

# **The Development and Assessment of the Semantic Fields Model of Visual Saliency.**

Benjamin Stone, B.Soc.Sci. (Hons.)

School of Psychology, The University of Adelaide

*Thesis submitted for the degree of*

*Doctor of Philosophy*



February 2010

## Contents

Signed Statement . . . . .	xviii
Acknowledgments . . . . .	xx
Dedication . . . . .	xxiii
Summary . . . . .	xxiv
<b>Chapter 1. Introduction and Literature Review</b>	<b>1</b>
1.0. Overview . . . . .	1
1.1. Display-based Web Page Search . . . . .	1
1.2. Semantic-based Web page Search . . . . .	6
1.3. Combining display-based and semantic-based information . . . . .	12
1.4. Introducing the Semantic Fields model . . . . .	12
1.5. Postlude . . . . .	16
<b>Chapter 2. Exegesis</b>	<b>18</b>
2.0. Overview . . . . .	18
2.1. Paper 1: Validating pupil dilation as a measure of cognitive load . . . . .	18
2.2. Paper 2: Using LSA Semantic Fields to estimate visual salience on Web pages . . . . .	22
2.3. Paper 3: Improving the semantic component of the Semantic Fields model . . . . .	25
2.4. Paper 4: Assessing the improved Semantic Fields model estimates of visual salience on Web pages . . . . .	30
2.5. Summary . . . . .	36
2.6. Further notes on papers . . . . .	36
<b>Chapter 3. Pupil Size and Mental Load (2004)</b>	<b>38</b>
<b>3.0. Abstract</b>	<b>39</b>

<b>3.1. Introduction</b>	<b>40</b>
3.1.1. The call for a measure of cognitive processing load . . . . .	40
3.1.2. Pupil dilation as a measure of cognitive load . . . . .	40
3.1.3. Pupil dilation as a measure of cognitive load in eye-tracking experiments that visually present stimuli on CRT monitors . . . . .	41
3.1.4. Experimental research hypotheses . . . . .	42
<b>3.2. Methodology</b>	<b>43</b>
3.2.1. Participants . . . . .	43
3.2.2. Apparatus & Procedure . . . . .	44
<b>3.3. Results</b>	<b>45</b>
3.3.1. Overall . . . . .	48
3.3.2. Rows . . . . .	48
3.3.3. Columns . . . . .	49
<b>3.4. Discussion</b>	<b>51</b>
<b>Chapter 4. Using LSA Semantic Fields to Predict Eye Movement on Web Pages</b>	
<b>(2007)</b>	<b>55</b>
<b>4.0. Abstract</b>	<b>56</b>
<b>4.1. Introduction</b>	<b>57</b>
4.1.1. Combining approaches . . . . .	57
4.1.2. Latent Semantic Analysis (LSA) . . . . .	58
4.1.3. LSA - Semantic Fields (LSA-SF) . . . . .	58
<b>4.2. Method</b>	<b>59</b>

4.2.1. Participants . . . . .	59
4.2.2. Apparatus . . . . .	60
4.2.3. Procedure . . . . .	61
4.2.4. Calculating the LSA Semantic Fields . . . . .	63
<b>4.3. Results</b>	<b>64</b>
4.3.1. Worst case to the best case scenarios . . . . .	65
4.3.2. Hyperlink-based LSA-SF . . . . .	66
4.3.3. All text LSA- SF . . . . .	67
<b>4.4. Discussion</b>	<b>68</b>
4.4.1. Other sources of heat . . . . .	68
4.4.2. The addition of rules . . . . .	68
4.4.3. Benefits of the LSA-SF to eye-tracking research . . . . .	69
4.4.4. Summary . . . . .	70
<b>Chapter 5. Comparing Methods for Single Paragraph Similarity Analysis (in press)</b>	<b>71</b>
<b>5.0. Abstract</b>	<b>72</b>
<b>5.1. Introduction</b>	<b>73</b>
5.1.1. Different types of textual language unit . . . . .	74
5.1.2. The dual focus of this paper . . . . .	78
<b>5.2. Semantic models, human datasets and domain-chosen corpora</b>	<b>80</b>
5.2.1. Semantic models . . . . .	80
5.2.2. The Datasets . . . . .	83
5.2.3. Domain-chosen corpora: WENN (2000-2006) & Toronto Star (2005) . . . . .	85

