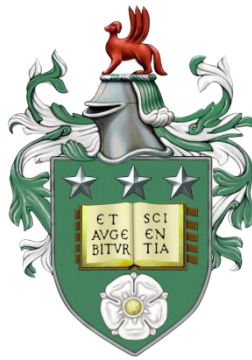


# **The role played by language in the interpretation of emotional facial expressions**

Emma Sally Portch



Submitted in accordance with the requirements of the degree of Doctor of Philosophy

The University of Leeds  
School of Psychology

January, 2016



## Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Various publications stem from research contained in this thesis. Each publication is referenced below, alongside details of its location within this thesis. In all publications where the candidate is the first author the candidate was solely responsible for the production of the content, with named authors providing support through review and modification only, unless otherwise stated.

*Portch, E., Brown, C., Havelka, J., Cook, H. & Simpson, R. (2013). Is the semantic satiation paradigm a useful method for investigating the role of language in the interpretation of facial expressions?*

Poster presentations at the Experimental Psychology Society (Lancaster, UK), the ESRC White Rose DTC Spring Conference (Leeds, UK) and the Faculty of Medicine and Health Conference (Leeds, UK), based on data presented in **Chapter 3 (experiment 2a)**.

*Portch, E., Brown, C., & Havelka, J. (2013). Shaping emotion: What role might emotion words play?*

Poster and oral presentations given at the Psychology Postgraduate Student Affairs Group (Lancaster, UK), the 18th Conference of the European Society for Cognitive Psychology (Budapest, Hungary) and the 'Language@Leeds' Interdisciplinary conference (Leeds, UK), based on data presented in **Chapter 3 (experiments 2a and 2b)**.

**Portch, E., Havelka, J., Brown, C. & Hutter, R. R. C. (2014/2015).** *Does language play a direct or indirect role in emotion perception?*

Oral presentations given at the Psychology Postgraduate Student Affairs Group (Cardiff, UK), the Experimental Psychology Society (Leeds, UK), and the International Society for Research on Emotion (Geneva, Switzerland), based on data presented in **Chapters 2, 3 and 4 (experiments 1, 2a, 2b, 3 and 4)**.

**Portch, E., Havelka, J., Brown, C. & Hutter, R. R. C. (2015).** *Emotion perception at threshold: Does language help?*

Poster presentations given at the Psychology Postgraduate Affairs Group (Glasgow, UK) and the 15<sup>th</sup> European Conference on Facial Expression, Measure and Meaning (Geneva, Switzerland), based on data presented in **Chapter 6**.

**Portch, E., Havelka, J., Brown, C. & Giner-Sorolla, R. (2015).** *Using affective knowledge to generate and validate a set of emotion-related action words.* *PeerJ* 3:e1100  
<https://doi.org/10.7717/peerj.1100>.

A subset of the co-authored work in this manuscript is included in **Chapter 7** (detailed author contributions for the manuscript version are provided there). However, the candidate confirms that the written content presented in this Chapter differs substantially from that of the manuscript, and places greater emphasis on how emotion-related action words were derived for immediate purposes (the experimental work reported in **Chapter 6**).

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

## Acknowledgements

First I would like to thank my supervisory team for their consistent support and advice. Each member has significantly contributed to my PhD, yet in very different ways. To Charity, for having a focused, details-oriented approach to all aspects of my work, for helping me to develop as a writer, and for providing me with the initial opportunity to come to Leeds. To Jelena, for always having a cleverly devised analogy to help me consider the 'wider picture' and for supplying countless cups of coffee. To Russ, for always providing helpful and detailed feedback on drafts and for your genuine concern regarding my declining mental health. I would also like to thank the various undergraduate students who helped collect data for this project: Hannah Cook and Rebecca Simpson (**experiment 2a, Chapter 3**); Dominic Manwani and Emma Butterworth (**experiments 1 and 3, Chapters 2 and 3**); Madeleine Steeds, Hannah Peacock and Laurence Rogers (**experiment 4, Chapter 4**).

When I started my PhD I thought that I would finally be surrounded by people just like myself; pale, exhausted introverts who had spent the last four years chained to their laptops. Fortunately, this was not the case. I feel privileged to have shared this experience with such a talented and inspirational group of people (Katy, Bex, Radka, Jess, Callum, Lara, Antonia and Dharsh). You've all taught me a great deal about the value of having a good social-support network and work-life balance. We've also managed to create numerous memories alongside our studies, including Mexican/Fav Fridays, Wednesday circuits, 'Circle Time' and our disastrous attempts at shared pet ownership (RIP batch one and two).

There remain a number of other staff and students at the School of Psychology who I would like to thank. I am grateful to Professor Anna Madill and Dr Amanda Waterman for providing advice when it was most needed and to Dr Gina Koutsopoulou, Julie Clark and Kathryn Meaney for making real efforts to welcome me to the department when I was still finding my feet. Kremena and Suzi were always on hand to advise and support me through the next hurdle and George and Ed accompanied me on my (innumerable) cigarette breaks, whilst providing many other welcome distractions.

Special thanks are reserved for Helen, my longest-standing office mate. I appreciate the many things you've done for me, which include, but are not limited to: taking me out for daily coffee dates, being available day and night for PhD-related moaning, keeping me company on late nights and weekends, feeding and clothing me, and decorating my desk (a special mention to Katy here for her exceptional photo-shopping skills – who knew that they would translate so well to thesis formatting!). I would also like to thank Helen's wonderful parents, who provided well-timed words of encouragement and gifts during the write-up period. I really could not have produced this thesis without you and I cannot express how much I will miss you.

Last, I am grateful to those who have loved and supported me throughout my time in Higher Education; my mother (Noonie), my partner (Andrew), and my best friend (Anna). For maintaining my carefully-calibrated diet of caffeine and nicotine, for picking up the slack when I had to work long hours, for treading on eggshells when a big deadline approached and for 'talking me down' when it (frequently) all got too much. I can only imagine the sense of relief you must feel now that my days as a student are finally numbered.

I end with the first PhD-related advice I received from a very dear, wise friend:

*'Doing a PhD is like sitting in a room with the same person for 3+ years. You might like them well enough to begin with, but you won't be able to stand the sight of them by the end.'*

*Kountouriotis, G. (2011)*

## Abstract

This thesis examines the role played by language in the interpretation of emotional expression. Language labels may *indirectly* influence such tasks, organising and reactivating a useful repository of semantic knowledge (e.g. Barrett, Lindquist & Gendron, 2007). This proposal was explored using a series of semantic satiation experiments (Lindquist, Barrett, Bliss-Moreau & Russell, 2006). Participants repeated words 3 or 30 times before deciding whether two faces matched in emotional expression. Word type was manipulated across experiments (emotion labels, neutral labels and non-words); an *indirect* account would only predict reduced accuracy when participants experience semantic inaccessibility, achieved via massed repetition of an emotional label. However, reduced discrimination was found both after 30 (vs. 3) repetitions of any word, and two non-linguistic activities. Findings then suggest that the massed repetition decrement arises via a non-semantic mechanism, such as response uncertainty (e.g. Tian & Huber, 2010). However, an emotion-specific effect of language was also consistently observed. Participants showed facilitated performance after 3 and 30 repetitions of an emotion label, but only when it matched both expressions in the pair. This may suggest that language labels *directly* influence early emotion perception (Lupyan, 2007, 2012), or provide strategic support during paired discrimination (e.g. Roberson & Davidoff, 2000). A perceptual threshold procedure was used to test the *direct* assumption. Participants repeated an emotion or neutral label before deciding whether a briefly presented face did, or did not, display an emotional expression. In comparison to the neutral baseline, participants showed no facilitation in performance following exposure to emotion labels that were ‘weakly’ or ‘strongly’ congruent with the subsequently presented expression. Overall, findings inconsistently support the notion that language shapes the interpretation of emotional expression. This prompts discussion of how task demands may influence language-driven recruitment of conceptual knowledge, and the time-course across which these linked elements influence interpretation.

## Table of Contents

<b>Intellectual Property and Publication Statements .....</b>	<b>iii</b>
<b>Acknowledgements.....</b>	<b>v</b>
<b>Abstract.....</b>	<b>vii</b>
<b>Table of Contents.....</b>	<b>viii</b>
<b>List of Tables.....</b>	<b>xvii</b>
<b>List of Figures.....</b>	<b>xviii</b>
<b>Chapter 1.....</b>	<b>1</b>
<b>General Introduction.....</b>	<b>1</b>
1.1 Perceiving emotion: How much useful information is displayed by the face? .....	1
1.1.1 The ‘basic’, or ‘discrete’ approach to emotion perception .....	1
1.1.2 Cross-cultural findings and the ‘universality hypothesis’ .....	2
1.1.3 Do expressions possess ‘category-diagnostic’ features and do these features aid emotion perception?.....	4
1.1.4 Can emotional expressions be perceptually ambiguous? .....	5
1.2 Do language labels influence emotion perception?.....	6
1.3 Construction accounts: Language labels and the activation of meaningful, semantic knowledge.....	8
1.3.1 The role of ‘external context’ in the categorisation of emotional expressions .....	8
1.3.2 Does language provide an ‘internal context’ for the categorisation of emotional expression? .....	10
1.3.3 How might semantic knowledge act as a top-down constraint for the categorisation of emotional expression? .....	13
1.4 The Label-feedback account (Lupyan, 2012): Language labels and the activation of perceptual forms of conceptual knowledge .....	15
1.4.1 Central assumptions of the label-feedback approach: language labels, perceptual knowledge and categorisation.....	15
1.4.2 Training paradigms: can ‘semantically-empty’ labels guide categorisation attempts? .....	16
1.4.3 Is the label-feedback account useful when assessing the language-mediated perception of emotional expressions?.....	18
1.5 Summary: Reconciling the construction and label-feedback accounts .....	21



1.5.1	The conceptual umbrella: flexible activation of semantic and perceptual forms of knowledge .....	22
1.5.2	Task demands and preferential activation of different sources of knowledge .....	23
1.5.3	Task demands and the activation of conceptual knowledge: Implications for the present work .....	24
<b>Chapter 2</b>	.....	<b>27</b>
	<b>Investigating the role of language in the perception of emotional expression: does the semantic satiation paradigm provide a useful method? .....</b>	<b>27</b>
2.1	General Introduction .....	27
2.1.1	Sematic satiation: paradigms and mechanisms .....	27
2.1.2	Semantic satiation and emotion perception: Lindquist et al. (2006) .....	28
2.2	Experiment 1: Replicating Lindquist et al. (2006, experiment 3) .....	33
2.2.1	Introduction .....	33
2.2.2	Method .....	35
2.2.2.1	Participants .....	35
2.2.2.2	Materials .....	36
2.2.2.2.1	<i>Words</i> .....	36
2.2.2.2.2	<i>Face Pairs</i> .....	36
2.2.2.3	Apparatus .....	37
2.2.2.4	Design .....	37
2.2.2.5	Procedure .....	39
2.2.3	Results .....	40
2.2.3.1	Data preparation .....	40
2.2.3.2	Main Analyses .....	41
2.2.3.3	Proportional Accuracy Analysis .....	43
2.2.3.3.1	<i>Repetition-based effects (predicted)</i> .....	43
2.2.3.3.2	<i>Other significant and marginal effects</i> .....	44
2.2.3.4	'Conditional Accuracy' Analysis .....	46
2.2.3.4.1	<i>Repetition-based effects (predicted)</i> .....	46
2.2.3.4.2	<i>Other significant and marginal effects</i> .....	46
2.2.3.5	Timed out trial analysis .....	47
2.2.3.5.1	<i>Repetition-based effects (predicted)</i> .....	47
2.2.4	Discussion .....	48
2.2.4.1	Exploring repetition-based decrements: category-based, construction vs. alternative accounts .....	48

2.2.4.2	Facilitative effects: exploring the interaction between word relevance and expression matching .....	50
2.2.4.3	Outline of subsequent experiments (2a, 2b, 3 and 4) .....	54
<b>Chapter 3</b>	.....	<b>57</b>
<b>Introducing baseline conditions to explore semantic vs. non-semantic explanations for the massed-repetition decrement (experiments 2a, 2b and 3)</b>	.....	<b>57</b>
3.1	Introduction.....	57
3.1.1	Introducing a neutral baseline (experiments 2a and 2b) .....	58
3.1.2	Specific predictions for E2a and E2b .....	60
3.2	Experiment 2a: Intermixed presentation of emotion and neutral labels .....	61
3.2.1	Method .....	61
3.2.1.1	Participants .....	61
3.2.1.2	Materials .....	62
3.2.1.2.1	<i>Words</i> .....	62
3.2.1.2.2	<i>Faces pairs</i> .....	63
3.2.1.3	Design.....	63
3.2.1.4	Procedure.....	63
3.2.2	Results .....	65
3.2.2.1	Main Analysis .....	65
3.2.2.2	Proportional Accuracy Analysis.....	66
3.2.2.2.1	<i>Repetition-based effects (predicted)</i> .....	66
3.2.2.2.2	<i>Other significant and marginal effects</i> .....	67
3.2.2.3	Conditional Accuracy Analysis.....	70
3.2.2.3.1	<i>Repetition-based effects (predicted)</i> .....	70
3.2.2.3.2	<i>Other significant and marginal effects</i> .....	70
3.2.2.4	Timed out Trial Analysis .....	73
3.2.2.4.1	<i>Repetition-based effects (predicted)</i> .....	73
3.2.2.4.2	<i>Other significant and marginal effects</i> .....	73
3.2.3	Results summary .....	74
3.3	Experiment 2b: Blocked presentation of emotion and neutral labels .....	76
3.3.1	Method .....	76
3.3.1.1	Participants .....	76
3.3.1.2	Materials .....	76
3.3.1.2.1	<i>Words</i> .....	76
3.3.1.2.2	<i>Faces</i> .....	76

3.3.1.3	Design.....	76
3.3.1.4	Procedure.....	77
3.3.2	Results .....	78
3.3.2.1	Main Analysis .....	78
3.3.2.2	Proportional Accuracy Analysis.....	79
3.3.2.2.1	<i>Repetition-based effects (predicted)</i> .....	79
3.3.2.2.2	<i>Other significant and marginal effects</i> .....	80
3.3.2.3	Conditional Accuracy Analysis.....	82
3.3.2.3.1	<i>Repetition-based effects (predicted)</i> .....	82
3.3.2.3.2	<i>Other significant and marginal effects</i> .....	82
3.3.2.4	Timed-out trial analysis.....	83
3.3.2.4.1	<i>Repetition-based effects (predicted)</i> .....	83
3.3.2.4.2	<i>Other significant and marginal effects</i> .....	84
3.3.3	Results summary .....	84
3.4	Integrated Discussion (intermixed and blocked experiments) .....	85
3.4.1	Carry-over effects and trial presentation .....	85
3.4.2	Repetition-based decrements: construction vs. non-semantic accounts.....	86
3.4.3	Facilitative effects: exploring the interaction between word type and expression matching.....	88
3.5	Experiment 3: Intermixed presentation of neutral and non-words.....	90
3.5.1	Introduction.....	90
3.5.2	Method .....	92
3.5.2.1	Participants .....	92
3.5.2.2	Materials .....	92
3.5.2.2.1	<i>Words</i> .....	92
3.5.2.2.2	<i>Faces</i> .....	93
3.5.2.3	Design.....	93
3.5.2.4	Procedure.....	94
3.5.3	Results .....	95
3.5.3.1	Main Analysis .....	95
3.5.3.2	Proportional Accuracy Analysis.....	97
3.5.3.2.1	<i>Repetition-based effects (predicted)</i> .....	97
3.5.3.3	Conditional Accuracy Analysis.....	97
3.5.3.3.1	<i>Repetition-based effects (predicted)</i> .....	97

3.5.3.4	Timed out trial Analysis.....	98
3.5.3.4.1	<i>Repetition-based effects (predicted)</i> .....	98
3.5.3.4.2	<i>Other significant and marginal effects</i> .....	98
3.5.4	Discussion .....	99
3.5.4.1	Repetition-based decrements: construction vs. non-semantic accounts .....	100
3.5.4.2	Non-replication of facilitative effects.....	101
<b>Chapter 4</b> .....		<b>103</b>
<b>Using ‘non-linguistic’ baseline measures to explore non-semantic mechanisms for the massed repetition decrement (experiment 4)</b> .....		<b>103</b>
4.1	Introduction.....	103
4.2	Method .....	107
4.2.1	Participants.....	107
4.2.2	Materials.....	107
4.2.2.1	Face pairs .....	107
4.2.3	Design .....	107
4.2.4	Procedure .....	109
4.3	Results .....	110
4.3.1	Main Analysis.....	110
4.3.2	Proportional Accuracy Analysis .....	111
4.3.2.1	<i>Repetition-based effects (predicted)</i> .....	111
4.3.2.2	<i>Other significant and marginal effects</i> .....	111
4.3.3	Conditional Accuracy Analysis .....	112
4.3.3.1	<i>Repetition-based effects (predicted)</i> .....	112
4.3.3.2	<i>Other marginal and significant effects</i> .....	113
4.3.4	Timed-out trial analysis .....	113
4.3.4.1	<i>Repetition-based effects (predicted)</i> .....	113
4.3.4.2	<i>Other significant and marginal effects</i> .....	114
4.4	Discussion .....	114
4.4.1	Repetition-based decrements: exploring the potential role of residual cognitive fatigue .....	114
4.4.2	Non-replication of facilitative effects .....	116
4.4.3	Summary of key findings .....	117

<b>Chapter 5</b> .....	<b>119</b>
<b>What can findings from the semantic satiation paradigm tell us about the role of language in emotion perception?</b> .....	<b>119</b>
5.1 Summary of main findings across experiments 1-4 .....	119
5.2 Generalised massed repetition effects .....	120
5.2.1 Lindquist et al. (2006) and the ‘category-based’ construction account .....	121
5.2.2 Exploring the interaction between repetition and expression-matching: the impact of trial presentation .....	122
5.2.3 The potential role of non-semantic mechanisms: residual cognitive fatigue, response uncertainty and label inaccessibility .....	123
5.3 Facilitation when all three stimuli match: A labelling effect? .....	127
5.3.1 Assessing the utility of a ‘direct’, label-feedback account (Lupyan, 2007, 2012) .....	128
5.3.2 Assessing the utility of an ‘indirect’, strategic labelling account .....	131
5.4 Methodological limitations and future directions .....	132
5.4.1 The impact of restricted response windows on accuracy rates .....	132
5.4.2 The immediate repetition paradigm: limitations and solutions .....	134
5.4.3 Delineating between the perceptual vs. post-perceptual influence of language on emotion perception: future directions .....	135
<b>Chapter 6</b> .....	<b>139</b>
<b>Using a threshold method to assess whether language directly shapes initial perception of emotional expressions (experiment 5)</b> .....	<b>139</b>
6.1 Introduction .....	139
6.1.1 Assessing the <i>Direct</i> vs. <i>Indirect</i> influence of language: the importance of methodology .....	139
6.1.2 Manipulating the ‘type’ and ‘relevance’ of linguistic cues .....	141
6.1.3 Specific predictions for the present work .....	144
6.2 Method .....	146
6.2.1 Participants .....	146
6.2.2 Materials .....	146
6.2.2.1 <i>Word stimuli</i> .....	146
6.2.2.2 <i>Face stimuli</i> .....	148
6.2.2.3 <i>Masking stimuli</i> .....	151
6.2.2.4 <i>Apparatus</i> .....	151
6.2.3 Design .....	152
6.2.3.1 <i>Pre-experimental, threshold trials</i> .....	152

6.2.3.2	<i>Experimental Trials</i> .....	152
6.2.4	Procedure .....	154
6.2.4.1	<i>Pre-experimental trials</i> .....	154
6.2.4.2	<i>Experimental trials</i> .....	154
6.3	Results .....	155
6.3.1	Data preparation .....	155
6.3.2	Accuracy Analysis (Hit rates, False alarms, $d'$ and $c$ ) .....	156
6.3.2.1	Hit Rates .....	160
6.3.2.2	False alarms.....	160
6.3.2.3	Sensitivity ( $d'$ ).....	160
6.3.2.4	Response Bias ( $c$ ).....	160
6.3.3	Reaction time analysis .....	160
6.4	Discussion .....	162
6.4.1	Summary of key predictions and findings .....	162
6.4.2	Methodological limitations: cue type and presentation duration .....	164
6.4.3	Methodological limitations: face stimuli.....	166
6.4.4	Theoretical implications .....	167
6.4.5	Conclusions and future work.....	169
<b>Chapter 7</b>	.....	<b>171</b>
<b>Do emotion-related, action words proportionally evoke discrete affective knowledge? Using a categorical method to generate and validate a stimulus set</b>	.....	<b>171</b>
7.1	Introduction.....	171
7.1.1	What are emotional verbs and why are we interested in them? .....	172
7.1.2	How can we quantify affective grounding for verbs? .....	173
7.1.3	Why might categorical ratings provide a particularly useful way to quantify the emotional meaning of verbs? .....	175
7.1.4	Summary of the present work.....	177
7.2	Study 1- Identifying action words that proportionally evoke affective knowledge .....	178
7.2.1	Method .....	178
7.2.1.1	Participants .....	178
7.2.1.2	Procedure.....	179
7.2.2	Results .....	179
7.2.2.1	Data modifications and modal exemplars .....	179
7.2.3	Discussion .....	184

7.3	Study 2- Validating associations between action words and emotion labels .....	185
7.3.1	Method .....	186
7.3.1.1	Participants .....	186
7.3.1.2	Design and Procedure .....	186
7.3.1.3	Materials .....	187
7.3.2	Results .....	188
7.3.2.1	Data preparation .....	188
7.3.2.2	Analysis.....	188
7.3.3	Discussion .....	190
7.4	General Discussion .....	191
	<b>Chapter 8.....</b>	<b>195</b>
	<b>General Discussion .....</b>	<b>195</b>
8.1	Key findings .....	195
8.2	Theoretical implications: the construction and label-feedback accounts. ....	196
8.2.1	Strand 1: The semantic satiation work (Chapters 2–5) .....	197
8.2.2	Strand 2: The threshold work (Chapter 6).....	199
8.3	Using language labels strategically to inform categorical decisions .....	201
8.4	Task demands and the selective activation of conceptual knowledge .....	202
8.5	Modifying the threshold paradigm to assess the time-course of language-mediated effects on emotion perception .....	206
8.6	Future strategies to assess the influence of semantic knowledge on emotion perception.....	207
8.7	Self-generation and rating of emotion-related verbs: theoretical implications and future directions .....	208
8.8	Potential applications.....	210
	<b>References.....</b>	<b>212</b>
	<b>Appendices .....</b>	<b>xv</b>
	Appendix A: Volunteer information sheet provided during experiments 1, 2a, 2b and 3.....	xv
	Appendix B: Health and safety exercises used in experiment 2b .....	xvii
	Appendix C: Volunteer information sheet provided for experiment 4.....	xix
	Appendix D: Volunteer information sheet provided for experiment 5. ....	xxi
	Appendix E: Full list of modal exemplars provided in the feature generation work (Chapter 7, Study 1).....	xxiii
	Appendix F: Mean scores for the full set of 36 label/action word ratings (Chapter 7, study 2).....	xxviii





