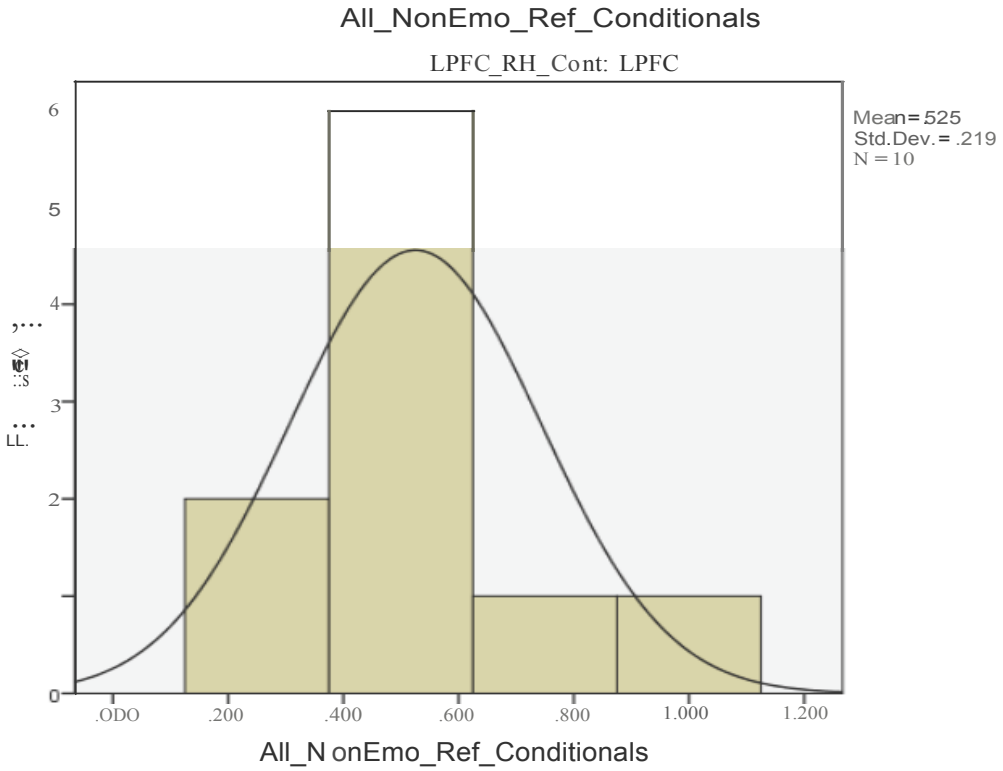
Emotion by Group Charts

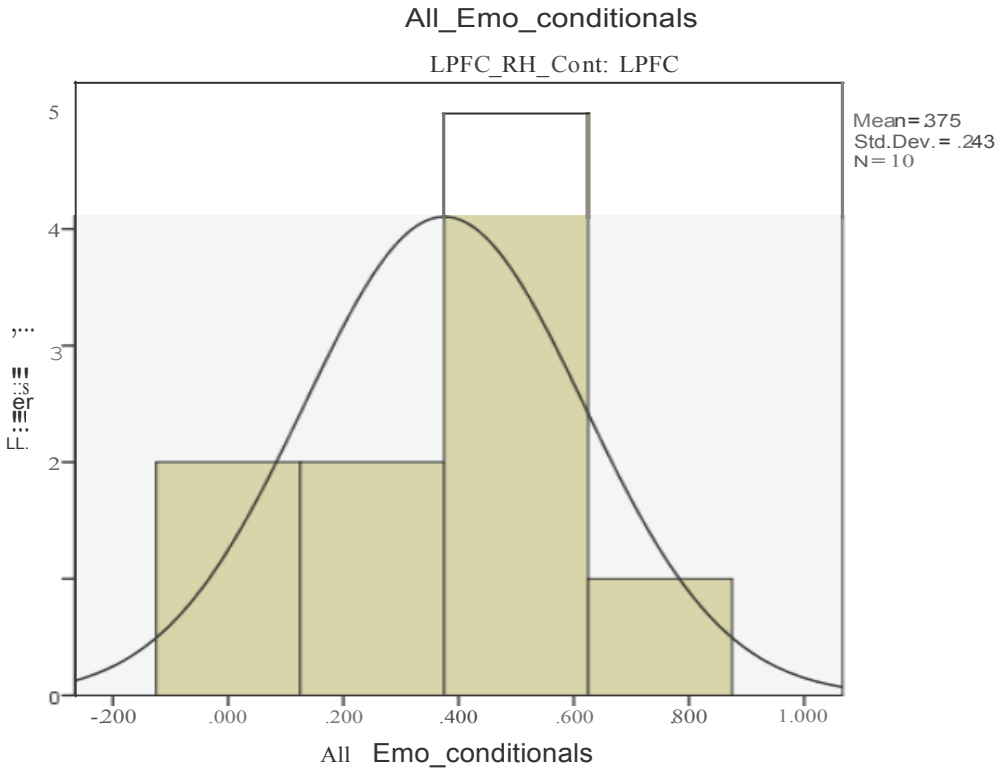

All_NonEmo_NRef_Conditionals
LPFC_RH_Cont: Controls
Mean= .507
Std.Dev.= .144
N = 34
30
20
10
0
000
200
400
600
800
1.000
1.200