<b>5.3. Study One. Comparison of models on domain-chosen corpora</b>	<b>85</b>
5.3.1. WENN dataset & WENN Corpus . . . . .	86
5.3.2. Lee dataset & Toronto Star Corpus . . . . .	86
5.3.3. Summary of Study One . . . . .	89
<b>5.4. Study Two: Corpus Preprocessing</b>	<b>90</b>
5.4.1. Removing numbers & single letters . . . . .	91
<b>5.5. Study Three: A better knowledge base?</b>	<b>95</b>
5.5.1. Wikipedia Sub-corpora . . . . .	96
5.5.2. All models compared on Wikipedia sub-corpora . . . . .	101
<b>5.6. Study Four: Corpora that include the dataset paragraphs</b>	<b>104</b>
<b>5.7. Overall Summary</b>	<b>105</b>
<b>5.8. Discussion</b>	<b>112</b>
<b>Chapter 6. Semantic Models and Corpora Choice when using Semantic Fields to     Predict Eye Movement on Web pages (submitted)</b>	<b>117</b>
<b>6.0. Abstract</b>	<b>118</b>
<b>6.1. Introduction</b>	<b>119</b>
6.1.1. Semantic Fields (SF) . . . . .	120
6.1.2. Focus of this paper . . . . .	120
<b>6.2. Method</b>	<b>121</b>
6.2.1. Participants . . . . .	121
6.2.2. Apparatus . . . . .	122

6.2.3. Procedure . . . . .	123
6.2.4. Semantic Fields Models . . . . .	125
6.2.5. Corpora . . . . .	127
6.2.6. Baseline models to estimate eye-position . . . . .	129
<b>6.3. Results</b>	<b>132</b>
6.3.1. Did the participants complete their tasks successfully? . . . . .	132
6.3.2. Were the participants paying attention? . . . . .	133
6.3.3. Ten models compared using the Bayesian Information Criterion . . . . .	136
6.3.4. How well does the VEC-SF model using the WIKI-WEB corpus predict the eye data? . . . . .	138
<b>6.4. Discussion</b>	<b>140</b>
<b>6.5. Conclusions</b>	<b>142</b>
<b>Chapter 7. General Conclusion</b>	<b>143</b>
7.1. Final Statement . . . . .	148
<b>References</b>	<b>149</b>
<b>A. Paper 1: Statement of contributions</b>	<b>163</b>
<b>B. Paper 2: Statement of contributions</b>	<b>167</b>
<b>C. Paper 3: Statement of contributions</b>	<b>169</b>
<b>D. Paper 4: Statement of contributions</b>	<b>172</b>
<b>E. Appendices from Paper 3 (Chapter 5)</b>	<b>174</b>

E.1. Examples of similar and dissimilar paragraphs as rated by humans for the WENN dataset . . . . .	174
E.2. Examples of similar and dissimilar paragraphs as rated by humans for the Lee dataset . . . . .	176
E.3. Standard stop-list . . . . .	177
E.4. Corpora Parameters . . . . .	178
E.5. Study One results tables . . . . .	178
E.6. Stop-list used by Pincombe 2004 . . . . .	178
E.7. Study Two result tables . . . . .	180
E.8. IMDB-based Lucene query for Wikipedia . . . . .	181
E.9. Lee-based Lucene query for Wikipedia . . . . .	182
E.10. Study Three result tables . . . . .	183
E.11. Study Four results tables . . . . .	186
<b>F. Appendices from Paper 4 (Chapter 6)</b>	<b>188</b>
F.1. Goal pages with Semantic Field maps generated using Vectorspace and WIKI- WEB . . . . .	188
<b>G. Paper 1 - Original Article</b>	<b>197</b>
<b>H. Paper 2 - Original Article</b>	<b>204</b>
<b>I. Paper 3 - Original Article</b>	<b>211</b>
<b>J. Paper 3 - Supplementary Material file</b>	<b>272</b>
<b>K. Paper 4 - Original Article</b>	<b>301</b>