## List of Tables

Table 2.1 Mean proportional accuracy rates reported by Lindquist et al. (2006, experiment 3, adapted from Table 7).....	30
Table 2.2 Results of Lindquist et al's. (2006) third experiment. ....	31
Table 2.3 Examples of the eight trial types used in experiment 1 (adapted from Lindquist et al. 2006).....	38
Table 2.4 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Relevance and Expression Matching (SEM presented in parenthesis).....	43
Table 3.1 Examples of the eight trial types used in experiment 2a (adapted from Lindquist et al., 2006).....	64
Table 3.2 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis). ' .....	66
Table 3.3 Mean, proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis). ....	79
Table 3.4 Examples of trial types from each of the eight conditions included in experiment 3 (adapted from Lindquist et al., 2006).....	94
Table 3.5 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis). ' .....	96
Table 6.1 Mean correct and incorrect categorisation percentages for selected KDEF faces, as presented in Goeleven et al. 2008 (SD presented in parenthesis). .....	150
Table 6.2 Mean hit rates, false alarms, $d'$ and $c$ (criterion) measures as a function of Word Type, Word Relevance and ISI length ( $N \leq 72$ , SEM presented in parenthesis).....	158
Table 6.3 Results of the Word Type $\times$ Word Relevance $\times$ ISI length mixed factorial ANOVA, for hit rates, false alarms, $d'$ and $c$ (criteria).....	159
Table 6.4 Mean reaction times as a function of Word Type, Word Relevance and ISI length (SEM presented in parenthesis). ....	161
Table 6.5 Mean correct reaction times as a function of Word Type, Word Relevance and ISI length.....	161
Table 6.6 Mean, correct reactions times by Word Type and Word Relevance (SEM presented in parenthesis).....	162
Table 7.1 Top three, most frequently generated action words, per emotion label. ....	183
Table 7.2 Mean typical and atypical ratings, $t$ , $p$ and $d$ statistics, per emotion label. ....	189

## List of Figures

Figure 2.1 Example of the temporal sequence for an judgment-irrelevant label, expression match trial.....	40
Figure 2.2 Mean proportional accuracy rates, by word relevance and expression matching (error bars represent SEM).....	45
Figure 2.3 Mean conditional accuracy rates by Word Relevance and Expression Matching (error bars represent SEM).....	47
Figure 2.4 Trial types and comparisons available in experiment 1.....	53
Figure 3.1 Mean proportional accuracy rates by Expression Matching and Word Type (error bars represent SEM).....	68
Figure 3.2 Mean proportional accuracy rates by Expression Matching and Repetition (error bars represent SEM). ....	69
Figure 3.3 Conditional accuracy rates by Expression Match and Repetition (error bars represent SEM).....	71
Figure 3.4 Mean conditional accuracy rates by Expression Matching and Repetition (error bars represent SEM). ....	72
Figure 3.5 Mean, proportional accuracy rates by Word Type and Expression Matching (error bars represent SEM).....	81
Figure 3.6 Mean proportion of timed out trials, by Repetition and Expression Matching (error bars represent SEM).....	99
Figure 4.1 Mean, proportional accuracy rates by Activity and Expression Matching (error Bars represent SEM).....	112
Figure 6.1 The temporal sequence for a single experimental trial. ....	155
Figure 7.1 Mean typical and atypical ratings, per emotion label (error bars represent 95% CI). ....	190

## Chapter 1

### General Introduction

#### 1.1 Perceiving emotion: How much useful information is displayed by the face?

##### 1.1.1 The 'basic', or 'discrete' approach to emotion perception

The 'basic' or 'discrete' view has accrued widespread support. This approach suggests that humans have evolved to both produce and decode a set of adaptive behaviours, associated with a select number of emotional states (e.g. Ekman & Friesen, 1975, 1976, 1978, Izard, 1992; Shariff & Tracy, 2011). Six emotional states are commonly considered 'basic' or 'universal': 'anger', 'fear', 'happiness', 'sadness', 'surprise' and 'disgust' (Ekman & Friesen, 1975, though see Ekman, 2003, for discussion of 'contempt'). Expression decoding is assumed to be effortless, as each basic state is associated with a set of non-overlapping 'behaviours', driven by pre-programmed physiological circuitry (e.g. Shariff & Tracy, 2011). This is particularly the case for behaviours that are automatically produced earlier, rather than later, in the time-course of an emotional experience (e.g. facial expressions, initial vocalisations and physiological reactions e.g. Zajonc, 1980; Ekman, 1992; Ekman & Cordaro, 2011, though see Matsumoto & Hwang, 2014).

The 'basicness' assumption has been extensively tested with facial expressions, as they represent an easily 'observable' by-product of emotion. The successful development of the Facial Action Coding System (FACS, Ekman & Friesen, 1978) is a good example of such work. This is an anatomical classification system, which proposes to accurately map changes in the 40+ action units contained in the face. Different configurations of these action units should correspond to displays of the basic emotional expressions (Ekman & Friesen, 1978). FACS-coded stimuli have since become the gold standard in emotion recognition research, widely used both by 'universalists' and 'non-universalists' to assess how humans achieve discrete categorisation of expression (e.g. Smith, Cottrell, Gosselin & Schyns, 2005; Jack,

Blais, Scheepers, Schyns & Caldara, 2009; Jack, Garrod, Yu, Caldara & Schyns, 2012; Matsumoto & Hwang, 2014).

### **1.1.2 Cross-cultural findings and the ‘universality hypothesis’**

The ‘universality hypothesis’ forms an important component of the ‘basic’ approach: that individuals from *all cultures* adopt the same facial configurations to express a limited set of basic emotions (e.g. Ekman, 1992). Cross-cultural research directly tests this claim. When conducted by supporters of the ‘basic’ emotion approach, ‘cross-cultural’ studies include a number of common, methodological elements; all of which have been associated with the potential to inflate agreement rates (e.g. Russell, 1994; Nelson & Russell, 2013; Yik, Widen & Russell, 2013). For example, categorisation follows a forced-choice format, in which participants match each of the displayed emotional expressions to a single basic emotion label, included within a wider list. To meet the universality assumption, participants from different cultures must agree that actors display intended expressions between 70-90% of the time (the pre-defined criterion specified by Haidt & Keltner, 1999). The results of several studies meet this criteria, when sampling across Western and non-Western populations (e.g. Ekman & Friesen, 1971; Ekman, 1992; Elfenbein & Ambady, 2002). However, two large cross-cultural reviews, conducted by Russell (1994) and Nelson and Russell (2013) found inconsistent support for the universality hypothesis.

Even when exclusively sampling studies that included these common methodological elements, categorisation agreement rates varied according to the cultural background and native language of the individual (there was higher agreement in samples that spoke Indo-European languages, Nelson & Russell, 2013). Agreement rates also varied according to expression; participants from different cultures generally showed high levels of agreement when categorising expressions related to happiness and, to a lesser extent, surprise (89.6% for ‘happiness’ and above 70% for ‘Surprise’ in Russell, 1994). Indeed, expressions of happiness often receive significantly higher cross-cultural agreement rates than other expressions (e.g. Jack et al., 2012; Damjanovic, Roberson, Athanasopoulos, Kasai & Dyson, 2010; Gendron, Roberson, van der Vyver., 2014b). However, the universality criterion went unmet when other expressions were considered; western vs. non-western participants showed higher within-sample agreement rates when categorising negative expressions, such as ‘anger’ and ‘disgust’ (Nelson & Russell, 2013).

The universality hypothesis also lacks support when tested using a robust, 'two-culture' approach (e.g. Russell, 1994; Norenzayan & Heine, 2005; Gendron, Roberson, van der Vyver & Barrett, 2014a, 2014b). Here categorisation rates of other and same-race faces are compared across western participants and non-western participants; the latter group having little exposure to Western displays of emotion in everyday life. Gendron et al. (2014b) used a 'free-sort' task with American and rural Namibian participants (the Himba ethnic group), in which they were required to sort posed photographs into piles of people 'feeling the same way'. While American participants were generally able to construct six piles, corresponding to each basic emotion label, the Himba participants made fewer piles and often conflated negative emotional expressions (see also Gendron et al., 2014a for similar categorisation behaviour with vocalisations of emotion). Lower agreement rates in the non-western sample may indicate that Western displays of emotion do not match those they encounter in same-race faces in everyday life (e.g. Norenzayan & Heine, 2005; Jack et al., 2009).

The universality hypothesis has also failed to find support when different methodologies were employed. Jack et al. (2012) asked participants to indicate when a 4D simulation of a moving face displayed an expression that could be categorised as 'emotional' (e.g. Yu, Garrod & Schyns, 2012). Arguably participants should only stop the simulation when the perceptual input matched their experience-driven mental representation of what an emotional face looks like (e.g. Jack, Caldara & Schyns, 2012). When face simulations depicted white, western models, both Western and East Asian participants showed a clear behavioural pattern, indicating that expressions were emotional when they broadly corresponded to FACS-coded configurations of the six basic expressions. In contrast, when simulations of East Asian faces were shown, participants only categorised four (vs. six) clusters of facial behaviours as emotional. Using an eye-tracking technique Jack et al. (2009) also reported differences in the range and regions of gaze fixations employed by Western and East Asian participants when they categorised FACS-coded expressions, in same and other-race actors. Where East Asian participants spent longer looking toward the eye regions, which may have contributed towards mistaken categorisations, Westerners showed no bias in their fixations. Together, these findings may indicate that the way people express, and decode, emotion varies according to their cultural background. Also that FACS-coded expressions, whether displayed by same or other-race faces, may only

support a Western conception of emotion (e.g. Jack et al., 2009, 2012; Gendron et al., 2014b).

### **1.1.3 Do expressions possess ‘category-diagnostic’ features and do these features aid emotion perception?**

Taken together, this evidence casts doubt on the ‘universality hypothesis’. However, even ‘non-universalists’ would agree that important information can be derived from the face (e.g. Russell, 1994; Nelson & Russell, 2013). Various researchers argue that at least some discrete emotional expressions hold ‘diagnostic’ features; when emotional categorisation is the task goal, attention may be directed towards these features to provide a ‘perceptual short-cut’ for categorisation (e.g. Ellison & Massaro, 1997; Schyns, Bonnar & Gosselin, 2002; Calvo & Nummenmaa, 2008; Calvo & Marrero, 2008, though see Calder, Keane, Young & Deane, 2000, who propose a holistic processing strategy for emotional faces). Evidently what constitutes a ‘category-diagnostic’ feature for a particular expression will vary, dependent on the cultural background of the observer and sender (e.g. Jack et al., 2009; Damjanovic et al., 2010). Highlighting differences in approach, categorisation via this process is thought to be more ‘strategic’ and ‘feature-driven’ (e.g. Calvo, Fernandez-Martin & Nummenmaa, 2012), than that proposed by advocates of the ‘basic’ approach, who emphasise the automaticity and ease with which categorisation occurs (e.g. Shariff & Tracy, 2011).

Mixed evidence emerges from the few studies that simultaneously explore the existence of diagnostic features for all six, basic expressions. Smith et al. (2005) used a ‘bubbles’ technique (Gosselin & Schyns, 2002), where a randomly changing pattern of Gaussian blur and clear *windows* were applied, at different spatial bandwidths, to photographs of actors displaying FACS-coded emotional expressions. Participants were tasked with categorising expressions and category-diagnostic features/regions were identified as those that were most likely to be visible during correct categorisations. ‘Bubble’ patterns varied considerably by expression. Although overlaps in diagnostic region were found for some expressions, Smith et al. (2005) comment that consideration of that region will yield different categorisations, dependent on the exact *perceptual* information contained therein (e.g. the teeth will be visible when a mouth displays happiness, but closed when it displays surprise). Similarly, Schurgin, Nelson, Lida, Ohira, Chiao & Franconeri et al. (2014) used an eye-tracking technique to detect strategic fixations to different regions of the face

when participants made emotionality decisions ('does this face show an emotional or neutral expression?'). In partial support of Smith et al. (2005), Schurgin et al. (2014) found patterns of first (and, to a lesser extent, sustained) fixations that differed according to which emotional expression had been displayed (although 88% of total fixations were toward the eye, mouth and nasion regions). However, it remains unclear whether Schurgin et al.'s. (2014) results reflect spontaneous vs. strategic differences in fixation pattern for each expression, as participants were asked to respond to blocks that contained only one type of emotional expression. As such, fixated regions may simply have been diagnostic for differentiating that particular expression from a neutral face. This is unlikely to reflect the nature of emotion categorisation in everyday life.

Despite inconsistencies when all expressions are sampled, there is strong evidence to suggest that 'smiles' represent a category-diagnostic feature of the 'happy' expression (e.g. Calvo & Numenmaa, 2008; Calvo & Marrero, 2008; Damjanovic et al., 2010; Gendron et al., 2014b). This perhaps explains why happy expressions are most accurately categorised in cross-cultural research (e.g. Damjanovic et al., 2010; Gendron et al., 2014a). Using the 'bubbles' technique (Gosselin & Schyns, 2002), researchers confirmed that the mouth region (particularly the upper lip) was the most dominant predictor of accuracy when participants judged whether faces displayed a happy or neutral expression (although some data also suggested that participants considered a conjunction of information, formed by both the mouth and eye regions; Gosselin & Schyns, 2001; Schyns et al., 2002; Joyce, Schyns, Gosselin, Cottrell & Roisson., 2006). Using different methods, Calvo et al. (2012) showed that participants frequently categorised blended expressions as 'happy' when a smiling mouth was paired with a set of eyes that were associated with a different emotional expression. This effect persisted when the blended expressions were inverted, suggesting that mistaken categorisations were predominantly based on a featural analysis of the mouth.

#### **1.1.4 Can emotional expressions be perceptually ambiguous?**

An important criticism can be applied to research used to investigate category-diagnostic features in emotional expression, as well as cross-cultural work; that they make use of FACS-coded facial stimuli (e.g. Ekman & Friesen, 1978; Smith et al., 2005; Jack et al., 2009, 2012). As such, it remains unclear whether the category-diagnostic features identified in such work actually appear in spontaneous displays of emotion, and/or similarly aid categorisation attempts in everyday life (e.g. Russell & Barrett, 1999; Russell, 2003; Barrett,

Lindquist & Gendron, 2007; Matsumoto, Keltner, Shiota, O'Sullivan & Frank, 2008; Nelson & Russell, 2013).

The few studies that use spontaneously produced, or low intensity emotional expressions, find substantially lower agreement rates than those that use prototypical stimuli (e.g. Hess, Blairy & Kleck, 1997; Matsumoto et al., 2002; Naab & Russell, 2007; Matsumoto, Olide, Schug, Willingham & Callan, 2009; Bould & Morris, 2008; Bould, Morris & Wink, 2008). For example, Naab and Russell (2007) and Matsumoto et al. (2009) found categorisation rates between 31-35% both when single-culture and cross-cultural groups of participants categorised spontaneous emotional expressions. However, Matsumoto and Hwang (2014) argue that some instances of low agreement may be an artefact of the methods used to reduce expression intensity. For example, when neutral and expressive endpoints of the same face are morphed together, the action units in different regions of the morphed image may display expressive properties to varying extents, therefore, producing an 'anatomically incorrect' expression, unlikely to be encountered in real life (e.g. Matsumoto et al., 2002; Matsumoto & Hwang, 2014). To counter this limitation, Matsumoto and Hwang (2014) reduced the intensity of their stimuli by carefully varying each of the action units associated with the FACS prototype for that particular expression. They found that participants were still able to accurately categorise these new, subtle expressions, at rates higher than chance. However, agreement rates fluctuated according to the expression shown, predictably higher for expressions of happiness and surprise, but lower, and more variable, for negative emotional expressions.

## 1.2 Do language labels influence emotion perception?

The previous discussion suggests that it is unclear to what extent categorisation of emotional expression is informed by the *perceptual* information contained in a face. According to Barrett (2006b) this leaves us with an 'emotion paradox': why do we have the strong sense that we observe a set of basic emotional states in others (and ourselves) if the available *perceptual* evidence fails to drive these conclusions? Barrett (2006b; Barrett et al., 2007; Scherer & Ellgring, 2007) suggest that there must be other information, available at the time of perception, to support categorisation. This *contextual* information may drive categorisation independent of the available *perceptual* evidence, or may simply bias the way we process and interpret that information. Drawing upon a previous example, an



individual's culture may act as a *contextual* factor, appearing to dictate which diagnostic features of an expression are attended for categorisation (e.g. Jack et al., 2009, 2012).

The influence of *context* is a critical facet of the construction approach (e.g. Russell, 2003; Barrett, 2006b, 2009, 2013). Although other *contextual* factors are briefly identified and discussed below (see section 1.3.1), this thesis will focus on the possible role played by one particular factor: *language*. This proposal is strongly at odds with the 'basic' approach, which views language as a direct product of social construction, and emotions as hard-wired, and therefore, independent of one another (e.g. Izard, 1994; Ekman & Cordaro, 2011).

Cross-cultural work, already discussed, provides preliminary evidence that an individual's language may influence their categorisation decisions. Commonly these tasks require that participants match each expression to one possible emotion label (e.g. 'sad'), included within a wider list of basic labels (Ekman & Friesen, 1975; Ekman, 2003). Importantly, decreases in agreement levels are found when this list is modified in some way (e.g. Nelson & Russell, 2013; Yik et al., 2013). Modifications may involve expanding the set of labels to include 'non-basic' options (e.g. amusement), removing one or more of the basic labels (e.g. Frank & Stennett, 2001), including a 'not sure' option (e.g. Frank & Stennett, 2001; Yik et al., 2013) or allowing participants to select more than one label to categorise each expression (e.g. Russell, 1993; Kayyal & Russell, 2012). Categorisation agreement is at its lowest when participants are provided with no list at all and must 'freely label' the faces shown, using either one or more labels (e.g. Rosenberg & Ekman, 1995; Haidt & Keltner, 1999; Widen, Christy, Hewett & Russell, 2011; Gendron et al., 2014a). Cross-cultural differences are also commonly exacerbated when 'free-labelling' methodologies are employed; discarding the 'basic' options, Japanese and American participants commonly provide labels indicative of internal states, or mental experiences (e.g. 'thinking', Russell, 1993; Gendron et al., 2014b), whereas Namibian participants preferred to categorise on the basis of behavioural terms (e.g. 'smiling' Gendron et al., 2014b, see also Gendron et al., 2014a). This may suggest that providing 'basic' language labels drives or constrains the categorisations that participants usually form (e.g. Barrett et al., 2007; Nelson & Russell, 2013; Yik et al., 2013).

Arguably, language may influence emotion perception via a number of different mechanisms. That numerous possibilities exist can be seen in the diversity of ways in which ‘language’ has been operationalised in previous research. While some researchers make use of single words to evoke an emotional *context* (e.g. basic emotion labels, connotative/emotion-laden words, emotion-related action words e.g. Lindquist, Gendron, Barrett & Bliss-Moreau, 2006; Foroni & Semin, 2009; Halberstadt, Winkielman, Niedenthal & Dalle, 2009), others use whole sentences or emotion-relevant scenarios (experimenter or participant-generated e.g. Carroll & Russell, 1996; Halberstadt & Niedenthal, 2001; Anderson, Siegel, Bliss-Moreau & Barrett, 2011). A range of these findings are pertinent when discussing the two possible accounts of language-mediated perception that are most relevant for this thesis: the construction approach (e.g. Barrett, 2006a, 2006b; Barrett et al., 2007) and the label-feedback hypothesis (e.g. Lupyan, 2007, 2012)<sup>1</sup>. Below, the central tenets of each theory are outlined, including the way in which links between language and *conceptual* knowledge are viewed and how these interlinked components are proposed to influence perception of emotion-relevant stimuli. In a final sub-section, attempts are made to reconcile the two accounts; similarities and differences are highlighted and methods are identified that would be best placed to examine each proposal. Finally, we discuss the extent to which work included in this thesis will have comparative implications for each account.

### **1.3 Construction accounts: Language labels and the activation of meaningful, semantic knowledge.**

#### **1.3.1 The role of ‘external context’ in the categorisation of emotional expressions**

Constructionists argue that we use a variety of information to discretely categorise facial expressions as belonging to one emotional state, above others (see also componential

---

<sup>1</sup> It is acknowledged that several other accounts exist to describe the influence of language on the perception of emotional expression. Most notably the ‘embodied’ position, which suggests that exposure to emotion words produce congruent facial mimicry which, in turn, improves processing of expressive stimuli (e.g. Foroni & Semin, 2009; Halberstadt et al., 2009; Niedenthal, Winkielman, Mondillion & Vermeulen, 2009). However, these accounts were not directly tested in this thesis, so will not be discussed further.

models e.g. Scherer, 1984). Various empirical findings support the contribution of ‘context’ (e.g. Barrett et al., 2007; Barrett & Kensinger, 2010; Barrett, Mesquita & Gendron, 2011). This ‘context’ may be intrinsic to the actions of the person experiencing emotion, and presented concurrent with their facial expression (e.g. gestures, behaviours and vocalisations, Aviezer et al., 2008). Conversely, ‘context’ may be external, based on factors present in the environment in which emotion perception occurs, for example, when emotion-related colours, such as red, as presented simultaneous with the face (e.g. Fetterman, Robinson & Meier, 2012; Young, Elliot, Feltman & Ambady, 2013) or situational/biographical information about the individual experiencing emotion (e.g. Fernández-Dols, Wallbott & Sanchez, 1991; Carroll & Russell, 1996). Importantly, various findings suggest that contextual information has a powerful influence, driving discrete categorisation of facial expression, either when the face displays a neutral/ambiguous expression (e.g. Halberstadt & Niedenthal, 2001; Fernández-Dols, Carrera, Barchard & Gacitua, 2008; Halberstadt et al., 2009) or when the face displays a different, discrete expression (e.g. Carroll & Russell, 1996; Aviezer et al., 2008). For example, when manipulating ‘intrinsic’ context, participants were more likely to perceive a facial expression of ‘disgust’ as ‘pride’, when the face was paired a muscular body adopting a confident pose (Aviezer et al., 2008). Using an ‘external’ contextual manipulation, a significantly higher number of participants labelled a ‘fearful’ face as ‘angry’ when it was presented alongside a scenario that would commonly elicit anger; such as ‘receiving a parking ticket’ (Carroll & Russell, 1996). Overall, both types of finding suggest that different forms of contextual information can drive interpretation, particularly when faces and contextual information share valence (they are both positive or negative, respectively) and degree of arousal (how intensely each piece of information signals the presence of emotion; e.g. Carroll & Russell, 1996). These findings allow constructionists to make two further arguments.