All_NonEmo_NRef_Conditionals

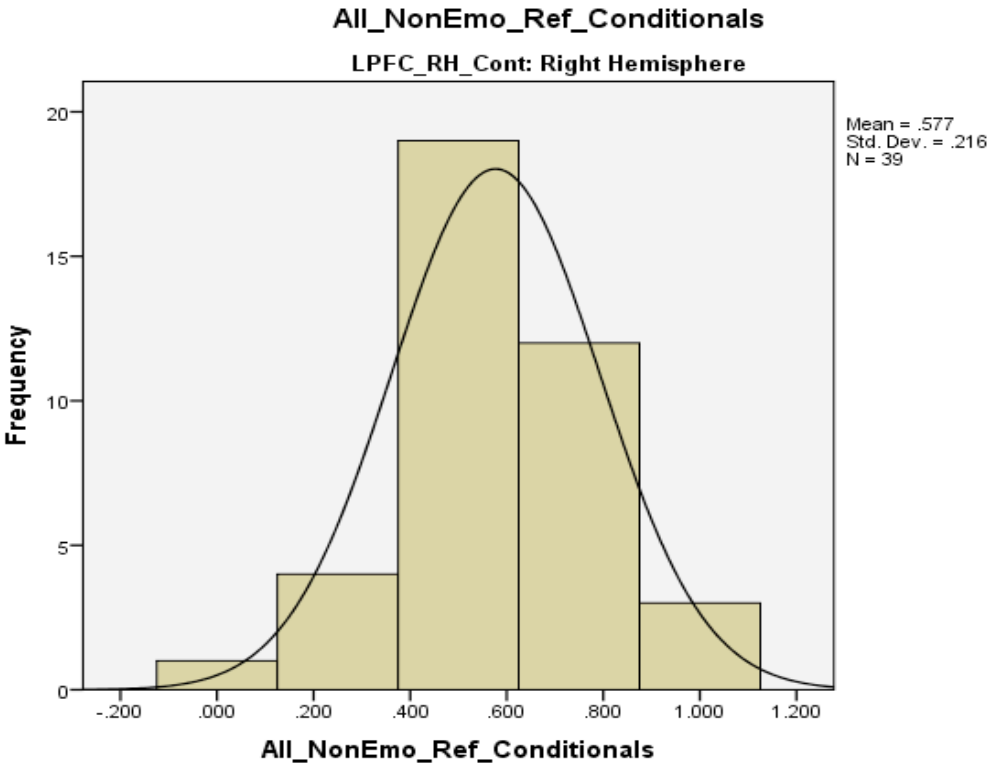

All_NonEmo _Ref_ Conditionals
LPFC_RH_Cont: Controls
Mean= .522
Std.Dev.= .178
N = 34
25
20
15
10
5
0
-.200
000
.200
.400
.600
800
1000
1.200
All_NonEmo_Ref_Conditionals