## List of Tables

2.1	Semantic Fields model with LSA and the TASA corpus, compared to Semantic Fields model with the semantic component held constant at one. The number of times higher Semantic Field values (all elements) were recorded for actual participant eye-points compared to eye-points generated in 1000 random trials. Best and Worst case calibration of eye-points are presented. . . . .	24
3.1	Descriptive statistics of the participants' Median Pupil Widths . . . . .	46
3.2	Approximate guide for converting Pixels x 10 into millimeters (mm). . . . .	47
3.3	Results of related samples t-tests used to compare average Median Pupil Widths recorded while participants were viewing the 'X' Stimulus in rows 0 to 4. . . . .	48
3.4	Results of related samples t-tests used to compare average Median Pupil Widths recorded while participants were viewing the 'X' Stimulus in columns 0 to 4. . . . .	49
4.1	Eye-based LSA-SF values compared to LSA-SF Overall Mean Values (OMVs) of the 1000 random trials associated with each page view. . . . .	65
6.1	Percentage of overlap between the expected landing Web page chosen by the experimenter and those chosen by the 49 participants. . . . .	133
6.2	Comparison of Bayesian Information Criteria (BIC) statistics calculated from log-likelihoods generated for all ten models. . . . .	137
E.4.1	Corpus parameters for the Toronto Star corpus, WENN corpus, and sub-corpora drawn from Wikipedia (1000 and 10000 documents) for both WENN and Lee datasets. . . . .	178



- E.5.1  $t$  values calculated using Williams' formula (T2) comparing within model correlations, where models have dimensionality or topics, with the human ratings contained in the WENN dataset. None of the models' performance significantly improved when dimensionality was increased (alpha 0.05). Significant decreases in performance would be indicated by negative values equal to or greater than 1.96. So, in no case was increased dimensionality associated with significant decrements to model performance. . . . . 179
- E.5.2  $t$  values calculated using Williams' formula (T2) comparing within model correlations, where models have dimensionality or topics, with human ratings contained in the Lee dataset. Significant decreases in performance would be indicated by negative values equal to or greater than 1.96. So, in no case was increased dimensionality associated with significant decrements to model performance. . . . . 179
- E.7.1 Correlations ( $r$ ) between similarity assessments of human raters and those made using LSA, Topic Model (Topics), Topic Model with Jensen-Shannon equation (Topics-JS), SpNMF at 50, 100, and 300 dimensions, and also the Overlap, Vectorspace and CSM models. The ALL columns display correlations based on corpora that contain both numbers and single letters (as used in Study One), conversely the NN-NSL columns are based on corpora with No Numbers and No single Letters (NN-NSL). Correlations exclude Same-Same document comparisons. . . . . 180

E.7.2 *t* values calculated using Williams’ formula (T2) comparing within model correlations, where models have dimensionality or topics, with human ratings contained in the WENN dataset used in Study Two. All corpora have had single letters and numbers removed. Significant decreases in performance would be indicated by negative values equal to or greater than 1.96. In no case was increased dimensionality associated with significant decrements to model performance. . . . . 181

E.7.3 *t* values calculated using Williams’ formula (T2) comparing within model correlations, where models have dimensionality or topics, with human ratings contained in the Lee dataset in Study Two. All corpora have had single letters and numbers removed. Significant decreases in performance would be indicated by negative values equal to or greater than 1.96. In no case was increased dimensionality associated with significant decrements to model performance. . . . . 181

E.10.1 Human to model correlations when estimating paragraph similarity on the WENN dataset, complex models using Wiki(pedia) 1000 & Wiki 10000 document corpora and the WENN Corpus (NN-NSL). Correlations exclude Same-Same paragraph comparisons. . . . . 184

E.10.2 Human to model correlations when estimating paragraph similarity on the Lee dataset, complex models using Wiki(pedia) 1000 & 10000 document corpora and the Toronto Star (NN-NSL) corpus. Correlations exclude Same-Same paragraph comparisons. . . . . 185

E.10.3 Examples of dimensions created by SpNMF on the 10000 document Wikipedia corpus generated for the Lee dataset where document length has been truncated at 100 words. . . . . 186