First, that although *perceptual* information from the face can influence judgments of emotionality, it is often not strong enough, or diagnostic in its’ own right, to inform discrete categorisation (e.g. Fernandez-Dols et al., 1991; Carroll & Russell, 1996; Mauss & Robinson, 2009; Barrett & Kensinger, 2010; Lindquist, Gendron, Barrett & Dickerson, 2014). Indeed, constructionists commonly claim that it is only possible to accurately extract valence and arousal information from facial expression, which stands in direct contrast to a ‘basic’

account (e.g. Lindquist, 2009; Barrett & Kensinger, 2010; see also dimensional theories e.g. Barrett & Russell, 1999; Russell & Widen, 2002; Russell, 2003).

Second, that in order for ‘external’ contextual information to have a bearing on emotional interpretation, it must be intrinsic to emotional meaning itself (e.g. Fehr & Russell, 1984; Shaver, Schwartz, Kirson & O’Connor, 1987; Russell, 1991a). Specifically, this kind of ‘non-*perceptual*’ information must have been stored as ‘emotionally-relevant’ by the participant during previous, emotional experiences, and must be readily accessible for comparison when interpreting new, emotional information. To take precedence over facial information the information must also be easily ‘categorisable’, or related to a discrete emotional state (e.g. Fernández-Dols et al., 1991; Widen, 2013). Non-*perceptual* information may be episodic (e.g. ‘how X acted the last time they were sad’) and/or *semantic* in nature (e.g. ‘failing an exam usually makes people feel sad’; e.g. Fernández-Dols et al., 1991; Lindquist, 2013; Lindquist et al., 2014).

### **1.3.2 Does language provide an ‘internal context’ for the categorisation of emotional expression?**

The proposal above suggests that we build a repository of emotion-related, semantic information through our everyday experiences. In order for this information to aid categorisation of emotion expression, when re-activated, it must be effectively organised. Indeed, it is argued that different pieces of semantic information are anchored to specific hubs, which correspond to particular emotional states (e.g. ‘what we know about sadness’). This type of ‘glue’ is important when there may be so much individual variability in the emotion-relevant information that a person integrates into their script (e.g. Somerville & Whalen, 2006; Barrett, Gendron & Huang, 2009; Lindquist, 2009). From infancy language labels (e.g. ‘sad’) are presumed to play this important role (e.g. Li & Gleitman, 2002; Xu, 2002; Barsalou & Weimer-Hastings, 2005). Implicating language allows constructionists to add other important facets to their theory. First, organisation by language labels facilitate the crucial activation of *semantic* knowledge. When appropriate *semantic* knowledge is activated people can form comparisons between externally presented, contextual information, and previously experienced information (e.g. ‘Y has just had a paper accepted. When X had a paper accepted they were very happy, which likely suggests that Y is experiencing happiness.’). In this sense, language labels facilitate a process by which

observers are directed towards, and make meaningful, emotion-relevant information in their environments (e.g. Barrett & Kensinger, 2010).

Driven by the ‘basic’/‘discrete’ view, that the structural information contained within a face is sufficient for emotion perception, many researchers simply present static, decontextualized emotional faces for categorisation (e.g. see the methodological description of cross-cultural work, above; Russell & Widen, 2002; Nelson & Russell, 2013). In the absence of contextual information, language may play a second role. Given the usefulness of *semantic* knowledge under ‘normal’, comparative instances of emotion perception (described above), participants likely engage in automatic knowledge activation when presented with decontextualized stimuli (e.g. Barrett et al., 2007). Although *semantic* knowledge cannot be used for the same type of comparison process (comparing current emotional input to that previously experienced), it may help the observer by providing information needed to ‘force’ categorisation of expression, useful when the available *perceptual* information is insufficient for such a decision (e.g. Russell & Widen, 2002; Whalen & Somerville, 2002). In other words, language labels (e.g. ‘sad’) are likely responsible for automatic activation of *semantic* knowledge, therefore providing a third, ‘internal context’ for the categorisation process (e.g. Lindquist et al., 2006; Barrett et al., 2007; Barrett & Kensinger, 2010; Gendron; Lindquist, Barsalou & Barrett, 2012). This may explain why participants show higher agreement in their categorisation of discrete facial expressions when the task involves matching the face to a label, rather than free label generation (e.g. Russell, 1994; Barrett et al., 2007; Nelson & Russell, 2013; Yik et al., 2013, see above). In addition to providing a constrained choice, labels facilitate participants’ ability to access internal *semantic* knowledge; representing an analogue to the more effortful, spontaneous label generation and knowledge retrieval that would usually be required when decontextualized stimuli are presented (Halberstadt & Niedenthal, 2001; Fernandez-Dols et al., 2008; Lindquist et al., 2014).

Interestingly, Halberstadt and Niedenthal (2001) and Halberstadt (2005) harness this ‘internal’ contextual knowledge, using an analogue of Carroll and Russell’s (1996) paradigm. In Halberstadt and Niedenthal’s (2001) study participants viewed a blended emotional face, which was equal parts fearful and angry, and a third of participants were asked to verbally explain why the actor might explicitly feel ‘fearful’ or ‘angry’. As no external context was provided by the experimenters, this manipulation invites explicit

activation of the participants' own *semantic* knowledge base. Following their description, participants were shown a video of that same face morphing between expressions of 'fear' and 'anger', and were asked to stop the video when they recognised the face they had previously described. Those who had described why the actor might be 'angry' were significantly more likely to stop the video nearer the 'angry' end of the spectrum, typically choosing an 'angrier' face than that presented during encoding. Perceptual biases were significantly stronger for participants in the description condition, in comparison to those who had simply seen an emotion label alongside the face, or passively viewed the faces, during encoding (baseline condition). Findings suggest that label-activated *semantic* knowledge biased perception of the test face, particularly when participants actively 'engaged' with knowledge, through description, rather than spontaneously activating knowledge in response to a label. Although several researchers argue that these findings may instead reflect explicit memory for the description and/or label provided during encoding (an angry face must be chosen to match the description given, see Halberstadt, Winkielman, Niedenthal & Dalle, 2009), participants exhibited poor explicit memory for labels when re-shown the original test faces (Halberstadt & Niedenthal, 2001; Halberstadt, 2005). Perceptual biases were also no stronger when participants were explicitly asked to remember their descriptions for a later test (Halberstadt & Niedenthal, 2001). These findings effectively demonstrate that activated *semantic* knowledge can influence perception of emotional expression.

Documented categorisation decrements in populations who lack access to *semantic* knowledge provide further evidence for the automatic activation, and use, of this information during emotion perception. For example, very young children (Russell & Widen, 2002; Widen & Russell, 2008) and patients with semantic dementia (Calabria, Cotelli, Adenzato, Zannetti & Miniussi, 2009; Lindquist et al., 2014). Specifically, when asked to sort emotional faces into piles of people who 'felt the same way' these participants often performed the task on the basis of valence and arousal, grouping together discrete expressions which are similarly positive/negative or showed 'emotion' to a similar degree of intensity. Interestingly both groups of participants were able to access basic emotion labels, which suggests that these labels are redundant for discrete categorisation when they are 'semantically-empty' (e.g. Ridgeway, Waters & Kuczaj, 1985; Widen, 2013, though see section 1.4.2 below). In support, children's ability to discretely categorise emotional expressions develops at a similar trajectory as their ability to do the

same for stories that describe an emotionally-relevant event (Widen & Russell, 2003, 2008; Widen, 2013). One can argue that these stories directly tap the *semantic* information that children gradually acquire during their own emotional experiences.

In sum, constructionists suggest that language labels (e.g. 'Sad') play a referential role, responsible for organising and re-activating an individual's repository of *semantic* knowledge about emotion (Barrett, Mesquita & Gendron, 2011; Widen, 2013). As such, emotion perception in both 'contextualised' and 'decontextualized' tasks is primarily supported by this *semantic* knowledge base, either because knowledge facilitates a comparison process between current and previously experienced emotion-relevant situations, or because knowledge 'biases' the categorisation of perceptually ambiguous expressions. Language labels are important insofar as they *indirectly* support knowledge recruitment.

### **1.3.3 How might semantic knowledge act as a top-down constraint for the categorisation of emotional expression?**

An important question remains. Given its' importance, how exactly does meaningful, *semantic* knowledge influence the visual perception of emotional expression? As constructionists argue that facial expressions themselves provide little diagnostic information, there are two possible ways in which knowledge can influence this process. First, *semantic* knowledge may influence face processing by strongly biasing weak *perceptual* markers of emotionality, or constraining their interpretation (e.g. Halberstadt & Niedenthal, 2001; Barrett et al., 2007; Nook, Lindquist & Zaki, 2015). Herbert, Sfarlea and Blumenthal's (2013) findings support this conclusion. These researchers asked participants to passively view emotional faces alongside sender-relevant labels (e.g. 'his fear') and found heightened amplitudes, indicative of enhanced early structural processing, when compared to unlabelled faces (P100, N170). A similar conclusion was reached in a binocular rivalry paradigm, in which pictures of emotional faces and houses were displayed simultaneously to the same eye (Anderson et al., 2011). Participants experienced longer periods of 'face dominance' when that face had been previously paired with affective, or socially-relevant gossip (e.g. 'threw a chair at a classmate') vs. non-affective, socially-irrelevant gossip, about the actor (e.g. 'drew the curtains in the room'). These effects were strongest for emotional faces that showed a negative expression, when paired with negative gossip (see also Abdel Rahman, 2011). Taken together, these findings suggest that

meaningful, *semantic* knowledge enhances early and late visual processing of emotional faces. ‘Enhanced’ processing may constitute, and result from, a general increase in the attention paid to the face during encoding, or from specific deployment of attention towards the few emotional markers (perceptual) within the expression (Collins & Olson, 2014).

As a second option, *semantic* knowledge may actively ‘construct’ a new percept of the face, independent of the *perceptual* information contained therein. Using fMRI, Fox, Moon, Iaria & Barton (2009) found neural adaptation when two sequentially-presented emotional faces were given the same basic label (e.g. ‘sad’), even when these faces differed perceptually (e.g. one face displayed a ‘sad’ expression, and the other, an ‘angry’ expression). The opposite was also true; neural adaptation was not found when the faces were *perceptual* representations from the same emotion category (e.g. two frowning faces), but were given two different emotion labels (e.g. ‘sad’ and ‘angry’). Similarly, when Thielscher and Pessoa (2007) asked participants to categorise the expression displayed upon a neutral face, dissociable neural signatures were displayed dependent upon whether participants (mistakenly) labelled the expression as ‘disgusted’ or ‘fearful’. Taken together, these findings suggest that label-activated *semantic* knowledge may literally change the way we perceive a face, creating a convincing, new emotional percept that allows us to abstract across ambiguous perceptual information (Thielscher & Pessoa, 2007).

These two possible explanations also find relative support in research that explores how the processing of facial identity may be influenced by learning biographical information about that actor (e.g. the actor’s occupation, hobbies, marital status, political attitudes; Abdel Rahman & Sommer, 2012). Researchers suggest that biographical information is spontaneously retrieved when we encounter previously seen faces, just as emotion-relevant *semantic* information may be retrieved when we attempt to categorise emotional expressions (e.g. Gobbini, Leibenluft, Santiago & Haxby, 2004; Gobbini & Haxby, 2007; Todorov, Gobbini, Evans & Haxby, 2007). Some physiological findings support the notion that participants show enhanced *perceptual* processing of faces paired with biographical information, in comparison to true unfamiliar faces, newly presented at test (e.g. Abdel Rahman, 2011; Abdel Rahman & Sommer, 2012). However, others show reduced markers of *perceptual* processing for trained-familiar vs unfamiliar faces, suggesting that linked biographical knowledge reduces the high *perceptual* demands associated with making



assessments of facial identity (e.g. Galli, Fuerra & Viggiano, 2006; Heisz & Shedden, 2009; Herzmann & Sommer, 2010; but see Paller, Gonslaves, Grabowecky, Bozic & Yamada, 2000).

In sum, the available evidence does not definitively explain how meaningful, *semantic* knowledge guides visual perception of emotional faces. Support exists for both of the possible constructionist positions; that meaningful knowledge enhances *perceptual* processing of the weak emotional markers contained within a face, or that knowledge biases processing, creating a new percept of the face, independent from *perceptual* information contained therein. Although the precise mechanisms are still under debate, findings converge to support the notion that emotion-related semantic knowledge exerts a powerful top-down influence on the way we perceive of emotional stimuli.

#### **1.4 The Label-feedback account (Lupyan, 2012): Language labels and the activation of perceptual forms of conceptual knowledge**

Predominantly explored in relation to object perception, the label-feedback account offers a different explanation for language-mediated perception (Lupyan, 2007, 2012). After outlining the basic assumptions and evidence for this position, the potential applications for emotion perception are discussed.

##### **1.4.1 Central assumptions of the label-feedback approach: language labels, perceptual knowledge and categorisation**

Critically, the label-feedback account relies on the proposal that objects hold a number of category-diagnostic features (i.e. features that are unique to objects belonging to a particular category, Harnad, 1987; Lupyan, 2007, 2012). Under normal circumstances then, people can judge category-membership based on visual inspection of the object alone. However, Lupyan (2007, 2012) argue that language labels may temporarily ‘boost’ visual categorisation. Lupyan (2012) state that label-based facilitation relies on the special relationship that forms between category-labels, and category-diagnostic perceptual features, contained within an object. These relationships are likely formed during everyday life, when we frequently use ‘basic-level’ labels to name referent objects, visible in our environment (e.g. Lupyan, 2008b). Therefore, the label-feedback account proposes that

simple exposure to a category label fosters a strong, but transient set of '*perceptual* expectancies'; when we hear the label 'chair' we expect to visually perceive an object that adheres to our prototypical conception of a chair (i.e. possessing four legs and a seat). When our expectations are then met, and a picture of a 'chair' presented, the category-diagnostic features of that 'chair' are accentuated, while our attention is simultaneously drawn towards those features. These facilitative effects are likely reliant upon a temporary perceptual warping process, whereby objects from the labelled category temporarily appear more 'typical', or in stronger possession of category-diagnostic features, and objects from a different category appear more disparate, or lacking in category-diagnostic features (e.g. Lupyan & Spivey, 2010b; Lupyan, 2012). Assumptions regarding *perceptual* expectancies rely on the notion that visual systems are partially predictive; upon label exposure, visual areas associated with the efficient processing of the particular category-diagnostic features of that object are temporarily primed (e.g. Bar et al., 2006; Kveraga, Ghuman & Bar, 2007; Lupyan & Spivey, 2010a, but see Eger, Henson, Driver & Dolan, 2007). Various findings support this notion, showing that labels only, or more strongly, facilitate the perception of 'typical' objects, which unambiguously meet our *perceptual* expectancies (e.g. Lupyan & Spivey, 2010a; Lupyan & Swingley, 2012; Edmiston & Lupyan, 2015). Typicality effects have been shown when using a variety of stimuli (e.g. line drawings and pictures of objects, numbers, letters) and presentation techniques (e.g. visual search paradigms; threshold presentation; target verification).

#### **1.4.2 Training paradigms: can 'semantically-empty' labels guide categorisation attempts?**

Training paradigms provide evidence in support of the label-feedback approach (e.g. Lupyan, Rakison & McClelland, 2007; Lupyan & Casasanto, 2014). Using unfamiliar and/or alien objects (e.g. YUFO stimuli, Gauthier, James, Curby & Tarr, 2003) allow researchers to control the degree of pre-existing knowledge that participants bring to the task (e.g. James & Gauthier, 2004; Abdel Rahman & Sommer, 2008). In Lupyan et al. (2007) and Lupyan and Casasanto (2014) participants were trained to associate 'families' of these alien objects either with a nonsense label (e.g. 'foove') or no-label at all. Following training, participants were required to make a categorical decision about the stimuli; deciding whether previously seen, or new aliens, did or did not belong to a particular family of 'aliens'. Crucially, labels were not presented during discrimination, nor did the task require explicit

retrieval of a label. Findings suggest that participants in the label vs. no-label training conditions were better to 'categorically perceive' stimuli.

Findings from both of these studies were taken as evidence to suggest that 'semantically-redundant' labels are sufficient to drive categorical perception of objects. In addition, Lupyan and Casasanto (2014) introduced further conditions to the paradigm. Based on pilot testing, two types of nonsense labels were chosen; one that participants implicitly associated with the particular *perceptual* features exhibited by a family of aliens (e.g. 'foove' is more likely to relate to a rounded vs. pointy object; congruent nonsense-label condition), and one that they did not (e.g. the label 'foove' was instead paired with a pointy object; incongruent nonsense-label condition). They also included a condition in which participants were trained to associate one family of aliens with a real-word label, that described a *perceptual* feature shared by all members of the family (e.g. 'pointy'). Lupyan and Casasanto (2014) found no difference in performance between participants in the real-word label and congruent nonsense-label conditions, but comparatively worse performance in the incongruent nonsense-label condition. These findings in particular suggest that labels *directly* facilitate object perception because they allow quick identification of category-diagnostic features; participants showed inferior performance in a categorisation task when the label assigned to that family of objects did not 'match' the *perceptual* features displayed .

However, a caveat may apply to this work. During the training and verification stages of both Lupyan et al's. (2007) and Lupyan and Casasanato's (2014) experiments, participants were asked to differentiate between novel aliens by deciding whether they would adopt 'approach' or 'avoidance' strategies on exposure. Training participants in this way may mean that they were afforded additional *semantic* and/or affective information about these aliens, bringing into question whether these paradigms truly offered a 'label-only' condition. Although participants in the 'no-label' condition also received approach/avoidance training, it remains unclear whether facilitation in the label condition results from the 'direct' *perceptual* advantage that labels provide, or as an 'indirect' consequence of labels facilitating access to affective/*semantic* information (see also Curby, Hayward & Gauthier, 2004; Collins & Curby, 2013). This argument was addressed in the experiments where novel object perception is compared after participants are trained to associate these objects with limited *semantic* information (e.g. a selection of features, such

as 'sticky', 'loud' and 'nocturnal') or a nonsense label, only. Some of these studies report behavioural and physiological findings consistent with processing advantages (perceptual or categorical) when objects are paired with *semantic* information/features vs. semantically-empty, nonsense labels (e.g. James & Gauthier, 2004; Abdel Rahman & Sommer, 2008; Collins & Curby, 2013).

### **1.4.3 Is the label-feedback account useful when assessing the language-mediated perception of emotional expressions?**

Some findings suggest that a label-feedback approach has potential to explain the facilitative role of language in the perception of emotional expression. As previously discussed (see section 1.1.3), various researchers suggest that basic emotional expressions possess at least some category-diagnostic features, though the number of which may vary according to the particular expression (e.g. Schyns et al., 2002; Smith et al., 2005; Schurgin et al., 2014). Enhancing the applicability of the label-feedback approach, findings suggest that people can discretely categorise at least some basic expressions based on visual consideration of diagnostic features alone (e.g. 'smiles' of happiness; Ellison & Massaro, 1997; Fiorentini & Viviani, 2009; Calvo & Nummenmaa, 2008; Calvo & Marrero, 2009; Calvo et al., 2012). Additionally, some researchers suggest that this type of featural analysis may occur before *semantic*, or affective, processing of the face (Storbeck, Robinson & McCourt, 2006; Calvo et al., 2012). Taken together, these findings support one of the basic tenets of the label-feedback approach; namely that participants can use diagnostic, perceptual features to categorise expressive stimuli under normal circumstances.

The face processing literature also provides evidence in support of other central tenets of the label-feedback hypothesis. These studies explore the ability to both categorise faces based on identity (e.g. Kikutani, Roberson & Hanley, 2008; Puri & Wojciuluk, 2008; Esterman & Yantis, 2010) and emotional expression (Roberson, Damjanovic & Pilling, 2007; Hanley & Roberson, 2011). The research cited below uses a variant of Etcoff and Magee's (1992) X-AB task to examine categorical perception of faces (e.g. Roberson, Davidoff & Braisby, 1999; Roberson & Davidoff, 2000). Here participants must assess whether face A or face B constitutes an identical match for a previously presented face (face X). Participants generally show superior identification of the target face (X) when the subsequently presented pair of faces (AB) represent a categorical mismatch, rather than a categorical match. We can consider an example in which target face 'X' represents the

identity 'Steve'. Here participants are better to select an identical representation of 'Steve' from the AB pair when 'Steves' face is paired with a picture of 'Kevin', rather than a second, perceptually different picture of 'Steve'. This is known as the 'cross-category' advantage (e.g. Harnard, 1987). Importantly, researchers demonstrate the robustness of the cross-category advantage by varying stimulus typicality. Morphing techniques are commonly used to merge the two endpoint identities away from their 'absolute' representations ( e.g. 100% 'Steve' and 100% 'Kevin'). When morphing has taken place, a cross-category pair may include faces that are only *proportionally* more representative of 'Steve' and 'Kevin', respectively (e.g. Face A = 60% 'Steve'/40% 'Kevin'; Face B = 60% 'Kevin'/40% 'Steve' ). In contrast, a within-category pair might include two faces that are both relatively more representative of 'Steve', then 'Kevin' (e.g. Face A = 70% 'Steve'/30% 'Kevin'; Face B = 80% 'Steve'/20% 'Kevin').

Kikutani et al. (2008) used this methodology to assess categorical perception for unfamiliar facial identities. A cross-category advantage was only found when participants were first briefly exposed to names for the two 'endpoint' identities on the continuum (e.g. 'Steve' and 'Kevin'), and not when training involved simple exposure to unlabelled versions of the endpoint identities. Kikutani et al. (2008) suggested that these (semantically-empty) names provided an additional source of evidence for identifying the category boundary that existed between the two identities. Arguably, names may offer facilitation because they boost our perception of the 'category-diagnostic' features, unique to each of the two facial identities.