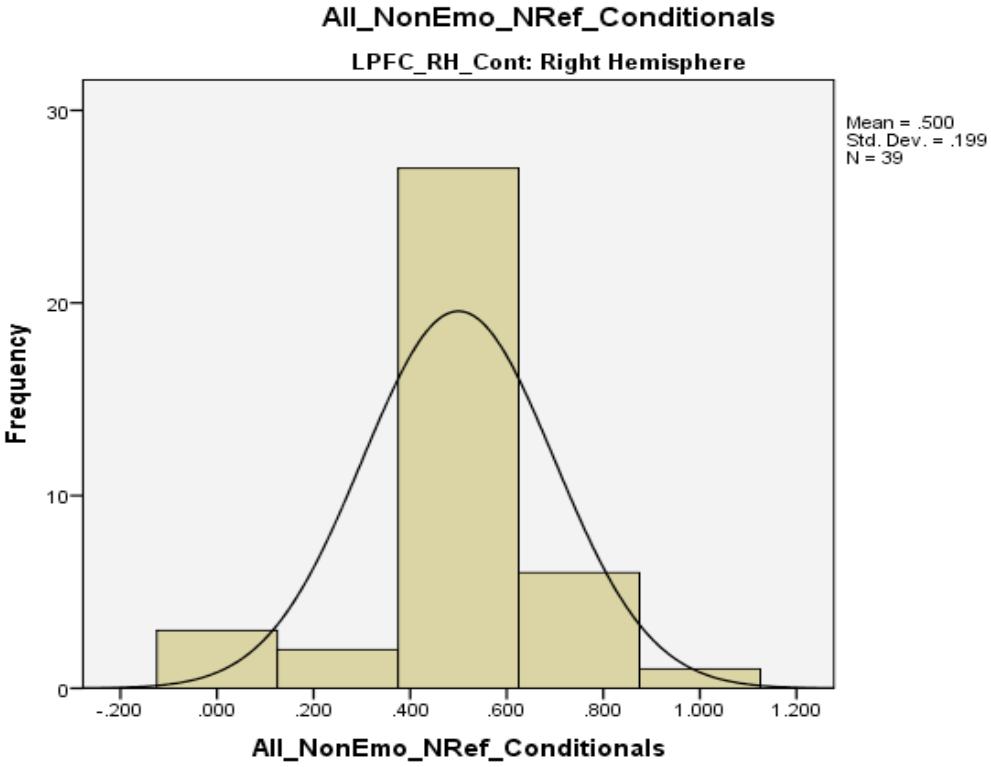

All_Emo_conditionals
LPFC_RH_Cont: Controls
Mean = .543
Std. Dev. = .249
N = 34
Frequency
12.5
10.0
7.5
5.0
2.5
0.0
.000
.200
.400
.600
.800
1.000
1.200
All_Emo_conditionals

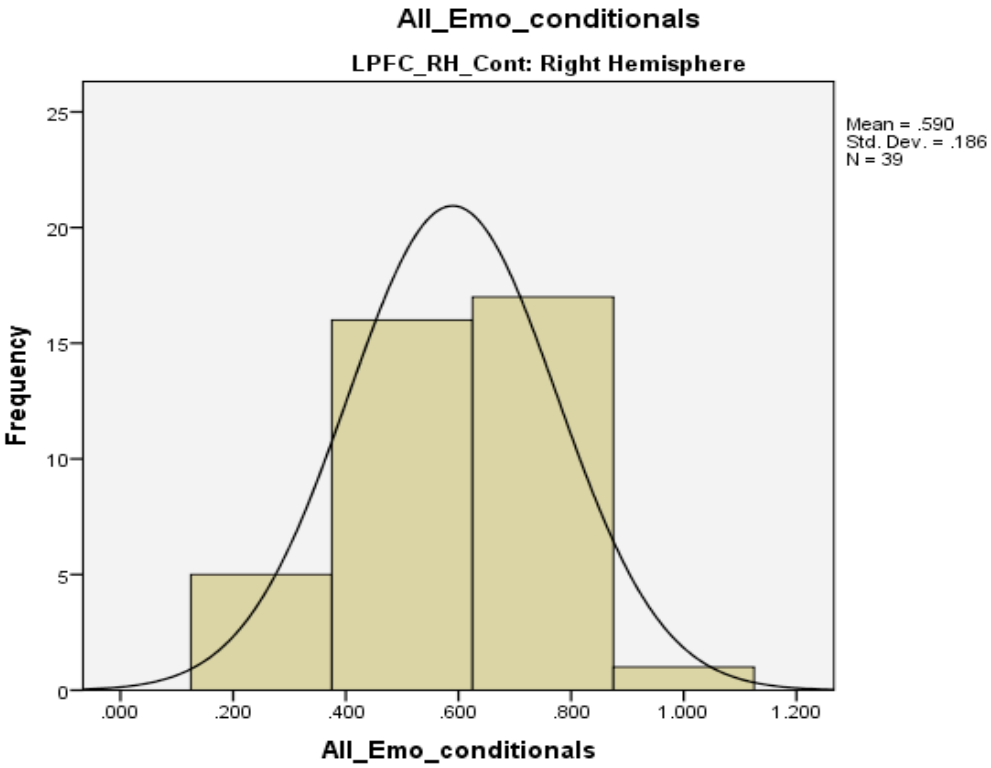

All_NonEmo_NRef_Conditionals
LPFC_RH_Cont: LPFC
Mean = .575
Std. Dev. = .121
N = 10
Frequency
10
8
6
4
2
0
.300
.400
.500
.600
.700
.800
.900
All_NonEmo_NRef_Conditionals