E.11.1 Comparison of models performance with standard Wikipedia 1000 corpora (Wiki 1000) and Wikipedia 1000 corpora including the 50 Lee paragraphs (Wiki 1050), using correlations between human and model estimates of paragraph similarity on the Lee dataset. Correlations exclude Same-Same paragraph comparisons. Significance tests were performed using Williams' T2 formula. . . . .	187
E.11.2 Comparison of models performance with standard Wikipedia 10000 corpora (Wiki 10000) and Wikipedia 10000 corpora including the 50 Lee paragraphs (Wiki 10050), using correlations between human and model estimates of paragraph similarity on the Lee dataset. Correlations exclude Same-Same paragraph comparisons. Significance tests were performed using Williams' T2 formula. . . . .	187

## List of Figures

1.1 Semantic Fields heat map of goal-oriented visual salience. Areas of greater estimated goal-oriented information salience have darker colors in this heat map.	13
2.1 Standardized pupil width during participants' fixations while they were performing 'add one' and 'subtract seven' mathematical tasks. Time spent on each task (2 minutes) has been delineated into deciles. . . . .	21
2.2 Graphic representation of the display-based models and the Semantic Fields model used in Paper 4. Areas of greater estimated goal-oriented information salience have darker colors in these heat maps. . . . .	32
2.3 Standardized pupil width during participants' fixations while they were performing goal-oriented Web page navigation. Data is for all three Web sites. Time spent searching each page is delineated into deciles. . . . .	34

2.4	Semantic Field values (Vectorspace with the Wikipedia sub-corpus) calculated for participant eye-points during goal-oriented Web page navigation. Data is for all three Web sites. Time spent searching each page is delineated into deciles.	35
3.1	The layout of the experimental room. . . . .	43
3.2	The visual stimulus is randomly moved to another cell in the experimental display every five seconds. . . . .	44
3.3	A Boxplot illustration of the differences between participants' overall average SUB7 and ADD1 Median Pupil Widths. . . . .	47
3.4	A Boxplot illustration of the differences between participants' average SUB7 and ADD1 Median Pupil Widths in each of the 5 rows in the experimental grid.	49
3.5	A Boxplot illustration of the differences between participants' average SUB7 and ADD1 Median Pupil Widths in each of the 5 columns in the experimental grid. . . . .	50
3.6	A histogram displaying the relative position of the outliers in the distribution of ADD1 in C0. . . . .	51
4.1	Example of a LSA-SF Map with a participant's eye data super-imposed using black dots. . . . .	63
4.2	Eye LSA-SF average minus the LSA-SF OMV for each page viewed by participants using the Link-based LSA-SF method (Best Case). . . . .	66
4.3	Eye LSA-SF average minus the LSA-SF OMV for each page viewed by participants using the All Text-based LSA-SF method (Best Case). . . . .	67

- 5.1 Correlations ( $r$ ) between the similarity ratings made on paragraphs in the WENN dataset by human raters and the those made by word overlap, LSA, Topics, Topics-JS (with Jensen-Shannon), SpNMF, Vectorspace, and CSM. All models, except word overlap used the WENN corpus. The effects of dimensionality reduction are displayed at 50, 100 and 300 dimensions for the more complex models that incorporate this reductive process. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 87
- 5.2 Correlations ( $r$ ) between the similarity ratings made on paragraphs in the Lee dataset by human raters and the those made by word overlap, LSA, Topics, Topics-JS (with Jensen-Shannon), SpNMF, Vectorspace, and CSM. All models, except word overlap used the Toronto Star corpus. The effects of dimensionality reduction are displayed at 50, 100 and 300 dimensions for the more complex models that incorporate this reductive process. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 88
- 5.3 Correlations between similarity estimates made by human and models on paragraphs in the WENN dataset. Models that employ a knowledge base used the WENN corpus. “ALL” depicts standard corpus preprocessing used in Study One, “NN-NSL” corpora have also had numbers and single letters removed. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 93