Two studies may also demonstrate the existence and use of label-driven, *perceptual* expectancies when participants process facial identity (Puri & Wojciuluk, 2008; Esterman & Yantis, 2010). Participants showed enhanced processing of distorted, and briefly presented, pictures of faces when they followed exposure to a relevant vs. irrelevant cue word (Puri & Wojciuluk, 2008). Facilitated processing was most apparent when relevant cue-words were 'exemplar' or 'identity' specific (e.g. Goldie Hawn), rather than 'category-specific' (e.g. face). Puri and Wojciuluk (2008) suggested that 'identity/exemplar' specific cues were particularly useful because they created precise *perceptual* expectancies for the participant, both accentuating and guiding the identification of a subset of category-diagnostic features that we associate with 'Goldie Hawn'. However, these findings could instead be indicative of *semantic* facilitation; hearing the name 'Goldie Hawn' may activate

meaningful *semantic* knowledge about that actress, facilitating perception of her face. Esterman and Yantis' (2010) findings may be more compelling. Using a similar paradigm, but exclusively presenting 'category-specific' cues (e.g. 'face'), they found anticipatory activation in the superior temporal sulcus for face vs. house stimuli, perhaps indicative of a type of 'face-specific' *perceptual* expectancy.

Research conducted by Roberson et al. (2007) is particularly useful when considering how the label-feedback account may apply to the categorisation of emotional expression. Roberson et al. (2007) investigated whether the 'category adjustment model' (Huttenlocher, Hedges & Vevea, 2000) could be used to explain performance in an 'X-AB' task, when emotional faces were used as stimuli. Applying the assumptions of the category adjustment model, when participants covertly label initial face 'X' during encoding (e.g. 'sad'), they should activate an internally-held, prototypical representation of what a 'sad' face looks like. Similar to the assumptions held by the label-feedback hypothesis then, this prototypical representation will strongly emphasise what constitute 'category-diagnostic' features of that particular expression. Due to the inherent memory demands of the staggered 'X-AB' procedure, it was argued that this activated representation would bias memory for encoded face 'X' across the retention interval, pushing it closer towards the internally-held exemplar for that particular emotion category (Roberson et al., 2007). As well as being consistent with the commonly found 'cross-category' advantage (Harnard, 1987), this account would also predict a processing asymmetry for 'within-category' trials. To illustrate, when both faces in the AB pair broadly represent a category-level match for face 'X', participants should be more likely to select the face that presents a more 'typical' categorical exemplar, which is perceptually closer to the activated prototype (e.g. a morph of 80% 'sad'/20% 'fearful' vs. a morph of 60% 'sad'/40% 'fearful'). In support, Hanley and Roberson (2011) report evidence of 'within-category' asymmetries across a range of categorical perception studies, that used both face stimuli (identity e.g. Kikutani et al., 2008; expression e.g. Roberson & Davidoff, 2000) and colour stimuli (e.g. Roberson & Davidoff, 2000).

When considering the data from Roberson and Davidoff (2000) and Roberson et al. (2007), Hanley and Roberson (2011) also reported that within-category asymmetries were significantly weaker when interference manipulations were applied that (a) reduced the likelihood of verbal labelling, and therefore prototype activation, during the encoding of

face 'X', or (b) reduced maintenance of the verbal label, and prototype, across the interval between presentation of faces 'X' and 'AB'. As such, these findings suggest that internal activation of a prototypical exemplar is tied to the generation and retention of the verbal, category label (e.g. 'sad'). Arguably, Huttenlocher et al's. (2000) category-adjustment model represents a memory-based analogue for Lupyan's (2007, 2012) label-feedback hypothesis. That the former model can explain the way in which emotional faces are categorised within a memory-based task (the staggered X-AB paradigm), may suggest that the label-feedback hypothesis is relevant for understanding how we categorise emotional faces in tasks that demands online, *perceptual* processing of stimuli.

### **1.5 Summary: Reconciling the construction and label-feedback accounts**

Construction and label-feedback accounts differ in several critical ways. Primarily, they make different claims about the way in which *perceptual* features aid the categorisation of emotional expression. Constructionists argue that, across instances of emotional experience that can be similarly categorised (e.g. 'sadness'), there are no/few *perceptual* regularities in the associated emotional expressions shown. As such, *perceptual* examination of the structural information contained within the face is insufficient to drive discrete categorisation (e.g. Barrett et al., 2007; Mauss & Robinson, 2009). In contrast, if Lupyan's (2007, 2012) label-feedback account is applicable to the instance of emotion perception, then category-diagnostic features should exist in emotional expressions, and help to inform discrete categorisation. This crucial difference drives respective arguments about the role that language *needs* to play, in order to aid perception.

For Constructionists, language and its' referent (meaningful, *semantic* knowledge) have a difficult task to perform, transforming an uninformative facial expression into a recognisable instance of a discrete emotional state. A stored repository of *semantic* knowledge will play an active role in discrete categorisation, whether it is spontaneously activated to constrain perception during a decontextualized task, or compared against information inherent in the current emotional context (e.g. 'X had an exam today. X was unhappy the last time she took an exam, and felt she had performed poorly. X might be unhappy now too.'). For constructionists then, emotion perception is an inferential process, driven by stored *semantic* knowledge and *indirectly* supported by language.

Emotion perception may be a relatively simpler task if a label-feedback account is applicable (e.g. Lupyan, 2007, 2012). This account would assume that (at least some) emotional expressions contain category-diagnostic features, meaning that bottom-up visual processing is often sufficient for discrete categorisation. Language labels simply ‘boost’ perception, actively accentuating and deploying attention toward category-diagnostic, expressive features that make each facial expression unique. Therefore, while language is not necessary for emotion perception it may play a *direct* role when present, due to the strong links that exist between language labels and *perceptual*, or categorical forms, of *conceptual* knowledge.

### 1.5.1 The conceptual umbrella: flexible activation of semantic and perceptual forms of knowledge

Arguably both accounts may be true. The meaningful, *semantic* knowledge and *perceptual* knowledge we hold about a referent fall under the same umbrella. Both may be considered components of a *conceptual* knowledge base (e.g. James & Gauthier, 2003, 2004; Lupyan & Thompson-Schill, 2012; Casasanto & Lupyan, 2015). In this way, construction and label-feedback accounts may be making a similar argument; that perception of referent expressions or objects are supported by the activation of relevant *conceptual* knowledge. However, two further points exemplify the differences between a construction and label-feedback approach.

First, evidence suggests that meaningful, *semantic* and *perceptual* forms of *conceptual* knowledge may not be simultaneously activated nor play the same, potentially facilitative role, in object perception (Roberson et al., 1999; Lupyan & Mirman, 2013; see also James & Gauthier, 2003, 2004). Indeed, Lupyan (2009) report dissociations in performance in an odd-one-out task when participants used *semantic*/thematic (i.e. ‘select the object that does not function similarly/cannot be commonly found in the same context’) or *perceptual* rules (i.e. “select the object that is not typically ‘red’ in colour”), to select an object from a triad. Participants were asked to complete the task under concurrent verbal interference (rehearsal of a number string). Verbal interference led to selective decrements when participants attempted to use a *perceptual* grouping rule, but not a *semantic*/thematic rule. When instructed to use a *semantic*/thematic rule to select the ‘odd-one-out’, participants were similarly accurate under concurrent verbal and visual interference conditions (a visuo-



spatial interference task). This highlights a second argument; that *perceptual* and *semantic* forms of *conceptual* knowledge may differ in how strongly they rely on a categorical label for recruitment. The findings reported above may suggest that *semantic* knowledge can still be recruited when verbal interference reduces access to and use of verbal labels, but that *perceptual* forms of knowledge cannot (e.g. Lupyan, 2009; Lupyan & Mirman, 2013; see also Gauthier et al., 2003). This presents a dissociation in the way that language labels activate *conceptual* knowledge (e.g. Lupyan & Casasanto, 2014).

### 1.5.2 Task demands and preferential activation of different sources of knowledge

Various researchers suggest that labels will selectively activate different forms of *conceptual* knowledge in an ‘ad-hoc’ fashion; a process largely dependent on the demands posed by the current task (e.g. Roberson et al., 1999; Barsalou, Simmons, Barbey & Wilson, 2003; James & Gauthier, 2004; Elman, 2004, 2009; Casasanto & Lupyan, 2015). Indeed, in a training paradigm where participants learnt to associate unfamiliar objects with only one type of *conceptual* knowledge (e.g. auditory; motion-related), selective neural activation was found when those objects were then presented during a perceptual matching task (e.g. activation in auditory areas for objects trained alongside information about the sounds that object makes, James & Gauthier, 2003).

The label-feedback hypothesis (Lupyan, 2007, 2012) largely focuses on the role of language labels in helping us to ‘categorically’ perceive stimuli, as labels may accentuate the features that make an object part of one category, but differentiate it from members of another (Harnad, 1987; Lupyan & Ward, 2013; Collins & Olson, 2014). Selective activation of *perceptual* knowledge may then be most useful when label exposure precedes discrimination or *perceptual* matching tasks (deciding whether a pair of stimuli do or do not belong to the same category e.g. Lupyan et al., 2007; Lupyan & Casasanto, 2014). This may particularly be the case when the pair of stimuli are presented simultaneously for discrimination; a task which poses strong *perceptual* demands but should not benefit from activated *semantic* knowledge about the stimuli (e.g. James & Gauthier, 2004; Lupyan, 2008b; but see Gauthier et al., 2003 for an example of *semantic* knowledge vs. label-only benefits in a simultaneous object discrimination task). Instead, when pairs of stimuli are presented sequentially, participants may engage in *semantic* processing of the first item,

promoting deeper encoding and consolidation of a representation that can then be compared efficiently to the second stimulus within the pair (e.g. James & Gauthier, 2004). The extent to which a matching/discrimination task invites *semantic* processing may increase linear to the length of the inter-stimulus interval between presentation of items within that pair (e.g. Lupyan, 2008b; Lupyan, Thompson-Schill & Swingley, 2010). Preferential activation of *semantic* knowledge may also occur when single stimuli are presented and participants engage in identification, target verification or old/new decisions about that item. *Semantic* processing may be further increased, and *perceptual* processing decreased, if that stimulus is perceptually degraded using image blurring or filtering techniques (Abdel Rahman & Sommer, 2008; Abdel Rahman, 2011).

Although an exhaustive list of task demands has not been considered, it is clear that researchers should be mindful of the ways in which methodological components can bias the type of language-mediated conceptual knowledge that participants bring to the task. In the section below, we speculate on the possible ways in which language-mediated conceptual biases might be triggered by, and influence performance, in the work presented in this thesis.

### **1.5.3 Task demands and the activation of conceptual knowledge: Implications for the present work**

Within this thesis, semantic satiation (experiments 1-4) and perceptual threshold paradigms (experiment 5) will be used to assess how language might influence the interpretation and perception of emotional expression. Each paradigm possesses different methodological components, which may bias the type of *conceptual* knowledge that participants preferentially activate to complete that task.

In the semantic satiation experiments (1-4; Chapters 2 – 5) language-based manipulations are used to temporarily reduce access to *semantic* knowledge about emotion, before participants attempt to discriminate between a pair of emotional expressions (e.g. Lindquist et al., 2006). Findings from this strand of work are expected to indicate the degree to which *semantic* forms of *conceptual* knowledge are important for emotion perception under ordinary circumstances (i.e. when *semantic* knowledge is available).

However, when describing this work we speculate as to whether, in the absence of *semantic* knowledge, participants may strategically rely on other forms of *conceptual* knowledge to successfully perform discrimination.

The threshold paradigm (experiment 5), described in Chapter 6, poses a different type of critical decision. Following exposure to an emotion word, participants must decide whether a briefly presented face does or does not display an emotional expression. A cue-relevance effect is predicted; participants may be better at deciding whether a face is 'emotional' when they have previously encountered an emotion word that matches (vs. mismatches) the expression displayed by the face. These facilitative effects would be anticipated whether emotion-words activate semantic or perceptual forms of emotion-related conceptual knowledge. However, other methodological task components may bias language-driven activation of different forms of conceptual knowledge. On the one hand, presentation of a single face may encourage label-driven activation of *semantic* knowledge (e.g. Abdel Rahman & Sommer, 2008; Abdel Rahman, 2011). However, on the other, 'perceptually degrading' stimuli, using brief presentation methods, may drive preferential activation of *perceptual* forms of *conceptual* knowledge (e.g. Marsolek, Schyner, Ritchie & Verfaellie, 2006).

Of course, it is often difficult to delineate between the contribution of *semantic* and *perceptual* forms of *conceptual* knowledge to task performance when participants possess both types of knowledge about a referent. In the present case, people likely have access to, and experience of, using basic labels to name emotional states, know what a subset of emotional expressions are likely to look like (*perceptual* exposure) and have developed an experience-based repository of emotion-related *semantic* knowledge. As such, while findings from both strands of work can be discussed in the context of the construction (e.g. Barrett, 2006b) and label-feedback accounts (Lupyan, 2007, 2012) only tentative claims can be made about the type of *conceptual* knowledge that influences respective task performance. More importantly, employing tasks which vary in critical decision, and use different manipulations to potentially regulate recruitment of these separable forms of *conceptual* knowledge, allows for a more robust assessment of the role of language in emotion perception. The present work thus aims to contribute towards a growing body of literature that examines this question (e.g. Lindquist et al., 2006, 2014; Gendron et al., 2012; Nook et al., 2015).



## Chapter 2

### Investigating the role of language in the perception of emotional expression: does the semantic satiation paradigm provide a useful method?

#### 2.1 General Introduction

##### 2.1.1 Semantic satiation: paradigms and mechanisms

A modified semantic satiation paradigm was used to assess whether a construction account could explain the role of language in emotion perception (e.g. Lindquist et al., 2006; Gendron et al., 2012). Semantic satiation effects are usually assessed under an immediate repetition paradigm (Tian & Huber, 2010). Participants repeat a category label, out loud, (e.g. 'Royalty') 3 or 30 times either before judging the semantic relationship between (a) the repeated label and a single lexical exemplar (e.g. 'King'; Smith, 1984), or (b) a pair of lexical exemplars, independent of their semantic relationship with the repeated label (e.g. 'King' and 'Queen', Smith & Klein, 1990; Balota & Black, 1997; Black, 2001). Findings from both types of experiment show that participants are slower to make a semantic judgment after 30 vs. 3 repetitions of a category label, but only when that label is related to, or matches, the exemplar(s) presented at test (e.g. repeat the word 'Royalty' then judge the word 'King' vs. 'Piano'). That satiation decrements are specifically tied to this type of trial suggests that massed repetition results in a priming reversal; after repeating a related category label 30 times participants lose the performance advantage they show after 3 repetitions of that label (e.g. Smith & Klein, 1990; Black, 2001, though see Tian & Huber, 2010).

Tian and Huber (2010) rigorously explored the mechanisms that produce satiation effects. Their investigation ruled out two possible explanations. First, that the node representative of repeated word becomes fatigued and can no longer be used to access related semantic nodes ('lexical satiation'; e.g. Esposito & Pelton, 1971) and second, that the repeated word itself remains intact, but the consistently re-activated semantic nodes, related to word meaning, are inhibited ('meaning satiation'; e.g. Smith & Klein, 1990; Balota & Black, 1997).

Instead Tian and Huber's (2010) results suggest that, while both the repeated word and related semantic knowledge remain intact, the connection between the elements is temporarily severed, slowing the spread of excitation between related nodes. This associative account supports the sluggish link hypothesis; that the satiation manipulation creates habituation that slowly spreads to inhibit access to related nodes and connections (Smith, 1984).

Lewis and Ellis (2000a) suggest that satiation effects are not restricted to word stimuli. These researchers showed similar increases in decision time when participants repeated a celebrities name out loud 30 (vs 3) times before deciding whether a subsequently presented photograph depicted that celebrity. Similar findings were obtained when Lewis and Ellis (2000a) repeatedly flashed images of a target celebrities face, before showing participants a photograph of a different celebrity, who did or did not possess a semantic relationship with the target (e.g. had co-starred in a popular film). Following 30 (vs. 3) face repetitions, participants were only slower to judge the semantic relationship between the two celebrities when they were semantically related (vs. unrelated) to one another. These experiments demonstrate that satiation is a general property of the semantic system (Lewis & Ellis, 2000a, though see Lewis & Ellis, 1999, 2000b); independent of the stimuli used to prime access to a semantic network, massed exposure to that stimulus inhibits the spread of excitation between the nodes representative of that stimulus and related knowledge.

### **2.1.2 Semantic satiation and emotion perception: Lindquist et al. (2006)**

Lindquist et al. (2006) made a novel attempt to empirically falsify a construction theory of emotion perception. To do so, they designed a semantic satiation paradigm that borrowed methodological elements from both Lewis and Ellis (2000a) and Smith and Klein (1990). In Lindquist et al. (2006, experiments 2 and 3), participants were required to repeat a basic emotion label out loud (e.g. 'Sad'), either 3 or 30 times, before deciding whether two emotional faces matched or mismatched each other in emotional expression (a simultaneous, perceptual discrimination task). According to the construction position, the interpretation of emotional expression relies on access to an internal repository of emotion-related, semantic knowledge (e.g. Barrett et al., 2007, Lindquist, 2009; Gendron et al., 2012). A perceptual discrimination task may place high demands on this system. Here participants need to meaningfully interpret each of the two facial expressions before deciding whether the degree of conceptual overlap between the two emotional states

warrants a match or mismatch decision. This is similar to the implicit decision, posed in previous research; that is, judging the semantic relationship between a pair of words, irrespective of their relationship to an initially repeated word (e.g. Smith & Klein, 1990; Balota & Black, 1997).

Lindquist et al. (2006) predicted that participants would show slowed or less accurate perceptual discrimination following 30 repetitions, or satiation, of an emotion label. This would reflect loss of access to decision-necessary semantic knowledge. Following previous semantic satiation research, Lindquist et al. (2006) also made a further prediction, based on their manipulation of 'word relevance'. They suggested that the satiation decrement would be strongest after massed repetition of an emotion label that matched, or was relevant to the judgment of both expressions presented for discrimination (e.g. repeat the word 'sad', then see two 'sad' faces). Here semantic knowledge discretely associated with those particular expressions would be rendered inaccessible, making it very difficult for participants to make a meaningful comparison. These specific predictions comprise a 'category-based' account (Lindquist et al., 2006). Performance in these trials would be explicitly compared against those in which the repeated label matched just one of the expressions displayed. Previous work using word pairs confirm that differences in performance after 30 vs. 3 repetitions of a judgment-relevant label are most pronounced when both stimuli in a pair are strongly related to the repeated label, rather than just one (e.g. Smith & Klein, 1990). As such, these trials are thought to comprise the 'satiation comparison' condition; a term coined by Lindquist et al. (2006) and used throughout the present thesis.

Lindquist et al's. (2006, Experiments 2 and 3) findings offer general support for their predictions; to illustrate, the descriptive statistics and main findings from Lindquist et al's. (2006) third experiment are displayed below, in Tables 2.1 and 2.2<sup>2</sup>. A set of apriori,

---

<sup>2</sup> For clarity, only the descriptive statistics and main findings from Lindquist et al's. (2006) third experiment are reported in Tables 2.1 and 2.2. The design of Lindquist et al's. (2006) second experiment differed from that used in the present semantic satiation work; here the 'expression matching' factor was omitted and the 'word relevance' factor had an additional level, producing a 2(repetition: 3 vs. 30) by 3 (word relevance: the repeated emotion label matches both expressions vs. one expression vs. neither expression) design. In addition, Lindquist et al. (2006) reported reaction times (vs accuracy rates) as the main dependent variable in their second study. This may

targeted comparisons showed that participants were both slower (experiment 2) and less accurate (experiments 2 and 3) in their discriminations after engaging in massed repetition of a ‘judgment-relevant’ (vs. irrelevant) emotion label, which matched both of the faces presented at test (e.g. the ‘satiation comparison’; see Tables 2.1 and 2.2). Although supportive of a ‘category-based’ construction account, these selective trends in the latency and accuracy data were weak. In contrast, there were robust trends in data from both experiments to suggest that participants were slower and less accurate in their decisions after 30 (vs. 3) repetitions of ANY emotion label (judgment-relevant or irrelevant), whether it matched both, one or none of the emotional expressions displayed at test (see Tables 2.1 and 2.2).

**Table 2.1 Mean proportional accuracy rates reported by Lindquist et al. (2006, experiment 3, adapted from Table 7).**

*Proportional accuracy rates are a function of repetition, word relevance and expression matching (SEM presented in parenthesis).*

Repetitions	Relevant Label		Irrelevant Label		Marginal Means
	Match	Mismatch	Match	Mismatch	
<b>3</b>	0.42 (0.040)	0.31 (0.030)	0.21 (0.030)	0.28 (0.040)	0.31 (0.030)
<b>30</b>	0.36 (0.040)	0.23 (0.030)	0.23 (0.040)	0.25 (0.030)	0.27 (0.030)
<b>Marginal Means</b>	0.39 (0.040)	0.27 (0.030)	0.22 (0.030)	0.26 (0.030)	-

reduce the utility of comparing the present results to those obtained in Lindquist et al’s. (2006) second experiment.



**Table 2.2 Results of Lindquist et al's. (2006) third experiment.**

*Results from the 2 (Repetition: 3 vs. 30) × 2 (Word Relevance: Relevant vs. Irrelevant) × 2 (Expression Matching: Match vs. Mismatch) within-participants ANOVA, conducted on accuracy rates. Pertinent post-hoc results are included and significant effects ( $p < .05$ ) emboldened.*

Effect	Statistic	DF	Value (F,t)	p	Effect size
<b>Repetition</b>	<b>F</b>	<b>(1, 35)</b>	<b>5.12</b>	<b>.03</b>	<b>0.076</b>
Expression Matching	F	(1, 35)	1.81	.19	0.020
<b>Word Relevance</b>	<b>F</b>	<b>(1, 35)</b>	<b>36.78</b>	<b>&lt;.001</b>	<b>0.001</b>
<b>Repetition × Word Relevance</b>	<b>F</b>	<b>(1, 35)</b>	<b>5.77</b>	<b>.02</b>	<b>0.053</b>
Repetition × Expression Matching	F	(1, 35)	1.29	.26	0.029
<b>Expression Matching × Word Relevance</b>	<b>F</b>	<b>(1, 35)</b>	<b>36.32</b>	<b>&lt; .001</b>	<b>0.010</b>
Repetition × Word Relevance × Expression Matching	F	(1, 35)	0.59	.45	<0.001
'Satiating comparison'	t	(35)	1.71	.10	unreported
Irrelevant word, 3 vs. 30 repetitions	t	(35)	-0.62	.54	unreported

Lindquist et al. (2006) acknowledged trends in their work which suggested that the massed repetition decrement was common both to the repetition of relevant and irrelevant emotion words. Lindquist et al. (2006) implicated the possible role of spreading activation/inhibition, commonly associated with semantic satiation outcomes (e.g. Collins & Loftus, 1975; Neely, 1977; Smith, 1984; Smith & Klein, 1990; Balota & Black, 1997). Researchers note that the suppression caused by massed repetition may only affect semantic nodes and/or connections very closely related to the repeated label early in the time-course of the effect. Following a transitory path, suppression may also affect more distant semantic nodes and/or connections, related to the category label at a superordinate level. In support, previous semantic satiation work has evidenced stronger satiation effects for associated items ('Wool') vs. repeated items ('Sheep') when both were presented during a delayed decision task (Kuhl & Anderson, 2011, see also Balota & Black, 1997).