All_NonEmo_Ref_Conditionals
LPFC_RH_Cont: LPFC
Mean=.525
Std.Dev.=.219
N=10
6
5
4
3
2
0
.ODO
.200
.400
.600
.800
1.000
1.200
All_N onEmo_Ref_Conditionals


All_Emo_conditionals
LPFC_RH_Cont: LPFC
Mean=.375
Std.Dev.=.243
N=10
5
4
3
2
0
-.200
.000
.200
.400
.600
.800
1.000
All Emo_conditionals


All_NonEmo_Ref_Conditionals
LPFC_RH_Cont: Right Hemisphere
Mean = .577
Std. Dev. = .216
N = 39
Frequency
All_NonEmo_Ref_Conditionals


All_NonEmo_NRef_Conditionals
LPFC_RH_Cont: Right Hemisphere
Mean = .500
Std. Dev. = .199
N = 39
Frequency
All_NonEmo_NRef_Conditionals


All_Emo_conditionals
LPFC_RH_Cont: Right Hemisphere
Mean = .590
Std. Dev. = .186
N = 39
Frequency
25
20
15
10
5
0
.000
.200
.400
.600
.800
1.000
1.200
All_Emo_conditionals

Appendix B. Effect Sizes

Correlation coefficient (r): Recommended by Field (2013) to measure the effect size of comparisons. The following guidelines are adapted from Field (2013) and Hopkins (1997)

- r = .10 to .30 is considered a small effect, accounting for only 1% ($.1^2$*100) to 9% of the total experimental variance
- r = .30 to . 50 is considered a medium effect, accounting for 9% - 25% of the total variance
- r = .50 to .70 is considered a large effect, accounting for 25% - 49% of the total variance
- r = .70 to .90 is considered a very large/huge/high effect size
- r = .90 to 1.00 is considered nearly infinite

Eta squared (η^2): Cohen suggested the following guidelines (Cohen, 1988):

- .01 to .05 = small effect
- .06 to .14 = medium effect
- .15 and larger = large effect

Appendix C: Discussion of the Binomial Probability Distribution with Regards to "Chance" Performance

Regarding the accuracy of all three groups in the different conditional tasks, it would seem that all three groups performed at near chance levels on several of the tasks ($M_{accuracy}$ = 46% - 59%), therefore a question that arises is "how can we assess the performance of these groups when they seem to be systematically performing at chance on the familiar and unfamiliar conditional tasks?" If these groups seem to be responding at chance, then each individual in a group can be thought of as flipping a coin and using it to choose an input response (see Caramazza, Capitani, Rey, and Berndt (2001) for a similar analysis of chance performance of aphasic patients on a binary response comprehension task). In this scenario, it can be expected that coin flips would yield 50% heads and 50% tails after a number of tosses, however, after a finite series of coin tosses, the resulting distribution is unlikely to yield an exact 50% split. In the case of single patients or participants, it is hard to make inferences from a small number of trials (i.e. less than 20), but as the number of trials increases; inferences from single cases become more reliable (Caramazza et. al., 2001). According to the binomial theorem distribution, with approximately 4-12 binary problems as in this study, a single subject cannot be expected to be answering significantly above chance alone unless their performance exceeds that of 75% or falls below 25% (see Figure 7 for an illustration of such a binomial probability distribution). Therefore, it is not prudent to expect to draw reliable inferences from a single subject on only 8 items. But when the number of subjects increases, the probability of getting chance performance using a binomial probability distribution significantly decreases. In the case of our data, for unfamiliar conditionals with a total of 8 items, and 30 or more subjects per group, the binomial distribution does

not contain any probabilities of chance performance that are greater than an alpha of .05 (see Figure 8 for an illustration of this using the number of controls and the total number of response items they solved as a group to construct a binomial distribution), therefore, given an alpha of .05 or less, any group performance average, can be considered significantly above chance.