- 5.4 Correlations between similarity estimates made by human and models on paragraphs in the Lee dataset. Models that employ a knowledge base used the Toronto Star corpus. “ALL” depicts standard corpus preprocessing used in Study One, “NN-NSL” corpora have also had numbers and single letters removed. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 94
- 5.5 Correlations between human judgments of paragraph similarity on the WENN dataset with estimates made using LSA (at 300 dimensions) using the WENN Wikipedia-based corpora containing 1000 and 10000 documents retrieved using Lucene with WENN-based query. Wikipedia documents have been truncated in four ways: first 100, 200, 300, and ALL words. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 99
- 5.6 Correlations between human judgments of paragraph similarity on the Lee dataset with estimates made using LSA (at 300 dimensions) using Lee Wikipedia-based corpora containing 1000 and 10000 documents retrieved using Lucene with Lee-based query. Wikipedia documents have been truncated in four ways: first 100, 200, 300, and ALL words. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 100
- 5.7 Correlations between human judgments of paragraph similarity on the WENN dataset with semantic model estimates made using Wikipedia Corpora with 1000 & 10000 documents and the WENN Corpus (NN-NSL). Error bars are the 95% confidence limits of the correlation. These results are also presented in Table E.10.1. Correlations exclude Same-Same paragraph comparisons. . . . 102

5.8 Correlations between human judgments of paragraph similarity on the Lee dataset with semantic model estimates made using Wikipedia Corpora with 1000 & 10000 documents and the Toronto Star (NN-NSL). Error bars are the 95% confidence limits of the correlation. These results are also presented in Table E.10.2. Correlations exclude Same-Same paragraph comparisons. . . . . 103

5.9 Correlations between human and model estimates of paragraph similarity on the Lee dataset using the standard Wikipedia 1000 corpora (Wikipedia 1000) and Wikipedia 1000 corpora including the 50 Lee documents (Wikipedia 1050). The overlap model has also been included in this bar graph to allow the reader another point of comparison. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . 106

5.10 Correlations between human and model estimates of paragraph similarity on the Lee dataset using the standard Wikipedia 10000 corpora (Wikipedia 10000) and Wikipedia 10000 corpora including the 50 Lee documents (Wikipedia 10050). The overlap model has also been included in this bar graph to allow the reader another point of comparison. Error bars are the 95% confidence limits of the correlation. Correlations exclude Same-Same paragraph comparisons. . . . . 107

5.11 Scatterplots of the two best similarity estimates calculated for both the WENN and Lee datasets compared to the average similarity estimates made by humans for each pair of paragraphs. On the WENN dataset, (A) LSA using the WENN corpus (NN-NSL), and (B) the Overlap model. On the Lee dataset, (C) Vectorspace using the Wikipedia 1050 (including Lee documents), and (D) the Overlap model. Note, on the Lee dataset, average human ratings have been normalized [0,1]. . . . . 111

6.1 Semantic Fields Map using Vectorspace and a corpus drawn from Wikipedia. Participant’s eye tracking data is super imposed using black dots. While the original SF model only used LSA, the SF models presented in this paper incorporate word overlap, Vectorspace, LSA, and SpNMF semantic models. Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 121

6.2 Textual Web page elements are highlighted in red, images that have “ALT” or descriptive text are included. . . . . 131

6.3 Standardized pupil width during participants’ fixations while they were performing goal-oriented Web page navigation. Time spent searching each page is delineated into deciles. . . . . 134

6.4 Semantic Field values (Vectorspace with the WEB-WIKI corpus) calculated for participant eye-points during goal-oriented Web page navigation. Time spent searching each page is delineated into deciles. . . . . 139

F.1.1 Mission Australia - Task 1, “Who is currently the Chief Operating Officer of Mission Australia?” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 188

F.1.2 Mission Australia - Task 2, “You are interested in working for Mission Australia. Search their Web site for the current job vacancies available at Mission Australia.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 189

F.1.3 Mission Australia - Task 3, “You are currently researching homelessness in young people and have heard that Mission Australia has recently published a report called ‘The voices of homeless young Australians’. Search the Mission Australia Web site for this report into youth homelessness.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. 190



F.1.4 Green Corps - Task 1, “You want to know more about Green Corps management. Find out who is the National Program Manager of Green Corps.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 191

F.1.5 Green Corps - Task 2, “Find what environmental and heritage benefits are contributed by Green Corps.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 192

F.1.6 Green Corps - Task 3, “Find the online Expression of Interest form to apply to become a Green Corps Partner Agency.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 193