Applied to the present work, massed verbal repetition of the label 'Sad' may only reduce access to semantic knowledge related to 'sadness' in the early stages of inhibition. Effects may later spread to inhibit the associated nodes and connections necessary for accessing

semantic knowledge about other emotional states. In particular, if valence is viewed as a superordinate grouping factor, then semantic knowledge about all discrete states with negative connotations might share links in a distributed network. This is entirely possible as Lindquist et al. (2006) atypically modified the semantic satiation paradigm, using only 'related' primes (negative emotion labels) and providing no baseline/'unrelated' trials (e.g. Smith & Klein, 1990; discussed further below). Here satiation of any negative emotion label might cause performance decrements through inhibition of nodes and links associated with 'negative' semantic knowledge.

However, alternative, non-semantic explanations may exist for the general, massed repetition decrement. Immediate repetition paradigms are often criticised because the 3 and 30 repetition conditions place unequal cognitive demands on the participant (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013). For example, participants may experience a greater degree of intra-trial, cognitive fatigue after 30 vs. 3 verbal repetitions, selectively slowing their responses (e.g. Saito, 1997; Tian & Huber, 2010). This may particularly be the case when a 'blinking technique' is used, whereby a new orienting response is created each time the word is repeatedly flashed on screen (Smith, 1984). Fatigue may damage performance by causing general disruptions to semantic processing; however, this would be an 'indirect' semantic explanation, differing from that associated with spreading activation (e.g. Smith & Klein, 1990; Black et al., 2013). In addition, participants may experience a higher degree of uncertainty about when a response is required from them in 30 (vs 3) repetition trials (e.g. Black, 2001; Tian & Huber, 2010). Participants likely develop an anticipatory advantage in response to 3 repetition trials; when a small number of items are presented, either simultaneously or sequentially, participants may be able to automatically process or 'subitise' the number of repetitions required before a decision (e.g. Camos & Tillmann, 2008). In other words participants likely find it very easy to predict when a response is required on half of the trials they encounter during the immediate repetition paradigm.

Previous immediate repetition paradigms include 'control' or 'unrelated' trial types to assess whether non-semantic factors can plausibly explain the massed repetition decrement (e.g. participants respond to the pair 'apple' and 'pear' after satiating the unrelated label 'royalty'; Smith & Klein, 1990; Balota & Black, 1997; Tian & Huber, 2010). In other words, while semantic explanations are possible when massed repetition decrements

occur only after repetition of ‘related’ words, other non-semantic factors are implicated when decrements are common to massed repetition of related/relevant and unrelated/irrelevant words (e.g. cognitive fatigue, response uncertainty, loss of an anticipatory advantage, Black, 2001; Tian & Huber, 2010; Black et al., 2013). Such comparisons are not possible in Lindquist et al.’s. (2006) work, as participants repeated emotion labels on every trial (whether they were ‘relevant’ or ‘irrelevant’ to expression discrimination). Rather than spreading activation, the observed massed repetition decrements may then be a consequence of alternative non-semantic factors.

In the first instance a direct methodological replication of Lindquist et al.’s. (2006, experiment 3) was conducted to assess the robustness of their results. It was important to examine whether participants show massed repetition decrements after 30 (vs 3) repetitions of (a) a judgment-relevant emotion label (matching at least one of the expressions shown at test), or (b) all emotion labels (relevant and irrelevant to discrimination). If performance decrements are inconsistently tied to massed repetition of judgment-relevant emotion labels then this would strengthen the need to include baseline/unrelated trials in subsequent work (e.g. Smith & Klein, 1990; Balota & Black, 1997). This subsequent work would explicitly assess the role of non-semantic factors (see Chapters 3 and 4).

## **2.2 Experiment 1: Replicating Lindquist et al. (2006, experiment 3)**

### **2.2.1 Introduction**

A methodological replication of Lindquist et al. (2006, Experiment 3) was conducted. Per trial, participants always repeated emotion labels. Label-relevance was manipulated across trials, so that repeated labels matched both emotional expressions, one expression or neither expression presented at test. It was important to assess whether repetition decrements occur after massed repetition of all emotion labels, or only those that are judgment-relevant (e.g. matching at least one of the expressions displayed at test, regardless of whether the faces themselves matched each other in expression). Following Lindquist et al. (2006, Experiment 3), participants were also given a very short response window in which to view and respond to each face pair (854ms). Inhibition of the semantic network, via satiation, is a transient effect. Although evidence concerning longevity of the effect is mixed, an active range between 400-900ms is considered likely

(e.g. Frenck-Mestre, Besson & Pynte, 1997; Kounios, Kotz & Holcomb, 2000; Gendron et al., 2012). Therefore, in order to assess whether semantic information is important for the initial perception of facial expression it was important that discrimination occurred while inhibition was active. Use of a short response window would also force participants to favour speed over accuracy in their decisions, reducing the likelihood of a speed-accuracy trade-off in the data (see Lindquist et al., 2006, Experiments 1 and 2) and reducing performance from ceiling (e.g. Draine & Greenwald, 1998). If participants favour speed over accuracy when responding, then satiation differences are more likely to be observed in accuracy rates, rather than reaction times (see data preparation section below)<sup>3</sup>.

Replication was an important first step as Lindquist et al. (2006) report an inconsistent pattern of results (see Tables 2.1 and 2.2). Across experiments (2 and 3) they suggested that increased latencies and decreased accuracy were most commonly observed after participants had repeated a judgment-relevant emotion label 30 (vs 3) times. These findings are consistent with a category-based, construction view; in order to interpret 'sad' expressions we require access to semantic knowledge discretely related to that emotional state. Arguably, trends in Lindquist et al.'s. (2006) data are more consistent with a general massed repetition decrement; participants tended to show increased latencies and decreased accuracy after 30 (vs. 3) repetitions of any emotion label, whether it was relevant to both, one or neither of the expressions displayed at test. Lindquist et al. (2006) implicated the mechanism of spreading activation (e.g. Collins & Loftus, 1975; Neely, 1977; Smith, 1984) to explain how 'global' decrements remained consistent with their construction position. However, it has been argued that other non-semantic explanations are equally plausible when using an immediate repetition paradigm. For example, after 30 vs 3 verbal repetitions of any emotion label participants may experience heightened, intra-trial, cognitive fatigue or may lose an anticipatory response advantage, associated with 3-

---

<sup>3</sup> In their second experiment, Lindquist et al. (2006) implemented a response window of 1,400ms. This was the median time taken participants took to make an accurate discrimination during their first experiment. Finding that this response window was too long to force a speed-accuracy trade-off, Lindquist et al. (2006) reduced the duration to 854ms for experiment 3. Key satiation differences in experiment 3 were only observed in the accuracy, rather than latency data, leading Lindquist et al. (2006) to conclude that this shorter response window was appropriate to create a speed-accuracy trade-off (e.g. Draine & Greenwald, 1998).

repetition trials (e.g. Smith & Klein, 1990; Black, 2001; Tian & Huber, 2010; Black et al., 2013).

The following data-driven predictions are made. First, the construction position (category-based account) would be supported by an interaction between repetition and word relevance. Specifically, this interaction should be driven by a difference in accuracy after 30 vs. 3 repetitions of a judgment-relevant emotion label (lower accuracy after massed repetition). There should be no difference in accuracy after 3 and 30 repetitions of an irrelevant emotion label. Focusing on trials in which judgment-relevant labels are repeated, planned comparisons will be used to assess whether any differences in accuracy following massed repetition are driven by performance in the 'satiation comparison' trials (when the repeated label matches both expressions at test). Consistent with a category-based account, participants should be less accurate to perform discriminations when knowledge discretely associated with both facial expressions is made inaccessible (see Smith & Klein, 1990 for analogous findings with word pairs). A second comparison will assess whether repetition-based differences arise in relevant, mismatch trials (when the repeated label matches just one expression at test). A construction account would predict comparatively smaller repetition-based decrements in these trials, as discrete semantic knowledge is rendered inaccessible for one face, only. Second, finding a main effect of repetition (common to 30 repetitions of both judgment-relevant and irrelevant emotion labels) would be consistent both with the construction position (spreading activation, Collins & Loftus, 1975; Neely, 1977) and several alternate, non-semantic accounts (e.g. cognitive fatigue, Black, 2001; Tian & Huber, 2010).

## **2.2.2 Method**

### **2.2.2.1 Participants**

Participants were 48 undergraduate students from the University of Leeds (45 females), who participated for course-related credits. The sample had a mean age of 19.63 years ( $SD = 1.12$ ; Range: 18-23). All participants were native English speakers and 43 were right-handed.

## 2.2.2.2 Materials

### 2.2.2.2.1 Words

Replicating Lindquist et al. (2006), four ‘basic’ emotion labels were used: ‘Sad’, ‘Fear’, ‘Anger’ and ‘Disgust’ (e.g. Ekman, 1992). These labels were equally sampled throughout the task. Words were always presented to participants in the centre of a white background in a black font, at point 18.

### 2.2.2.2.2 Face Pairs

Replicating Lindquist et al. (2006), facial identities were selected from the Pictures Of Facial Affect inventory (Ekman & Friesen, 1976). This inventory contains photographs of 13 actors adopting prototypic facial expressions and depicts the head and neck region, only. Four facial identities were chosen (two male and two female), each displaying the four emotional expressions of interest (‘Sad’, ‘Fear’, ‘Anger’ and ‘Disgust’) <sup>4</sup>.

A series of face pairs were created for the perceptual discrimination task. Pairings where both faces displayed the same emotional expression were termed ‘expression match’ pairs (e.g., both faces depicted a ‘Sad’ expression). Pairings could be mixed for sex but the same facial identity was never displayed within a pair. Given the four unique identities, six possible expression match pairings could be constructed for each emotion category (24 expression match pairs in total). The position of the photograph in the pair (left or right) also contributed to trial uniqueness. Switching the left/right positioning of the original pairs produced an additional 24 pairs (12 per emotion category, 48 pairs in total). Similarly, all possible, unique ‘expression mismatch’ pairs were calculated. Two restrictions were applied. First, that the two faces did not display the same facial expression; and second, that the two faces were not of the same facial identity. Again, trials were considered

---

<sup>4</sup> Validation data, accompanying the Pictures of Facial Affect, suggest that the participants commonly categorise the selected identities as displaying the intended emotional expressions (Ekman & Friesen, 1976). Averaging across all four expressions of interest, female identity NR was accurately categorised by 90.25% of participants (SD = 8.18); female identity MF (M = 91.75%, SD = 5.68); male identity PE (M = 89.25%, SD = 4.27); and male identity JJ (M = 88.25%, SD = 8.81).

unique if the two same photographs were displayed, but in a different left/right position. This resulted in a pool of 96 unique combinations of expression mismatch pairs, that could be drawn from when constructing trial pairs.

Faces within a pair were each sized to 9cm x 8cm. At test, expression pairs were displayed on a computer monitor, side by side, on a white background.

### **2.2.2.3 Apparatus**

All of the semantic satiation experiments (1, 2a, 2b, 3 and 4; Chapters 2–4) were conducted using E-Prime 2.0 professional software (Psychology Software Tools, Pittsburgh, PA)

### **2.2.2.4 Design**

A 2 (Word Relevance: Relevant vs. Irrelevant) × 2 (Repetition: 3 vs. 30) × 2 (Expression Matching: Match vs. Mismatch) within-participants design was used. Participants completed 128 trials in total, split equally across two trial blocks. There were 8 conditions and 16 trials per condition; trials from each condition were distributed equally across the two trial blocks, and presented in a random, intermixed order. The four different emotional expressions were also sampled equally across the experiment.

An equal number of expression match and mismatch pairs were assigned to four arbitrary face sets (16 each, 32 pairs in total). The allocation of expression pairs to sets was pseudo-random and attempts were made to avoid repetition of expression pairs across the four sets. Though in some cases this was unavoidable as there were an unequal number of unique matching and mismatching expression pairs (there were less expression match pairs). Across participants, these face sets were systematically rotated around the four different conditions, derived from levels of the remaining within-participant variables (Repetition and Word Relevance). Face set and condition pairings were counterbalanced and four possible orders of pairing were created. An equal number of participants ( $n = 12$ ) encountered each of the four possible face sets, paired with the same conditions.

Examples of each of the eight possible trial types are shown in Table 2.3, below. Three different proportional (accuracy) measures were calculated as dependent variables, per condition (see data preparation section, pages 37–38).

**Table 2.3 Examples of the eight trial types used in experiment 1 (adapted from Lindquist et al. 2006).**

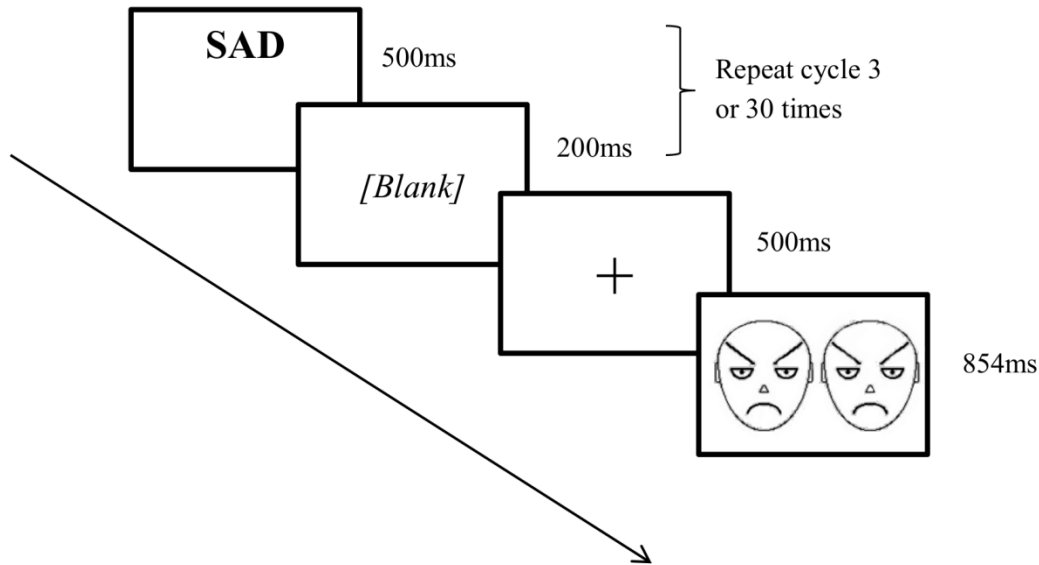
<i>Word Repetition</i>	<i>Type of Emotion Word</i>	<i>Expression Match</i>	<i>Trial Example</i>
<b>3</b>	<b>Relevant</b>	<b>Match</b>	3 repetitions of the word 'Sad'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Sad'; compare a 'Sad' and 'Angry' expression pair
	<b>Irrelevant</b>	<b>Match</b>	3 repetitions of the word 'Fear'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Fear'; compare an 'Angry' and 'Sad' expression pair
<b>30</b>	<b>Relevant</b>	<b>Match</b>	30 repetitions of the word 'Anger'; compare two 'Angry' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Anger'; compare an 'Angry' and 'Disgusted' expression pair
	<b>Irrelevant</b>	<b>Match</b>	30 repetitions of the word 'Disgust'; compare two 'Fearful' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Disgust'; compare an 'Angry' and 'Fearful' expression pair



### 2.2.2.5 Procedure

A perceptual discrimination task was used. During a single trial an emotion label was flashed on the middle of the screen, and participants were required to repeat this label, out loud, each time that it appeared. The word was presented either 3 or 30 times. Per repetition the label was displayed for 500ms, with a 200ms inter-stimulus interval (blank screen) before it was repeated. Following 3 or 30 word repetitions, participants viewed a blank screen for 200ms, followed by a fixation cross for 500ms, before they were required to complete the perceptual discrimination task. Here participants decided whether two photographed faces matched or mismatched each other in emotional expression. Participants were instructed to base their matching decisions solely on the relationship that existed between the two faces. In other words, decisions should be made irrespective of any shared/unshared emotional content between the repeated label and those faces. They were also encouraged to respond while the faces were still displayed on the screen (full participant instructions are presented in Appendix A). Participants indicated a 'match' or 'mismatch' decision using the '1' and '9' keys on the keyboard. Decision-key mapping and hand dominance were counterbalanced across participants. Participants were not required to switch response keys at any time during the task. The next trial began immediately after the 854ms response window had elapsed, regardless of whether the participant had submitted a response. The temporal sequence for a single trial is displayed in Figure 2.1.

Participants completed 8 practice trials before beginning the task, which included examples of each type of trial. Two different facial identities, selected from the POFA (Ekman & Friesen, 1976), were used in the practice compared to experimental trials.



**Figure 2.1 Example of the temporal sequence for an judgment-irrelevant label, expression match trial.**

*For reasons of stimulus reproducibility, schematic faces have been used in this figure. Participants encountered pairs of photographed stimuli from the Pictures of Facial Affect inventory throughout experiments 1, 2a, 2b, 3 and 4 (Ekman & Friesen, 1976).*

## 2.2.3 Results

### 2.2.3.1 Data preparation

Three accuracy totals were calculated, per participant. Each total accounts for different proportions of the three types of available response: a correct response (responding 'match' when facial expressions 'matched'), an incorrect response (responding 'match' when facial expressions 'mismatched') and a timed out response, where the participant did not respond within the 854ms window. First, a 'proportional' accuracy variable was calculated, which followed Lindquist et al's. (2006) operationalization. Here the proportion of incorrect and timed out responses were conflated to produce a 'total' error score. Accuracy was then expressed as a proportion of the total number of trials included in the experiment, using the following calculation:  $\text{Number of correct responses} / (\text{number of correct} + \text{incorrect} + \text{timed out responses})$ . Second, a 'conditional' accuracy variable was calculated. Here only responses submitted within the short response window were considered; correct responses were expressed as a proportion of the total number of trials to which participants had given a response [ $\text{Number of correct responses} / (\text{number of}$

correct responses + incorrect responses)]. Third, we calculated a ‘timed out trial’ accuracy variable. Here we focused on trials during which the participant had failed to submit a response within the time window. These trials were considered as a proportion of the total number of trials that the participant had encountered [expressed as: Number of timed out responses/ (number of timed out responses + correct responses + incorrect responses)]. In this experiment participants were timed out in response to 21.27% of all trials.

It was important to calculate multiple accuracy variables for methodological reasons. When taking into account all three types of response, use of a constrained response window is often associated with chance-level accuracy rates, as it forces the participant to make errors (e.g. Draine & Greenwald, 1998). Arguably, the conditional accuracy analysis considers only those trials in which participants demonstrably favoured speed over accuracy when responding. The results of this analysis will inform us whether the responses made by participants (proportion correct vs incorrect) were typically at chance level. A comparison of trends across the proportional and conditional accuracy analyses can be made to assess the extent to which a restricted response window forces two types of error; incorrect and timed out responses. This is not possible when there exists just one accuracy variable, which conflates the contribution of both errors, as in Lindquist et al. (2006). Across experiments (1, 2a, 2b, 3 and 4), single-samples t-tests were conducted to assess whether proportional and conditional accuracy rates, per condition, differed significantly from chance level (50% accurate).

### **2.2.3.2 Main Analyses**

Separate 2 (Word Relevance: Relevant vs. Irrelevant) × 2 (Repetition: 3 vs. 30) × 2 (Expression Matching: Match vs. Mismatch) within-participants ANOVAs were conducted for all three types of accuracy variable. The results of the proportional accuracy analysis are first reported, which replicate Lindquist et al’s. (2006) operationalisation of accuracy. Then analyses conducted on the conditional accuracy data and timed out trial data are reported.

For each of these analyses, effects and planned comparisons are first reported that address the predictions outlined in the introduction. The repetition x word relevance interaction is specifically examined to assess whether 30 vs 3 repetitions of a judgment-relevant (vs. irrelevant) emotion label lead to decrements in discrimination performance. Confirmatory

evidence from this interaction would support a category-based, construction position (Lindquist et al., 2006; Gendron et al., 2012). In particular, finding that repetition decrements are tied to the massed repetition of judgment-relevant, emotion labels when that label matches both of the emotional expressions displayed at test (i.e. the 'satiation comparison' trials; Smith & Klein, 1990; Lindquist et al., 2006, see Tables 2.1 and 2.2). In contrast, alternative semantic (spreading activation) or non-semantic explanations (cognitive fatigue, loss of an anticipatory advantage) will be favoured if decrements follow both massed repetition of judgment-relevant and irrelevant emotion labels (a main effect of repetition). Following the report of predicted effects and planned comparisons, we present additional significant and marginal effects. For both a-priori and additional analyses we consider effects significant if  $p < .05$ , and marginal if  $p \leq .06$  and  $> .05$ . Mean proportional accuracy, conditional accuracy and timed out trial rates, per condition, are displayed in Table 2.4, below.

**Table 2.4 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Relevance and Expression Matching (SEM presented in parenthesis).<sup>5</sup>**

Dependent Variable	Repetitions	Relevant Label		Irrelevant Label	
		Match	Mismatch	Match	Mismatch
Proportional Accuracy	3	0.53 (0.020)	0.47 (0.026)	0.46 (0.025)	0.51 (0.029)
	30	0.47 (0.027)	0.42 (0.029)	0.43 (0.025)	0.50 (0.026)
Conditional Accuracy	3	0.64 (0.026)	0.60 (0.023)	0.58 (0.026)	0.63 (0.029)
	30	0.59 (0.035)	0.56 (0.036)	0.56 (0.033)	0.63 (0.029)
Timed out responses	3	0.16 (0.020)	0.21 (0.030)	0.20 (0.030)	0.20 (0.028)
	30	0.24 (0.032)	0.25 (0.035)	0.23 (0.032)	0.22 (0.031)

### 2.2.3.3 Proportional Accuracy Analysis

#### 2.2.3.3.1 Repetition-based effects (predicted)

There was a significant main effect of repetition. Participants made fewer correct responses after 30 repetitions of any emotion label ( $M = 0.46$ ,  $SEM = 0.014$ ) than after 3 repetitions ( $M = 0.49$ ,  $SEM = 0.017$ ),  $F(1,47) = 8.61$ ,  $p = .005$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.16$ .