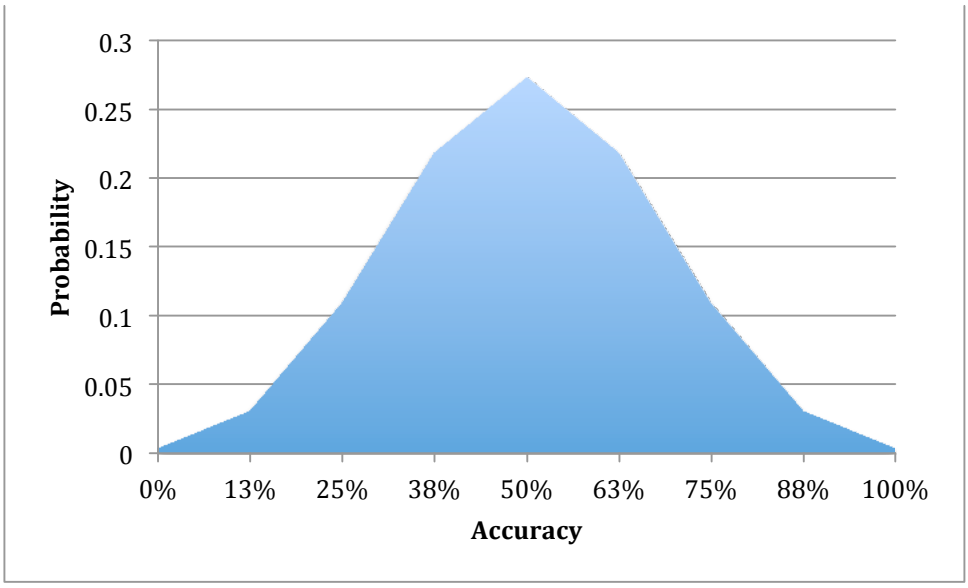


Figure 9. Binomial probability distribution for 8 items for a single subject. The x-axis represents accuracy across 8 trials, and the y-axis represents the binomial probability of this accuracy on a binary response task.

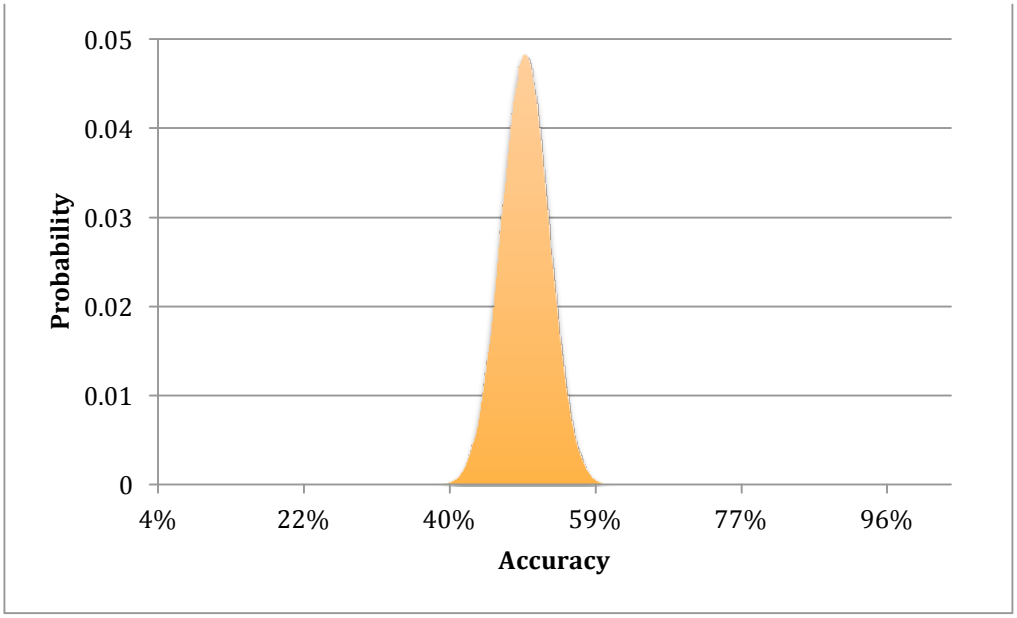


Figure 10. Binomial probability distribution for 8 unfamiliar items for the 34 controls. All possible group percentages fall below a .05 probability making it very unlikely that any response pattern is by chance alone. The x-axis represents accuracy across 8 trials, and the y-axis represents the binomial probability of this accuracy on a binary response task.

Thus, as illustrated by the discussion of differential performance on determinate and indeterminate items and the binomial probability distribution, the performance demonstrated in Figure 8 is significantly different from chance ($p < .05$), therefore, if chance cannot account for this performance pattern then we cannot dismiss it as a fluke. In the case of our data, most reasonable explanation is that the indeterminate forms pull overall performance rates towards "chance" levels.