F.1.7 White Lion - Task 1, “Find out who is the current President of White Lion.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 194

F.1.8 White Lion - Task 2, “You are interested in becoming a mentor for young people. Find out how to become one of White Lions mentors.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. 195

F.1.9 White Lion - Task 3, “You are interested in financial viability of White Lion as a business. Find out which Government Departments are supporters of the White Lion organization.” Areas of greater estimated goal-oriented information salience have darker colors in this heat map. . . . . 196

*Signed Statement*

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Benjamin Stone and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the Web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through Web search engines, unless permission as been granted by the University to restrict access for a period of time.

Stone, B. & Dennis, S. (submitted). Semantic models and corpora choice when using semantic fields to predict eye movement on web pages. *International Journal of Human-Computer Studies*.

Stone, B., Dennis, S., & Kwantes, P. J. (in press). Comparing methods for paragraph similarity analysis. *Topics in Cognitive Science*.

Stone, B., & Dennis, S. (2007). Using LSA semantic elds to predict eye movement on web pages. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Society* (pp. 665-670). Mahwah, NJ: Lawrence Erlbaum Associates.

Stone, B., Lee, M., Dennis, S., & Nettelbeck, T. (2004). Pupil size and mental load. *Ist*

*Adelaide Mental Life Conference.* Available at:

<http://www.psychology.adelaide.edu.au/cognition/aml/> Accessed April 2, 2009.

SIGNED:

DATE:

## *Acknowledgments*

Seven years is a long time to do anything in life, and certainly it is more than enough time to finish a Ph.D. thesis. It might go some way to helping the reader understand why this process has taken so long, if I mention that I have had as many supervisors as I have spent years completing this research.

In the beginning I worked with Dr Lynn Ward, Professor Ted Nettelbeck and Dr Brett Bryan. Both Lynn and Ted were from the School of Psychology and Brett was working with Geography department. Much of my first year was spent learning how to write computer programs. As an undergraduate, I had my first experience with computers' four years prior to this date, and it took until second year for me to worked up the confidence to submit an assignment that was not hand written. So, I had a fair way to catch up to improve my skill levels in this area. I thank Lynn, Ted and Brett for indulging me in this pursuit, as it has become a skill that I now use in my day to day work life and this thesis has relied heavily on these skills.

Brett is a good programmer and Geographical Information Systems expert. In my honors year we had developed a pupillometer using a video camera, and so the next logical step was to create the eye-tracker. Unfortunately, as was to become a recurring theme during my Ph.D., people have a life of their own to live, and the world does not revolve around me. At the end of my first year of candidature, Brett was offered a job at the CSIRO<sup>1</sup> and was unable to continue supervising me in this project. While I wish him all the best for this change in direction, it did stifle my plans to develop the eye-tracker and the Ph.D. thesis needed to be re-thought.

After taking a year off to work and plan a new Ph.D. project, I developed an interest in human behavior in Web based environments. Professor Michael Lee volunteered to supervise me in this new project, and Ted was kind enough to stay on as my secondary supervisor. The

---

<sup>1</sup>Australia's Commonwealth Scientific and Industrial Research Organisation.

approach of my research changed under Michael's supervision. Instead of developing an eye-tracker, Michael simply bought one. My research could now focus on the task of modeling users' behavior whilst engaged in Web tasks. After a year spent working with Michael, the University of California managed to entice him away from Adelaide, and again I found myself without a supervisor. I also wish him all the best with his future pursuits.

Fortunately for me, Michael was replaced in the Psychology department by Dr Simon Dennis who was kind enough to take on the role as my principal supervisor. Simon really has been the driving force behind my academic development, and I will always be indebted to him for the friendship, guidance, and patience he has shown towards me. During the next two years in Adelaide, Simon helped develop my skills as both a programmer and a research scientist. However, as my story has already revealed, talented people are always in demand. Ohio State University offered Simon an Associate Professorship in 2007, and again I was without a primary supervisor in Adelaide. That said, Simon has continued to be my mentor in this research project as an external supervisor, and I thank him for sticking by me.

At the end of 2007, both Dr Dan Navarro and Professor John Dunn were kind enough to step into the roles of principal- and co-supervisors, respectively. I wish to thank them both for the support they have offered me over the last couple of years.