<sup>5</sup> Single-samples t-tests revealed that five of the condition means for proportional accuracy did not differ significantly from chance level (0.50): the judgment-relevant, 3 repetition, expression match condition ( $p = .11$ ); the judgment-relevant, 3 repetition, expression mismatch condition ( $p = .24$ ); the judgment-relevant, 30 repetition, expression match condition ( $p = .32$ ), the judgment-irrelevant, 3 repetition, expression mismatch condition ( $p = .64$ ) and the judgment-irrelevant, 30 repetitions, expression mismatch condition ( $p = .96$ ).

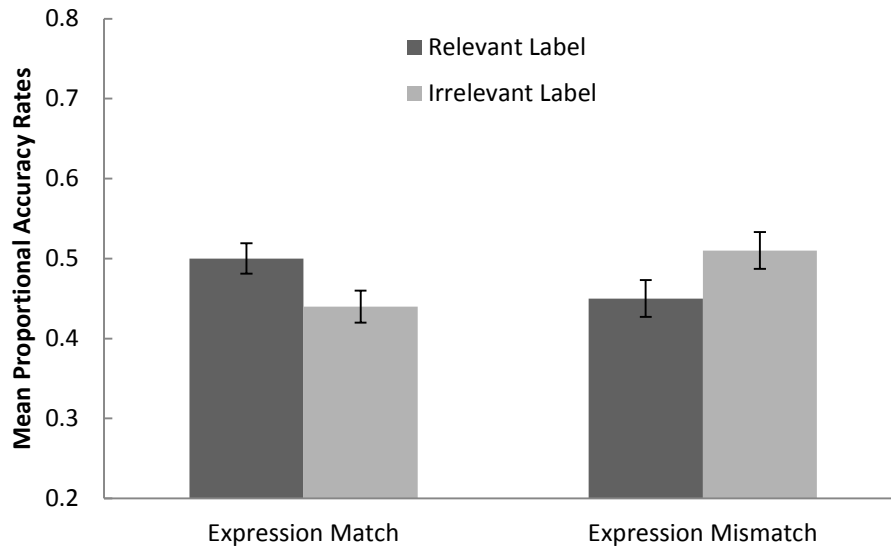
In addition, proportional accuracy means for two conditions fell significantly below chance levels: the judgment-relevant, 30 repetition, expression mismatch condition ( $p = .009$ ) and the judgment-irrelevant, 30 repetition, expression match condition ( $p = .025$ ). We discuss the potential implications of interpreting results driven by close-to-chance or below-chance levels in Chapter 5, pages 123–125.

There was no interaction observed between word relevance and repetition,  $F(1,47) = 2.96$ ,  $p = .092$ ,  $MSE = 0.012$ ,  $\eta_p^2 = 0.059$ . However, planned comparisons showed that participants gave significantly fewer correct responses after 30 repetitions of a judgment-relevant emotion label ( $M = 0.45$ ,  $SEM = 0.020$ ) than after 3 repetitions ( $M = 0.50$ ,  $SEM = 0.014$ ),  $t(47) = 3.43$ ,  $p < 0.001$ ,  $d = 0.42$ . In contrast, participants did not differ in their ability to provide correct responses after 30 repetitions of an judgment-irrelevant emotion label ( $M = 0.47$ ,  $SEM = 0.022$ ), than after 3 repetitions ( $M = 0.48$ ,  $SEM = 0.019$ ),  $t(47) = 0.95$ ,  $p = .35$ .

Targeted analysis of the 'satiation comparison' conditions revealed that participants gave significantly less correct responses after 30 repetitions of a judgment-relevant emotion label, than after 3 repetitions, when that label matched both of the expressions presented  $t(47) = 2.12$ ,  $p = .039$ ,  $d = 0.36$  (means presented in Table 2.4, in rows 1 and 2 of column 1). When the repeated label matched only one of the presented faces, participants did not differ in their ability to give a correct response following 30 and 3 repetitions,  $t(47) = 1.83$ ,  $p = .074$  (means presented in Table 2.4, in rows 1 and 2 of column 2).

#### 2.2.3.3.2 *Other significant and marginal effects*

There was a significant interaction between word relevance and expression matching:  $F(1, 47) = 14.68$ ,  $p < .001$ ,  $MSE = 0.023$ ,  $\eta_p^2 = 0.24$ .



**Figure 2.2 Mean proportional accuracy rates, by word relevance and expression matching (error bars represent SEM).**

There were two facets to this interaction. Paired samples t-tests showed that, after repeating a judgment-irrelevant emotion label, participants were significantly more likely to correctly respond to mismatching ( $M = 0.51$ ,  $SEM = 0.022$ ) vs. matching expressions pairs ( $M = 0.44$ ,  $SEM = 0.023$ ),  $t(47) = 2.25$ ,  $p = .016$ ,  $d = 0.46$ . However, after repeating a judgment-relevant emotion label, participants were similarly likely to give a correct response when they encountered mismatching expression pairs ( $M = 0.45$ ,  $SEM = 0.023$ ) and matching expression pairs ( $M = 0.50$ ,  $SEM = 0.019$ ),  $t(47) = 1.89$ ,  $p = .065$ .

In response to expression match pairs, participants were significantly more likely to give correct responses after repeating a judgment-relevant emotion label, than an irrelevant label,  $t(47) = 2.42$ ,  $p = .020$ ,  $d = 0.42$ . The opposite was true for expression mismatch pairs. Participants were significantly more likely to submit correct responses after repeating a judgment-irrelevant vs. relevant emotion label,  $t(47) = 3.51$ ,  $p < .001$ ,  $d = 0.39$ .

### 2.2.3.4 'Conditional Accuracy' Analysis

#### 2.2.3.4.1 Repetition-based effects (predicted)

There was no main effect of repetition,  $F(1, 47) = 2.96$ ,  $p = .092$ ,  $MSE = 0.027$ ,  $\eta_p^2 = 0.059$ . Participants were similarly accurate to respond following 30 repetitions ( $M = 0.59$ ,  $SEM = 0.033$ ), and 3 repetitions, of an emotion label ( $M = 0.61$ ,  $SEM = 0.026$ ).

There was no interaction observed between repetition and word relevance:  $F(1, 47) = 1.35$ ,  $p = .25$ ,  $MSE = 0.018$ ,  $\eta_p^2 = 0.028$ . However, planned comparisons (paired samples t-tests) showed that participants were marginally less accurate following 30 repetitions of a relevant emotion label ( $M = 0.58$ ,  $SEM = 0.019$ ) than 3 repetitions, ( $M = 0.62$ ,  $SEM = 0.016$ ),  $t(47) = 1.97$ ,  $p = .060$ . In contrast, conditional accuracy rates did not differ after participants had repeated an judgment-irrelevant emotion label 30 times ( $M = 0.59$ ,  $SEM = 0.019$ ) vs. 3 times ( $M = 0.61$ ,  $SEM = 0.019$ ),  $t(47) = 0.63$ ,  $p = .54$ .

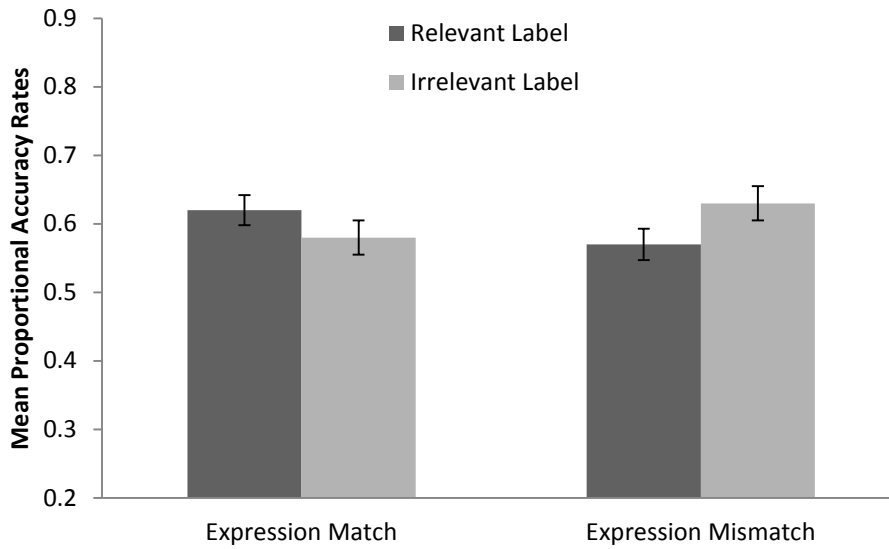
Specific analysis of the 'satiation comparison' conditions revealed no significant difference in conditional accuracy rates when participants encountered matching expression pairs, after repeating a judgment-relevant label 3 or 30 times,  $t(47) = 1.17$ ,  $p = .25$  (means presented in Table 2.4, in rows 3 and 4 of column 1). Similarly, when responding to mismatching expression pairs, participants were similarly accurate following 3 and 30 repetitions of a relevant label (which matched one of the expressions shown),  $t(47) = 1.08$ ,  $p = .29$  (means presented in Table 2.4, in rows 3 and 4 of column 2).

#### 2.2.3.4.2 Other significant and marginal effects

Mirroring the analysis of proportional accuracy rates, there was a significant interaction between word relevance and expression matching,  $F(1, 47) = 5.75$ ,  $p = .020$ ,  $MSE = 0.037$ ,  $\eta_p^2 = 0.11$ . Whilst the pattern of results followed that of the proportional accuracy scores, all post-hoc comparisons were non-significant.

Mean conditional accuracy rates, by word relevance and expression matching, are shown in Figure 2.3.





**Figure 2.3 Mean conditional accuracy rates by Word Relevance and Expression Matching (error bars represent SEM).**

### 2.2.3.5 Timed out trial analysis

#### 2.2.3.5.1 Repetition-based effects (predicted)

There was a significant main effect of repetition,  $F(1,47) = 6.38, p = .015, MSE = 0.022, \eta_p^2 = 0.12$ . Participants were significantly more likely to be timed out in trials which required 30 repetitions of an emotion label ( $M = 0.23, SEM = 0.033$ ) than 3 repetitions ( $M = 0.19, SEM = 0.027$ ).

There was no interaction observed between repetition and word relevance,  $F(1,47) = 1.44, p = .24, MSE = 0.010, \eta_p^2 = 0.030$ . However, paired-samples t-tests showed that participants were significantly more likely to be timed out in their responses following 30 repetitions of a judgment-relevant emotion label ( $M = 0.24, SEM = 0.036$ ) than 3 repetitions ( $M = 0.19, SEM = 0.023$ ),  $t(47) = 2.94, p = .005, d = 0.24$ . In contrast, participants showed no difference in the commission of timed out responses following 3 repetitions ( $M = 0.20, SEM = 0.029$ ) and 30 repetitions ( $M = 0.23, SEM = 0.032$ ) of a judgment-irrelevant emotion label,  $t(47) = 1.39, p = .17$ .

Analysis of the ‘satiation comparison’ conditions also showed that participants were significantly more likely to be timed out in response to matching expression pairs following 30 (vs 3) repetitions of a judgment-relevant emotion label,  $t(47) = 3.41$ ,  $p = .001$ ,  $d = 0.44$  (means presented in Table 2.4, in rows 5 and 6 of column 1). In contrast, participants were similarly likely to be timed out when responding to expression mismatch pairs after 3 or 30 repetitions of an emotion label that matched one of those expressions,  $t(47) = 1.44$ ,  $p = .16$  (means presented in Table 2.4, in rows 5 and 6 of column 2).

There were no other significant or marginal effects.

## 2.2.4 Discussion

A replication of Lindquist et al. (2006, Experiment 3) was conducted to assess whether satiation decrements were specific to massed repetition of ‘relevant’ emotion labels, with stronger effects predicted when that label matched both of the emotional expressions shown (the ‘satiation comparison’; Smith & Klein, 1990; Lindquist et al., 2006). If so, this would lend support to the category-based, construction position; that access to emotion-related semantic knowledge is crucial for interpreting emotional expressions and that this semantic knowledge is at least ‘activated’, if not organised, according to discrete, emotional state (e.g. Barrett et al., 2007). However, several alternative explanations are possible if decrements were common following massed repetition of any emotion label; one consistent with the construction position (spreading activation, Collins & Loftus, 1975; Neely, 1977) and others inconsistent, and non-semantic in nature (e.g. cognitive fatigue, loss of an anticipatory advantage, Black 2001; Tian & Huber, 2010; Black et al., 2013).

### 2.2.4.1 Exploring repetition-based decrements: category-based, construction vs. alternative accounts

Some findings appear to support for the spreading activation (e.g. Neely, 1977) and/or non-semantic account (e.g. Tian & Huber, 2010). Results of the proportional accuracy analysis suggested that participants were less accurate in their discriminations after repeating both judgment-relevant and irrelevant emotion labels 30 vs. 3 times (a main effect of repetition). The ‘proportion correct’ data represents the number of correct trials, as a proportion of the conflated number of errors that the participant produced (e.g. the summed number of incorrect and timed-out trials). However, similar, weaker trends were

also found in the conditional accuracy analysis; a measure that only accounts for the relative proportion of trials in which participants made a correct vs. incorrect response, and omits the contribution of timed-out responses. Participants were also significantly more likely to be 'timed out' in their attempts to respond following 30, as opposed, to 3 label repetitions. That the repetition decrement was stronger in the timed out trial vs. conditional accuracy analysis, may suggest that when the effect is again found in the proportional accuracy analysis, it is more strongly driven by failures to give a timely response, than occasions where an incorrect response was submitted. In other words, participants may still be able to discriminate between expressions following massed repetition, but may be slower to do so.

Across analyses, there were some findings supportive of the 'category-based' construction position (e.g. Lindquist et al., 2006; Gendron et al., 2012); that the interpretation of expression relies on access to semantic knowledge discretely related to that emotional state (e.g. the interpretation of 'sad' expressions rely on access to semantic knowledge about 'sadness'). Although no interactions were observed between repetition and word relevance for any of the three accuracy variables, further evidence became available through planned comparisons. Across all three accuracy measures these comparisons suggested that participants were less accurate (or were more likely to miss the response window) after 30 repetitions of a judgment-relevant label, than 3 repetitions (although only marginally so for the conditional accuracy scores). Where these repetition-based decrements occurred it was predicted that they would be strongest in the 'satiation comparison' trials, in which the judgment-relevant label matched both emotional expressions shown at test, rather than just one (e.g. Smith & Klein, 1990; Lindquist et al., 2006, see Tables 2.1 and 2.2). A-priori comparisons supported this conclusion in the proportional accuracy and timed-out trial data; participants were less accurate, and more likely to miss trials, after 30 (vs. 3) repetitions of a judgment-relevant label that matched both expressions, but not one. However, no differences emerged when these comparisons when conducted on the conditional accuracy data. All three analyses revealed that participants performed similarly after 3 and 30 repetitions of a judgment-irrelevant emotion label. In sum, a category-based construction position found moderate support, through planned comparisons, in the proportion correct analysis and timed out trial analysis, but not the conditional accuracy analysis. Again, this perhaps suggests that differences in the proportion correct data are more strongly driven by the proportion of

trials that the participant missed after 30 (vs. 3) repetitions of a judgment-relevant label, as opposed to the proportion of trials in which they made incorrect responses.

In sum, the present work provided some evidence for the category-based construction position. Supportive trends emerged in planned comparisons conducted on proportional accuracy and timed-out trial rates, but not the conditional accuracy rates. However, it is important to note that the strongest ‘test’ of the category-based account was not met; there was no interaction observed between word relevance and repetition in any of the three analyses. This adds to the inconsistent evidence of this interaction, reported in Lindquist et al.’s. (2006) work. Evidence in support of a ‘general’ massed repetition decrement, common to 30 (vs 3) repetitions of both judgment-relevant and irrelevant emotion labels, appeared to be more robust. A general massed repetition decrement was found for all three accuracy measures (albeit a marginally significant trend for the conditional accuracy data). As such, alternative semantic (spreading activation e.g. Neely, 1977) and non-semantic explanations (e.g. cognitive fatigue, Tian & Huber, 2010) may be favoured. Given the previously identified limitations of Lindquist et al.’s. (2006) task (e.g. exclusive use of emotion labels/no inclusion of baseline trials), it is not yet possible to delineate between semantic and non-semantic explanations for the general, massed repetition decrement. In chapters 3 and 4, four experiments (2a, 2b, 3 and 4) are described which introduce a variety of baseline conditions to meet this aim.

#### **2.2.4.2 Facilitative effects: exploring the interaction between word relevance and expression matching**

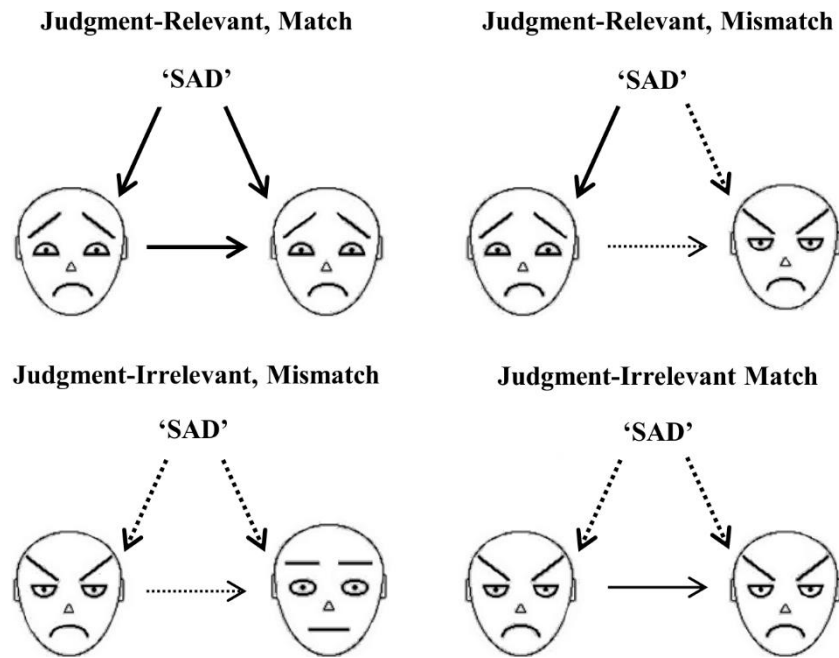
Two further facilitative effects are discussed. One effect emerged in the analyses that accounted for the correct decisions that participants made (proportion correct and conditional accuracy). Here facilitation was observed in expression match trials, but only after repetition of a judgment-relevant emotion label (e.g. trials in which all three stimuli matched; the ‘satiation comparison’ condition, Smith & Klein, 1990; Lindquist et al., 2006). This facilitation occurred relative to performance in the judgment-irrelevant, match trials, but not the judgment-relevant, mismatch trials (though note that trends in the post-hoc comparisons for both proportional and conditional accuracy rates suggested that participants were more likely to submit a correct response in the judgment-relevant, match trials vs. the judgment-relevant, mismatch trials). Importantly, there was no three-way interaction, suggesting that this facilitation was driven by enhanced performance at

both levels of label repetition. Interestingly, comparative facilitation in these conditions was also shown, but not discussed, in Lindquist et al. (2006, Experiments 2 and 3; see Tables 2.1 and 2.2).

Findings may replicate an anomalous trend, frequently observed in the semantic satiation literature (Smith & Klein, 1990; Black, 2001; Tian & Huber, 2010; Kuhl & Anderson, 2011; Black et al., 2013). In the ‘satiation comparison’ trials (Lindquist et al., 2006), although 30 repetitions usually result in worse performance than 3 repetitions, performance after 30 repetitions of a judgment-relevant label remains superior to that observed in baseline trials (i.e. those in which an unrelated or irrelevant category label is repeated). This pattern emerges in the present dataset. Accuracy rates, displayed in Table 2.4, show that performance in the judgment-relevant, 30 repetition match trials is superior to that in the judgment-irrelevant, 30 repetition match trials (0.47 and 0.43 for proportional accuracy rates; 0.59 and 0.56 for conditional accuracy rates). In other words, it is unclear why the satiation manipulation does not reduce performance in the ‘satiation comparison’ trials to baseline levels. Some researchers argue that massed repetition of a relevant label may cause concurrent and competitive priming processes, one positive and one negative (e.g. Smith & Klein, 1990; Huber & O’Reilly, 2003; Tian & Huber, 2010). The links between primes and their corresponding semantic nodes operate normally during early repetitions (the 3 repetition condition), and this positive priming may translate into better performance when the repeated emotion label matches both test expressions. However, later repetitions of the label may fatigue these connections, resulting in much slower or only partial semantic activation when the participant is repeatedly exposed to that judgment-relevant label (the 30 repetition condition). However, satiation may not offset all of the residual positive priming that occurred during the first few repetitions of a 30 repetition trial. This perhaps explains why performance in the 30 repetition, judgment-relevant, match trials is only worse in comparison to the 3 repetition, judgment-relevant match trials, and not to performance in judgment-irrelevant trials (match or mismatch). When we collapse across levels of repetition then, facilitation is still observed in judgment-relevant match trials, as performance after both 3 and 30 repetitions is comparatively heightened. Given that judgment-relevant, mismatch trials only include one expression that matches the repeated label, any positive priming (after 3 repetitions) or residual priming (after 30 repetitions) is unlikely to result in facilitation of a similar magnitude.

Facilitation was also observed in performance levels for the judgment-irrelevant, mismatch trials (e.g. repeat the label 'sad', then see an 'angry' and 'fearful' expression pairing). Analyses conducted on measures that account for the proportion of correct responses showed that participants performed better in judgment-irrelevant, mismatch trials, than judgment-irrelevant, match and judgment-relevant, mismatch trials (again, trends were less robust in the conditional vs. proportional accuracy analysis, though the same pattern of findings emerged). Analyses suggest that facilitated performance was present after both 3 and 30 repetitions of a judgment-irrelevant emotion label. Conversely, Lindquist et al.'s (2006, experiment 2 and 3) data showed that participants performed worst in trials preceded by a judgment-irrelevant label (see Table 2.1). Lindquist et al. (2006) used a verbal interference argument to explain the main effect of word relevance. After participants repeat an irrelevant label it may be comparatively difficult to covertly generate the 'relevant' emotion 'labels' that they require to discriminate between the two test faces (e.g. correctly applying the label 'sad' to a 'sad' expression). Covert label generation may be an automatic, heuristic step that we engage in when attempting to visually discriminate between categorical stimuli (e.g. Roberson & Davidoff, 2000; Roberson, Damjanovic & Pilling, 2007; Roberson, Damjanovic & Kikutani, 2010). This process may be equally disrupted, whether an judgment-irrelevant label is repeated a small or large number of times (single exposure to a judgment-irrelevant label was sufficient to cause disruption in Roberson et al., 2007). Arguably the process by which we access and 'apply' emotion labels may differ from the way in which we access and use emotion-related, semantic knowledge for discrimination. Crucially, that the present study found facilitation in judgment-irrelevant, mismatch trials, compared to judgment-relevant, mismatch trials may weaken an argument based on verbal interference (see further consideration of this explanation in Chapter 5).