Many thanks go to Dr Peter Kwantes, who co-authored the third paper presented in this thesis with Simon and myself. Also many thanks go to the Defence Research & Development Canada (grant number – W7711-067985), who funded the research presented in the third paper. Furthermore, there have been numerous reviewers who have helped improve the four papers presented in this dissertation. I wish to extend my sincere appreciation for their helpful comments and suggestions, which have greatly improved the work present here.

So, seven years and as many supervisors on, I find myself at the end of this journey with a few others to thank. First, I would like to thank my wife Tegwen who is the most patient and understanding person I know. Next, my parents Barbara and Peter, who are both owed many

thanks for their constant encouragement and support. Finally, thank you to all my friends and colleagues who have had continued to encourage and motivate me towards the completion of this work.

*Dedication*

To those that I love most, Tegs, Mum, Dad, Chris and Del.

*Summary*

The present thesis describes the development and assessment of the Semantic Fields Model of visual salience. The Semantic Fields model provides estimates of visual salience in relation to goal-oriented Web site search tasks. The development and assessment of this model is reported over seven studies that are presented in two journal articles and two peer-reviewed conference papers.

In Paper 1 (N=50), pupil dilation is validated as a measure of cognitive load for use in later studies. While it has been found previously that a participant's pupil dilation will be larger during more complex tasks, these experiments have not generally been conducted under the environmental condition of light radiated from a computer monitor. The findings of this experiment indicate that computer monitor radiance in our experimental setting did not interfere with the ability to discriminate successfully between task-related pupil dilation.

Paper 2 (N=49) introduces the Semantic Fields model for estimating the visual salience of different areas displayed on a Web page. Latent Semantic Analysis and the Touchstone Applied Science Associates (TASA) corpus were used to calculate Semantic Field values for any (x, y) coordinate point on a Web page based on the structure of that Web page. These Semantic Field values were then used to estimate eye-tracking data that was collected from participants' goal-oriented search tasks on a total of 1842 Web pages. Semantic Field values were found to predict the participants' eye-tracking data.

In Paper 3 (N=100), four studies are present in which improvements are made to the semantic component of the Semantic Fields model. Estimates of textual similarity generated from six semantic models were compared to human ratings of paragraph similarity on two datasets. Results suggest that when single paragraphs are compared, simple non-reductive models (word overlap and vector space) can provide better similarity estimates than more complex models (Latent Semantic Analysis, Topic Model, Sparse Non-negative Matrix



Factorization, and the Constructed Semantics Model). Various methods of corpus creation were explored to facilitate the semantic models' similarity estimates. Removing numeric and single characters, and also truncating document length improved performance. Automated construction of smaller Wikipedia-based corpora proved to be very effective even improving upon the performance of corpora that had been chosen for the domain. Model performance was further improved by augmenting corpora with dataset stimulus paragraphs.

In Paper 4 (N=49), ten models are compared in their ability to predict eye-tracking data that was collected from participants' goal-oriented search tasks on a total of 1809 Web pages. Forming the basis of six of these models, three semantic models and two corpus types are compared as semantic components for the Semantic Fields model. Latent Semantic Analysis, Sparse Non-Negative Matrix Factorization, vector space, and word overlap were used to generate similarity comparisons of goal and Web page text in the semantic component of the Semantic Fields model. Vector space was consistently the best performing semantic model in this study. Two types of corpora or knowledge-bases were used to inform the semantic models, the well known TASA corpus and other corpora that were constructed from the Wikipedia encyclopedia. In all cases the Wikipedia corpora out performed the TASA corpora. The non-corpus based Semantic Fields model that incorporated word overlap performed more poorly at these tasks. Three display-based models were also included as a point of comparison to evaluate the effectiveness of the Semantic Fields models. In all cases the corpus-based Semantic Fields models outperformed the solely display-based models when predicting the participants' eye-tracking data. Both final destination pages and pupil data (dilation) indicated that participants' were actively performing goal-oriented search tasks.

Based on this research, it is concluded that the Semantic Fields model provided useful estimates of visual salience during participants' goal-oriented search of Web sites.