There is an important similarity between the two types of trial in which heightened accuracy was found in the present work; judgment-relevant, match and judgment-irrelevant, mismatch conditions (see Figure 2.4).



**Figure 2.4 Trial types and comparisons available in experiment 1.**

*Solid arrows indicate a 'match' comparison, and dashes a 'mismatch' comparison. For reasons of stimulus reproducibility, schematic faces have been used in this figure. Participants were presented with pairs of photographed faces from the Pictures of Facial Affect throughout experiments 1, 2a, 2b, 3 and 4 (Ekman & Friesen, 1976).*

Despite asking participants to make decisions irrespective of the repeated label, we argue that there are three possible comparisons that participants can make in each trial: between the repeated label and the expression shown on Face A, the repeated label and the expression shown on Face B, and between the expressions displayed on Face A and Face B. Facilitated performance is common to conditions in which all three of these comparisons provide a congruent result; three 'matches' in the judgment-relevant, match condition and three 'mismatches' in the judgment-irrelevant, mismatch condition. If facilitation, common to both match and mismatch trials, arises due to the participant making a direct comparison between repeated labels and faces, then this may suggest that language labels have a more 'direct', or strategic, influence on emotion perception than Lindquist et al. (2006) propose. Attention is paid to analogous trial types in our subsequent work to further assess this possibility (Chapters 3 and 4).

### 2.2.4.3 Outline of subsequent experiments (2a, 2b, 3 and 4)

The current work replicates Lindquist et al. (2006) in an important way. Both sets of findings provide stronger evidence for a general, massed repetition decrement, than a decrement specifically tied to the satiation of judgement-relevant emotion labels. Critically, neither the design of Lindquist et al's. (2006) work, nor the methodological replication, would allow us to differentiate between accounts supportive of a general massed repetition decrement (spreading activation vs. non-semantic explanations). Thus in experiments 2a and 2b participants will perform the perceptual discrimination task both after repetition of judgment-relevant emotion labels and neutral labels, which lack any emotional connotations (e.g. 'Pen'). Neutral label trials represent an important baseline condition (e.g. Smith & Klein, 1990; Balota & Black, 1997). The following predictions are made. If discrimination between emotional expressions requires access to meaningful semantic information, then performance should only suffer after massed repetition of judgment-relevant, emotion labels (e.g. Lindquist et al., 2006; Barrett et al., 2007; Gendron et al., 2012). This would suggest that decrements result from semantic inaccessibility, supporting construction theories. However, if alternative, non-semantic factors drive the general massed repetition decrements in Lindquist et al's. (2006) work, then we would expect to find reduced discrimination accuracy both after 30 (vs. 3) repetitions of emotion and neutral labels.

An additional baseline is introduced in experiment 3. Participants will be required to repeat a set of non-words that lack both emotional meaning and any semantic grounding (e.g. 'Borbi'). This manipulation will help us to assess whether there remains a semantic element to any decrements that follow the massed repetition of neutral labels. According to which, massed repetition of any word that has meaning might interfere with general access to the individuals semantic network, via a distributed pattern of spreading inhibition (e.g. Pynte, 1991). If the massed repetition decrement selectively occurs after 30 repetitions of a neutral label then this would favour a semantic explanation for previous results. However, if decrements persist after massed repetition of both neutral labels and non-words this will provide further support for the role of alternative, non- semantic factors in previous work.

Last, in experiment 4 we specifically assess whether a massed repetition decrement is driven by one particular non-semantic factor: heightened cognitive fatigue (e.g. Black,



2001; Tian & Huber, 2010). Here participants were asked to engage in a non-linguistic manipulation: steady-rhythm, repetitive foot tapping. This manipulation should produce similar generic, intra-trial demands to that posed by consistent verbal repetition (e.g. Miyake, Emerson, Padilla & Ahn, 2004; Saeki & Saito, 2004), while leaving semantic (and linguistic) resources intact (see work that compares the effects of articulatory suppression and manual tapping e.g. Baddeley, 1986; Larsen & Baddeley, 2003). Prior to expression discrimination participants will either engage in 3 or 30 repetitive foot taps, or passively watch as the same number of symbols appear on the screen. If participants show selective performance decrements after 30 (vs. 3) instances of foot tapping then this may suggest that intra-trial cognitive fatigue fully or partially explains any previously observed massed repetition decrements. However, if participants show no performance decrements after 30 (vs 3) foot taps, or decrements after both 30 instances of tapping and passive watching, then additional non-semantic, non-linguistic factors are likely implicated in the massed repetition decrement (e.g. response uncertainty/loss of an anticipatory advantage; Tian & Huber, 2010; Black et al., 2013).



## Chapter 3

### **Introducing baseline conditions to explore semantic vs. non-semantic explanations for the massed-repetition decrement (experiments 2a, 2b and 3)**

#### **3.1 Introduction**

The results of experiment 1 very closely mirror those of Lindquist et al. (2006, experiments 2 and 3, see Tables 2.1 and 2.2). Across three sets of analyses some support was found for the ‘category-based’ construction account (proportional accuracy and timed-out trial analysis). However, there were stronger indications that participants were generally less likely to submit correct responses, or respond in a timely manner, after massed repetition of any emotion label (judgment-relevant or irrelevant). It is equally difficult to interpret general, repetition-based decrements in Lindquist et al.’s. (2006) work and that reported in Chapter 1; when participants repeat emotion labels on every trial a massed repetition decrement could either be indicative of a ‘spreading activation/inhibition’ mechanism, supportive of the construction account (e.g. Smith, 1984; Smith & Klein, 1990), or the influence of several, non-semantic factors (e.g. increased cognitive fatigue, Black, 2001; Tian & Huber, 2010; Black et al., 2013).

In three subsequent experiments (2a, 2b and 3) ‘baseline’ manipulations were added to differentiate between these two accounts. Following previous semantic satiation research, a set of ‘unrelated’/neutral trials were included in experiments 2a and 2b (e.g. Smith & Klein, 1990; Balota & Black, 1997; Black, 2001; Tian & Huber, 2010). To do this, trials that required repetition of judgment-irrelevant emotion labels were removed (i.e. labels that matched neither emotional expression at test, experiment 1) and replaced with trials that required the repetition of neutral, object labels (e.g. ‘pen’). A semantic, construction position would be supported if decrements in discrimination follow only the massed repetition of emotion labels (e.g. Lindquist et al., 2006). In contrast, a non-semantic

account would be favoured if decrements are common to the massed repetition of emotion and neutral labels (e.g. Tian & Huber, 2010).

Experiment 3 included a further set of baseline trials. Some researchers argue that semantic explanations are still possible when repetition-based decrements follow the satiation of related and unrelated labels (e.g. Pynte, 1991; Frenck-Mestre, Besson & Pynte, 1997). Here massed repetition of any label with a semantic basis (emotion and neutral labels) may cause inhibition of the semantic network at various focal points, indirectly blocking access to the emotion-related semantic knowledge required for expression discrimination. In experiment 3, participants will encounter trials in which they repeat non-words, with no semantic grounding (e.g. Sekiguchi, 1999), alongside neutral labels, with no emotional connotations. If repetition-based decrements follow equally the massed repetition of neutral labels and non-words then this would provide stronger support for various non-semantic vs. semantic explanations (e.g. cognitive fatigue, Tian & Huber, 2010).

### **3.1.1 Introducing a neutral baseline (experiments 2a and 2b)**

Following previous semantic satiation work, a set of neutral labels were first included as a baseline condition in experiment 2a (e.g. Smith & Klein, 1990; Balota & Black, 1997). Across trials then, participants would either make responses following repetition of a judgment-relevant emotion label, which matched one or both of the emotional expressions shown at test, or a neutral label, which matched neither expression. Neutral and emotion label trials were presented to participants in a random, intermixed order. Neutral labels (e.g. 'pen') have no emotional connotations, thus a semantic account would predict that their repetition should neither initially activate, nor later inhibit, access to emotion-related semantic knowledge (e.g. Smith, 1984; Smith & Klein, 1990). Finding a massed repetition decrement after exposure to neutral labels, or common following both exposure to emotion and neutral labels, will alternatively provide support for a non-semantic explanation (e.g. increased cognitive fatigue, Tian & Huber, 2010).

Experiment 2b differed from experiment 2a in one important way. Rather than presenting emotion and neutral label trials in an intermixed, random order, participants now encountered the two types of trial in separate experimental blocks. Block order was counterbalanced across participants (emotion block first, then neutral block; neutral block first, then emotion block). Modifying presentation format, and inserting a short, filled break

between trial blocks, aimed to eliminate possible two types of carry-over effect, described below (e.g. Blaxton & Neely, 1983; Smith & Klein, 1990; Holle, Neely & Heimberg, 1997; Pilotti, Antrobus & Duff, 1997; Black, 2001; McKenna & Sharma, 2004).

Various types of carry-over effect may be possible when emotional and neutral stimuli are presented in an intermixed order. First, findings from the emotional Stroop task (e.g. MacKay et al., 2004) suggest that slow-acting interference effects are possible when we process emotional words, which may influence neutral word processing when trial types are intermixed (e.g. McKenna, 1986; Holle et al., 1997; McKenna & Sharma, 2004). Semantic satiation, resulting from massed repetition of emotion labels, may similarly accumulate across trials (e.g. Tian & Huber, 2010; Kuhl & Anderson, 2011; English & Visser, 2014). In Kuhl and Andersons' (2011) task (replicated by English & Visser, 2014), participants satiated or primed a set of target words (e.g. 'sheep') and were later implicitly prompted to self-generate these target words in response to highly related probes (e.g. 'herd'). Participants continued to show slower generation latencies for satiated (and strongly associated) words up to a period of 10 minutes after the repetition phase (experiment 2). That satiation effects are long-lasting may suggest that semantic inaccessibility has the potential to interfere with responses across a number of trials, independent of the specific type of word that precedes discrimination (e.g. another emotion label or a neutral label; though see Brown, Zoccoli & Leahy, 2005; Renoult, Brodeur & Debruille, 2010). These effects may be possible because of the 'sluggish' nature of spreading activation (e.g. Smith, 1984).

Second, completion of neutral labels trials may impact on the strategy that participants adopt when they respond to trials preceded by an emotion label. Participants may quickly learn that neutral words are uninformative cues for the subsequent discrimination task, promoting a general state of inattention towards the to-be-repeated word (e.g. Black et al., 2013). As a result, participants might strategically attend only to the surface or phonological features of each neutral label to maintain a steady rate of repetition, foregoing a deeper analysis of label meaning and preserving semantic accessibility (e.g. Smith & Klein, 1990; Pilotti et al., 1997; Kounios et al., 2000; Black, 2001; Shivde & Anderson, 2011, though see Kuhl & Anderson, 2011). Importantly, participants may find it difficult to re-calibrate the strategies they adopt when a randomised testing format is used (e.g. Black, 2001; Pilling, Wiggett, Ozgen & Davies, 2003; Roberson et al., 2007). Therefore,

a shallow/phonological processing style may also be employed when a potentially informative cue is presented (e.g. an emotion label). Some researchers suggest that shallow processing of judgment-relevant labels (emotion) may weaken the potential to find satiation effects (e.g. Blaxton & Neely, 1983; Smith & Klein, 1990; Pilotti et al., 1997). However, English and Visser (2014) found weaker satiation effects when participants were explicitly instructed to remember the words they were repeating/satiating for a subsequent recognition test; arguably these instructions would promote deeper, semantic encoding of the stimuli, which should have resulted in enhanced satiation effects (e.g. Smith & Klein, 1990).

Independent of the potential nature of carry-over effects, it is clear that intermixing emotion and neutral label trials in experiment 2a may obscure performance differences after repetition of each type of label. Presenting the different label types in a blocked order, and counterbalancing block order across participants, should provide a more robust set of results that can be compared against those obtained in experiment 2a.

### **3.1.2 Specific predictions for E2a and E2b**

The same predictions are made for both experiments 2a and 2b. A 'category-based' construction account will be supported if an interaction between word type and repetition is found. Specifically, participants should be less accurate following 30 (vs 3) repetitions of an emotion label, but should show no difference in performance following 30 (vs. 3) repetitions of a neutral label. A-priori comparisons will be used to assess the presence of these trends. After repeating an emotion label we might also expect to find larger massed-repetition decrements in accuracy if that label matches both, rather than one of the emotional expressions shown at test (e.g. an expression match vs. mismatch pairing, the 'satiation comparison' condition, Lindquist et al., 2006). Observing these effects would strongly suggest that access to discrete, emotion-related semantic knowledge is necessary for the discrimination of emotional expressions (e.g. Barrett et al., 2007; Gendron et al., 2012). In contrast, non-semantic interpretations, such as increased cognitive fatigue or response uncertainty (e.g. Tian & Huber, 2010), would be favoured if a main effect of repetition is found, such that participants perform worse after 30 (vs 3) repetitions of any label (emotion or neutral). Arguably, finding a main effect of repetition in experiment 2a will no longer be consistent with a 'spreading activation/inhibition' explanation, as it was

when participants repeated only judgment-relevant and irrelevant emotion labels across trials (experiment 1, e.g. Collins & Loftus, 1975; Neely, 1977, though see experiment 3).

Findings from experiment 1 were suggestive of a further facilitative effect of language labels on emotion perception. Participants performed best in the subset of trials in which all three stimuli matched each other (repeat the label 'sad' then view a matching pair of 'sad' expressions)<sup>6</sup>. Facilitation was present at both levels of label repetition and may suggest that participants directly attend to overtly repeated labels to aid discrimination. In this subset of trials, when participants receive congruent feedback from two initial label-to-face comparisons (both 'matches') they may be primed to make the correct decision when they reach the critical face-to-face discrimination (a third 'match' decision). We explore this possibility in experiments 2a and 2b by examining the interaction between label type and expression matching. If findings replicate those from experiment 1 then participants should perform better in expression match trials preceded by an emotion label, than in either expression match trials preceded by a neutral label, or expression mismatch trials, preceded by an emotion label.

## **3.2 Experiment 2a: Intermixed presentation of emotion and neutral labels**

### **3.2.1 Method**

#### **3.2.1.1 Participants**

Participants were 48 undergraduate students from the University of Leeds (47 females), who participated for course-related credits. The sample had a mean age of 19.39 years (SD

---

<sup>6</sup> In experiment 1, a similar type of facilitation was also observed in trial types in which all three stimuli mismatched one another (e.g. repeat the label 'sad', then view an 'angry' and 'fearful' face). This exact trial type did not exist in experiments 2a and 2b, although a decision-level analogue was available (i.e. trials in which participants repeated a neutral label before responding to a mismatching pair of emotional expressions). Due to differences in the type of preceding label, it was unclear whether to expect relative facilitation in this type of trial (e.g. better performance in the neutral label, expression mismatch vs. neutral label, expression match trials).

= 1.27, Range: 18-24). All participants were native English speakers and 46 were right-handed.

### 3.2.1.2 Materials

#### 3.2.1.2.1 Words

Three emotion labels were retained from experiment 1 ('Anger', 'Fear' and 'Sad'), but the label 'Disgust' was removed. While some emotion taxonomies include the label 'Disgust' (e.g. Plutchik, 1980; Johnson & Laird Oatley, 1989; Ekman, 1992, 2003; Izard, 1992; Kassam, Markey, Cherkassky, Loewenstein & Just, 2013), others do not consider 'Disgust' to be basic in the same sense as other emotions (e.g. Shaver, Schwartz, O'Connor & Kirson, 1987; Robinson, Storbeck, Meier & Kirkeby, 2004; Reisenzein, 2009). In particular, Shaver et al. (1987) suggested that 'Disgust' and 'Anger' are too conceptually similar to be considered discrete emotional states. As such it is unclear whether there would be equal potential to observe repetition-based decrements when participants satiate the label 'Disgust', in comparison to other emotion labels, which hold a stronger, undifferentiated conceptual knowledge base.

Three neutral labels were also selected from the MRC Psycholinguistic database ('Pen', 'Desk' and 'Phone'; Coltheart, 1981; Wilson, 1988). Emotion and neutral labels were matched for number of letters (but not syllables, see experiment 3). Alike emotion labels, neutral labels had a common semantic theme (all objects commonly found in an office). The set of emotion and neutral labels did not differ significantly in their mean familiarity scores [Coltheart, 1981;  $M_{\text{emotion}} = 542.67$ ,  $SD = 41.61$ ;  $M_{\text{neutral}} = 562.33$ ,  $SD = 18.01$ ,  $t(4) = 0.54$ ,  $p = .63$ ]. Using Kucera-Francis scores (Francis & Kucera, 1982) there was no difference between label sets in written frequency [ $M_{\text{emotion}} = 70$ ,  $SD = 49.79$  and  $M_{\text{neutral}} = 45.67$ ,  $SD = 24.58$ , respectively,  $t(4) = 0.76$ ,  $p = .49$ ]. Nor was there a significant difference when using spoken frequency scores from the SUBTLEX-UK database [Van Hueven, Mandera, Keuleers & Brysbaert, 2014;  $M_{\text{emotion}} = 10237$ ,  $SD = 5525.70$  and  $M_{\text{neutral}} = 13464$ ,  $SD = 15410.57$ ,  $t(4) = 0.34$ ,  $p = 0.75$ ]. As predicted, neutral words had higher concreteness scores (Coltheart, 1981;  $M = 592.67$ ,  $SD = 27.79$ ) than emotion words ( $M = 333.67$ ,  $SD = 23.46$ ),  $t(4) = 12.34$ ,  $p < .01$ ].



#### 3.2.1.2.2 *Faces pairs*

Participants encountered face pairs consisting of three emotional expressions: 'Anger', 'Fear' and 'Sad' (sampled from the pairing sets developed for experiment 1). Types of emotion expression were equally distributed across blocks and trials from different conditions.

#### 3.2.1.3 **Design**

A 2 (Word Type: Emotion vs. Neutral) × 2 (Repetition: 3 vs. 30) × 2 (Expression Matching: Match vs. Mismatch) within-participants design was used. Participants completed 96 trials in total; 48 in each trial block. There were eight conditions, and 12 trials per condition (see trial types in Table 3.1). Participants encountered emotion and neutral label trials in a random, intermixed order throughout each experimental block. As before, three accuracy variables were calculated, per condition: proportional accuracy, conditional accuracy and the proportion of timed out trials (see specific calculations in Chapter 2, pages 37–38). Overall, timed out trials accounted for 20.51% of all responses.

#### 3.2.1.4 **Procedure**

The experimental procedure closely followed that used in experiment 1, and Lindquist et al. (2006, experiment 3); participants engaged in 3 or 30 verbal repetitions of a label before deciding whether two faces matched or mismatched each other in emotional expression. Participants were asked to disregard any potential relationships between the words they had repeated, and the emotional expressions displayed, when they made their matching decisions (see Appendix A for full participant instructions). Although the short response window was carried forward from experiment 1 (854ms), participants were now required to press the spacebar between the end of the current trial's response window, and the beginning of a new trial. This modification was introduced to provide participants with a longer preparation time for the subsequent trial, and meant that inter-trial intervals were self-paced and potentially variable.

**Table 3.1** Examples of the eight trial types used in experiment 2a (adapted from Lindquist et al., 2006).

<b>Word Repetition</b>	<b>Type of Word</b>	<b>Expression Match</b>	<b>Trial Example</b>
<b>3</b>	<b>Emotion</b>	<b>Match</b>	3 repetitions of the word 'Sad'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Sad'; compare a 'Sad' and 'Angry' expression pair
	<b>Neutral</b>	<b>Match</b>	3 repetitions of the word 'Pen'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Pen'; compare an 'Angry' and 'Sad' expression pair
<b>30</b>	<b>Emotion</b>	<b>Match</b>	30 repetitions of the word 'Anger'; compare two 'Angry' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Anger'; compare an 'Angry' and 'Disgusted' expression pair
	<b>Neutral</b>	<b>Match</b>	30 repetitions of the word 'Phone'; compare two 'Fearful' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Phone'; compare an 'Angry' and 'Fearful' expression pair

## 3.2.2 Results

### 3.2.2.1 Main Analysis

Separate 2 (Word Type: Emotion vs. Neutral)  $\times$  2 (Repetition: 3 vs. 30)  $\times$  2 (Expression Matching: Match vs. Mismatch) within-participant ANOVAs were conducted on proportional accuracy, conditional accuracy and timed-out trial rates. Bonferroni corrections were applied, per analysis. As in experiment 1, effects and planned comparisons that address specific predictions are first reported (see pages 56–57), followed by additional significant and marginal effects. Effects are considered significant if  $p < .05$ , and marginal if  $p \leq .06$  and  $> .05$ . Mean proportional accuracy, conditional accuracy and timed-out trial rates are presented, by condition, in Table 3.2 below.

**Table 3.2 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis).<sup>7, 8</sup>**

Dependent Variable	Repetitions	Emotion Label		Neutral Label	
		Match	Mismatch	Match	Mismatch
Proportional Accuracy	3	0.58 (0.023)	0.44 (0.023)	0.47 (0.026)	0.46 (0.026)
	30	0.57 (0.027)	0.36 (0.027)	0.50 (0.030)	0.40 (0.036)
Conditional Accuracy	3	0.70 (0.025)	0.55 (0.027)	0.56 (0.027)	0.57 (0.029)
	30	0.71 (0.025)	0.49 (0.035)	0.63 (0.036)	0.52 (0.040)
Timed out responses	3	0.16 (0.023)	0.20 (0.027)	0.17 (0.026)	0.20 (0.027)
	30	0.21 (0.030)	0.27 (0.036)	0.21 (0.030)	0.24 (0.033)

### 3.2.2.2 Proportional Accuracy Analysis

#### 3.2.2.2.1 Repetition-based effects (predicted)

There was a significant main effect of repetition,  $F(1, 47) = 5.95$ ,  $p = .019$ ,  $MSE = 0.017$ ,  $\eta_p^2 = 0.11$ . Participants were significantly less likely to provide a correct response after 30 repetitions of any label ( $M = 0.46$ ,  $SEM = 0.19$ ), than after 3 repetitions ( $M = 0.49$ ,  $SEM = 0.14$ ). Please note that this effect is qualified by a significant interaction with expression matching, reported below (page 65).

<sup>7</sup> Single-samples t-tests revealed that three of the condition means for proportional accuracy did not differ significantly from chance level (0.50): the neutral label, 3 repetition, expression match condition ( $p = .28$ ); the neutral label, 30 repetition, expression match condition ( $p = .86$ ) and the neutral label, 3 repetition, expression mismatch condition ( $p = .13$ ).

In addition, the proportional accuracy means for three conditions fell significantly below chance levels (0.50): the emotion label, 3 repetition, expression mismatch condition ( $p = .012$ ); the emotion label, 30 repetition, expression mismatch condition ( $p < .001$ ) and the neutral label, 30 repetition, expression mismatch condition ( $p = .008$ ).

<sup>8</sup> Single-samples t-tests revealed that two of the condition means for conditional accuracy did not differ significantly from chance level (0.50): the emotion label, 30 repetition, expression mismatch condition ( $p = .70$ ) and neutral label, 30 repetition, expression mismatch condition ( $p = .61$ ).

There was no interaction observed between repetition and word type,  $F(1,47) = 1.93$ ,  $p = .17$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.039$ . However, paired samples t-tests revealed that participants were significantly less likely to provide a correct response after 30 repetitions of an emotion label ( $M = 0.46$ ,  $SEM = 0.022$ ) than after 3 repetitions ( $M = 0.51$ ,  $SEM = 0.016$ ),  $t(47) = 2.61$ ,  $p = .012$ ,  $d = 0.39$ . There was no significant difference in the proportion of correct responses given after 3 ( $M = 0.46$ ,  $SEM = 0.019$ ) or 30 repetitions of a neutral label ( $M = 0.45$ ,  $SEM = 0.020$ ),  $t(47) = 0.93$ ,  $p = .34$ .

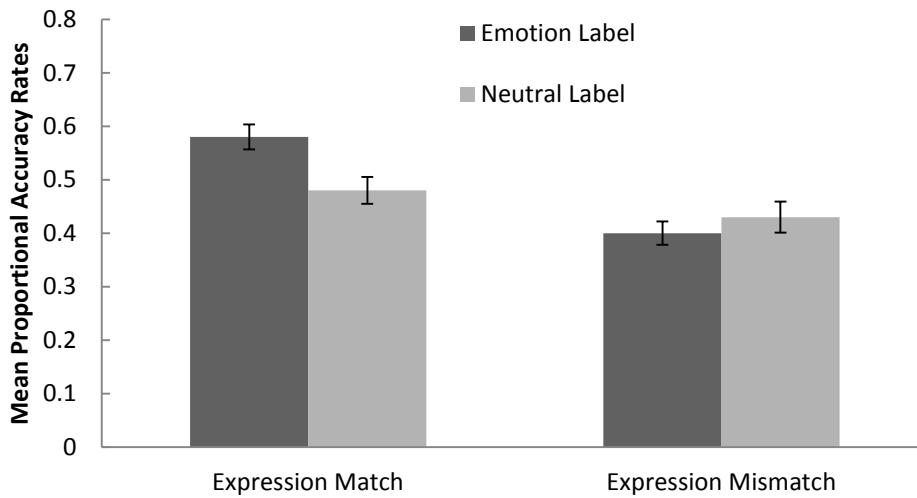
Differences following 3 and 30 repetitions of an emotion label were not driven by performance in the 'satiation comparison' trials. In trials where both expressions matched the repeated emotion label, participants showed no difference in the proportion of correct responses made after 3 repetitions, as opposed to 30 repetitions,  $t(47) = 0.63$ ,  $p = .54$  (see condition means presented in Table 3.2, rows 1-2 of column 1). However, when the emotion label matched only one expression in a pair, participants were significantly less likely to give a correct response following 30 vs. 3 repetitions,  $t(47) = 2.74$ ,  $p = .009$ ,  $d = 0.46$  (see condition means presented in Table 3.2, rows 1-2 of column 2).

#### 3.2.2.2.2 *Other significant and marginal effects*

There was a significant main effect of expression matching, such that participants were less likely to give a correct response to a mismatching expression pair ( $M = 0.41$ ,  $SEM = 0.022$ ) as opposed to a matching expression pair ( $M = 0.53$ ,  $SEM = 0.020$ ),  $F(1,47) = 14.14$ ,  $p < .001$ ,  $MSE = 0.092$ ,  $\eta_p^2 = 0.23$ .

There was also a significant main effect of word type,  $F(1,47) = 4.82$ ,  $p = .033$ ,  $MSE = 0.019$ ,  $\eta_p^2 = 0.093$ . Participants were significantly less likely to give a correct response after repeating a neutral label ( $M = 0.46$ ,  $SEM = 0.017$ ), as opposed to an emotion label ( $M = 0.49$ ,  $SEM = 0.016$ ).

The three main effects (Repetition, Word Type and Expression Matching) were qualified by two significant interactions. First, there was a significant interaction between Expression Matching and Word Type,  $F(1, 47) = 8.43$ ,  $p = .006$ ,  $MSE = 0.044$ ,  $\eta_p^2 = 0.15$  (see Figure 3.1).



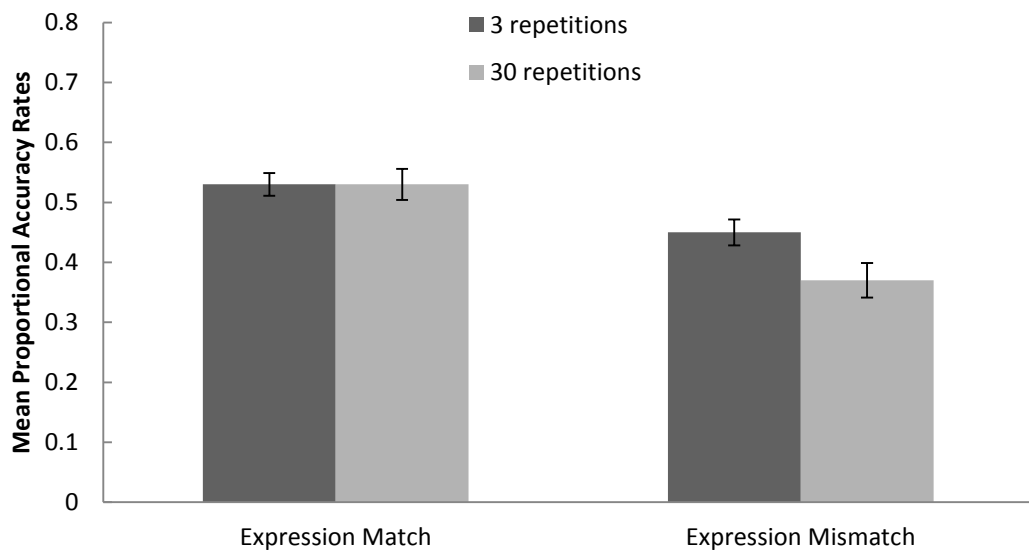
**Figure 3.1 Mean proportional accuracy rates by Expression Matching and Word Type (error bars represent SEM).**

When participants repeated an emotion label, they were significantly more likely to submit a correct response in response to matching expression pairs ( $M = 0.58$ ,  $SEM = 0.023$ ), as opposed to mismatching pairs ( $M = 0.40$ ,  $SEM = 0.022$ ),  $t(47) = 5.74$ ,  $p < .001$ ,  $d = 1.14$ . After repeating a neutral label, participants gave a similar proportion of correct responses for expression mismatch pairs ( $M = 0.43$ ,  $SEM = 0.029$ ) and expression match pairs ( $M = 0.48$ ,  $SEM = 0.026$ ),  $t(47) = 1.25$ ,  $p = .22$ . Additionally, when participants responded to expression match pairs, they were significantly more likely to give a correct response after repeating an emotion vs. neutral label,  $t(47) = 3.68$ ,  $p = .001$ ,  $d = 0.55$ . In response to expression mismatch pairs, there was no difference in the proportion of correct responses given following repetition of an emotion or neutral label,  $t(47) = 1.20$ ,  $p = .24$ .

It may be that participants used the repeated label as a heuristic ‘cue’ for the subsequent decision task (e.g. Smith & Klein, 1990; Black, 2001; Black et al., 2013). When repeating an emotion label, participants may have responded ‘match’ due to the shared emotional content between the label and test expressions. Thus better performance in emotion match trials may reflect a systematic response bias. However, additional analyses ruled out this possibility. After repeating an emotion label, participants were significantly more likely to correctly respond ‘match’ to matching expression pairs ( $M = 0.58$ ,  $SEM = 0.025$ ) than to incorrectly respond ‘match’ to mismatching expression pairs ( $M = 0.37$ ,  $SEM = 0.031$ ),  $t(47) = 8.70$ ,  $p < .001$ ,  $d = 1.24$ . In addition, participants did not demonstrate a reversed bias,

automatically responding ‘mismatch’ after repeating a neutral label with unshared emotional content (e.g. Black, 2001). After repeating a neutral label, participants were significantly more likely to correctly respond ‘mismatch’ to mismatching expression pairs ( $M = 0.43$ ,  $SEM = 0.031$ ), than they were to incorrectly respond ‘mismatch’ to matching expression pairs ( $M = 0.33$ ,  $SEM = 0.029$ ),  $t(47) = 4.70$ ,  $p < .001$ ,  $d = 0.52$ .

Second, there was a significant interaction between Expression Matching and Repetition,  $F(1,47) = 4.99$ ,  $p = .030$ ,  $MSE = 0.027$ ,  $\eta_p^2 = 0.090$  (see Figure 3.2).



**Figure 3.2 Mean proportional accuracy rates by Expression Matching and Repetition (error bars represent SEM).**

In general, a greater proportion of correct responses were given to matching compared to mismatching expressions pairs. This trend appeared after 3 repetitions: matching expression pairs ( $M = 0.53$ ,  $SEM = 0.019$ ) vs. mismatching expression pairs ( $M = 0.45$ ,  $SEM = 0.022$ ) vs,  $t(47) = 2.82$ ,  $p = .007$ ,  $d = 0.56$ , but was stronger in magnitude after 30 repetitions: matching expression pairs ( $M = 0.53$ ,  $SEM = 0.029$ ) vs. mismatching expression pairs ( $M = 0.37$ ,  $SEM = 0.026$ ),  $t(47) = 3.74$ ,  $p < .001$ ,  $d = 0.80$ . In addition, when participants encountered expression mismatch pairs, they were significantly less likely to give a correct response after repeating a label 30 times versus 3 times,  $t(47) = 2.81$ ,  $p = .007$ ,  $d = 0.45$ . When participants encountered expression match displays there was no difference in the proportion of correct responses given after 3 or 30 repetitions of a label,  $t(47) = 0.31$ ,  $p = .76$ .

### 3.2.2.3 Conditional Accuracy Analysis

#### 3.2.2.3.1 Repetition-based effects (predicted)

The main effect of repetition was non-significant,  $F(1, 47) = 0.39, p = .53, \text{MSE} = 0.018, \eta_p^2 = 0.008$ . Participants were similarly accurate following 30 label repetitions ( $M = 0.59, \text{SEM} = 0.034$ ) and 3 label repetitions ( $M = 0.60, \text{SEM} = 0.027$ ). Please note that the repetition variable significantly interacts with the expression match variable (see pages 68–69).

There was no interaction observed between repetition and word type,  $F(1, 47) = 1.31, p = .26, \text{MSE} = 0.022, \eta_p^2 = 0.027$ . Moreover, paired samples t-tests revealed that participants were similarly accurate in their discriminations after 3 repetitions ( $M = 0.62, \text{SEM} = 0.017$ ) and 30 repetitions of an emotion label ( $M = 0.60, \text{SEM} = 0.017$ ),  $t(47) = 1.26, p = .21$ .

Participants also showed no difference in accuracy after 3 ( $M = 0.57, \text{SEM} = 0.016$ ) or 30 repetitions of a neutral label ( $M = 0.58, \text{SEM} = 0.019$ ),  $t(47) = 0.43, p = .67$ .

Specific analysis of the ‘satiation comparison’ conditions revealed no significant difference in accuracy when participants responded to matching expression pairs, after repeating a judgment-relevant emotion label 3 or 30 times,  $t(47) = 0.46, p = .65$  (see condition means in Table 3.2, rows 3-4 of column 1). When the emotion label matched only one expression in a pair, participants were also similarly likely to give a correct response following 30 vs. 3 repetitions,  $t(47) = 1.71, p = .094$  (see condition means in Table 3.2, rows 3-4 of column 2).

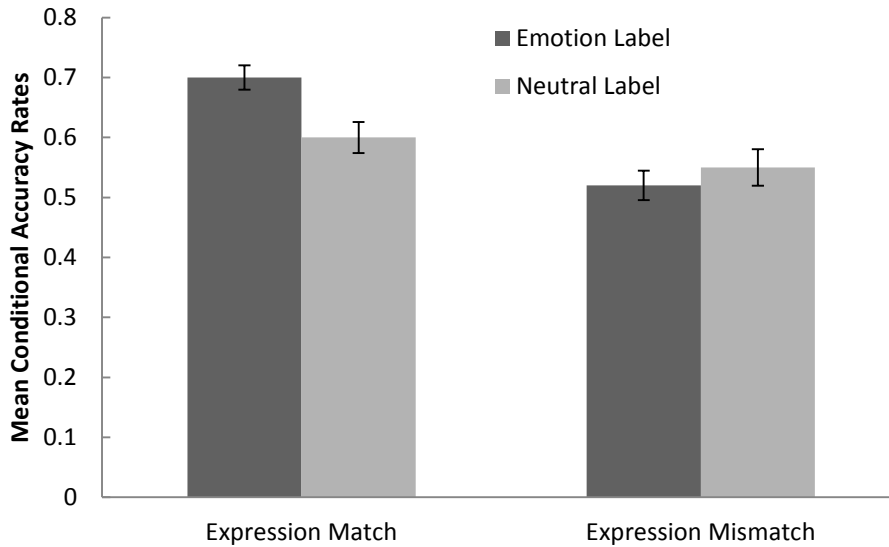
#### 3.2.2.3.2 Other significant and marginal effects

There was a significant main effect of expression matching,  $F(1, 47) = 11.03, p = .002, \text{MSE} = 0.12, \eta_p^2 = 0.19$ . Participants were significantly less accurate when required to respond to two faces which were mismatching ( $M = 0.53, \text{SEM} = 0.033$ ) as opposed to matching, in emotional expression ( $M = 0.65, \text{SEM} = 0.029$ ).

There was a significant main effect of word type. Participants were significantly less accurate in their decisions after repeating a neutral label ( $M = 0.57, \text{SEM} = 0.033$ ), rather than an emotion label ( $M = 0.61, \text{SEM} = 0.027$ ),  $F(1, 47) = 4.65, p = .036, \text{MSE} = 0.031, \eta_p^2 = 0.090$ .



These effects were qualified by an interaction between expression matching and word type,  $F(1, 47) = 7.88$ ,  $p = 0.007$ ,  $MSE = 0.05$ ,  $\eta_p^2 = 0.15$ . Marginal means are shown in Figure 3.3.



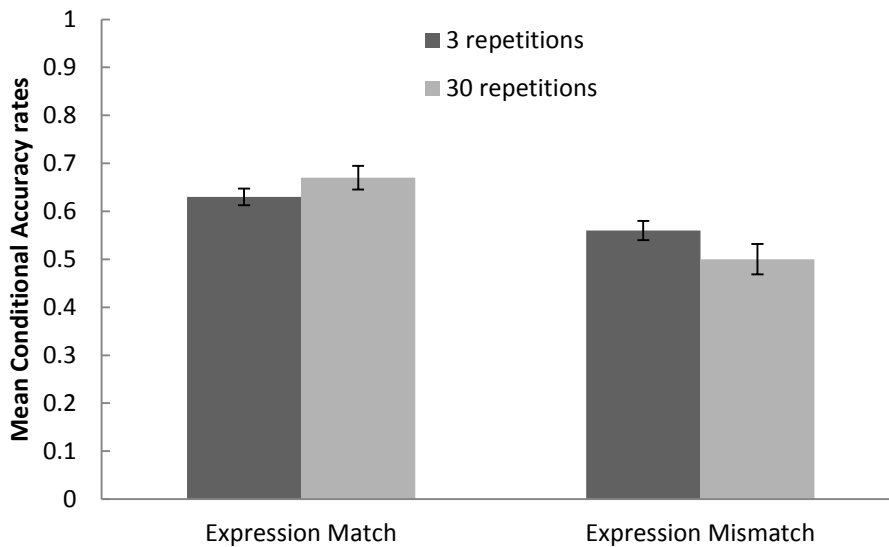
**Figure 3.3 Conditional accuracy rates by Expression Match and Repetition (error bars represent SEM).**

After repeating emotion labels, paired samples t-tests revealed that participants were significantly more accurate when responding to expression match ( $M = 0.70$ ,  $SEM = 0.020$ ) vs. mismatch pairs ( $M = 0.52$ ,  $SEM = 0.025$ ),  $t(47) = 5.30$ ,  $p < .001$ ,  $d = 1.17$ . After repeating neutral labels, participants performed similarly when responding to expression match ( $M = 0.60$ ,  $SEM = 0.026$ ) and expression mismatch pairs ( $M = 0.55$ ,  $SEM = 0.030$ ),  $t(47) = 1.02$ ,  $p = .31$ . In addition, when they encountered expression match pairings, participants were significantly more accurate after they had repeated an emotion vs. neutral label,  $t(47) = 3.86$ ,  $p < .001$ ,  $d = 0.63$ . Participants performed similarly when they encountered expression mismatch pairs, whether they had repeated emotion or neutral labels,  $t(47) = 0.88$ ,  $p = .39$ .

Additional analyses again ruled out the possibility that facilitation emerged due to a cue-driven response bias (e.g. Smith & Klein, 1990; Black, 2001; Black et al., 2013). After repeating an emotion label, participants were significantly more likely to correctly respond 'match' to matching expression pairs ( $M = 0.70$ ,  $SEM = 0.014$ ) than to incorrectly respond 'match' to mismatching expression pairs ( $M = 0.48$ ,  $SEM = 0.025$ ),  $t(47) = 7.85$ ,  $p < 0.001$ ,  $d$

=1.43. After repeating a neutral label, participants were significantly more likely to correctly respond 'mismatch' to mismatching face pairs ( $M = 0.55$ ,  $SEM = 0.030$ ), than they were to incorrectly respond 'mismatch' to matching face pairs ( $M = 0.40$ ,  $SEM = 0.026$ ),  $t(47) = 5.33$ ,  $p < .001$ ,  $d = 0.78$ .

There was a significant interaction between repetition and expression matching,  $F(1, 47) = 5.86$ ,  $p = .019$ ,  $MSE = 0.041$ ,  $\eta_p^2 = 0.11$  (see marginal means presented in Figure 3.4, below).



**Figure 3.4 Mean conditional accuracy rates by Expression Matching and Repetition (error bars represent SEM).**

After 3 repetitions, paired samples t-tests revealed that participants showed a significant decrease in accuracy when responding to expression mismatch pairs ( $M = 0.56$ ,  $SEM = 0.020$ ), vs. expression match pairs ( $M = 0.63$ ,  $SEM = 0.017$ ),  $t(47) = 2.27$ ,  $p = .028$ ,  $d = 0.54$ . After 30 repetitions, participants also responded less accurately to mismatching ( $M = 0.50$ ,  $SEM = 0.032$ ) vs. matching expression pairs ( $M = 0.67$ ,  $SEM = 0.025$ ),  $t(47) = 3.37$ ,  $p = .001$ ,  $d = 0.87$ . In addition, when participants encountered face pairs that matched in expression they were significantly less accurate to respond following 3 vs. 30 word repetitions,  $t(47) = 2.18$ ,  $p = .035$ ,  $d = 0.28$ . In contrast, when participants encountered face pairs that mismatched in expression they were significantly less accurate after 30 vs. 3 word repetitions,  $t(47) = 1.99$ ,  $p = .050$ ,  $d = 0.34$ .

### 3.2.2.4 Timed out Trial Analysis

#### 3.2.2.4.1 Repetition-based effects (predicted)

There was a significant main effect of repetition,  $F(1,47) = 13.56$ ,  $p = .001$ ,  $MSE = 0.018$ ,  $\eta_p^2 = 0.22$ . Participants were significantly more likely to be timed out in trials involving 30 repetitions of a label ( $M = 0.23$ ,  $SEM = 0.029$ ) than 3 repetitions of a label ( $M = 0.18$ ,  $SEM = 0.023$ ).

There was no interaction observed between repetition and word type,  $F(1,47) = 0.59$ ,  $p = .45$ ,  $MSE = 0.008$ ,  $\eta_p^2 = 0.012$ . Planned comparisons showed that participants were significantly more likely to be timed out in their attempts to respond following 30 repetitions of an emotion label ( $M = 0.24$ ,  $SEM = 0.030$ ), than 3 repetitions ( $M = 0.18$ ,  $SEM = 0.023$ ),  $t(47) = 3.40$ ,  $p = .001$ ,  $d = 0.34$ . Participants were also more likely to be timed out in their responses following 30 repetitions of neutral label ( $M = 0.22$ ,  $SEM = 0.030$ ) than 3 repetitions ( $M = 0.18$ ,  $SEM = 0.025$ ),  $t(47) = 2.73$ ,  $p = .009$ ,  $d = 0.21$ .

Specific analysis of the ‘satiation comparison’ conditions revealed that participants were significantly more likely to be timed out when their responses followed 30 repetitions of a matching emotion label, than 3 repetitions,  $t(47) = 2.34$ ,  $p = .024$ ,  $d = 0.22$  (see condition means presented in Table 3.2, rows 5-6 of column 1). When the emotion label matched only one expression in a pair, participants were also significantly more likely to be timed out when their responses followed 30 vs. 3 repetitions,  $t(47) = 2.79$ ,  $p = .008$ ,  $d = 0.32$  (see condition means presented in Table 3.2, rows 5-6 of column 2).

#### 3.2.2.4.2 Other significant and marginal effects

There was a significant main effect of expression matching,  $F(1,47) = 11.57$ ,  $p < .001$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.20$ . Participants were significantly more likely to be timed out when they encountered mismatching expression pairs ( $M = 0.23$ ,  $SEM = 0.027$ ), than matching expression pairs ( $M = 0.19$ ,  $SEM = 0.032$ ).

### 3.2.3 Results summary

Following 30 (vs. 3) repetitions of a label, participants were significantly less likely to give a correct response (proportional accuracy). Unqualified repetition-based decrements were not found in the conditional accuracy data. However, trends in both the proportional and conditional accuracy analyses suggested that repetition-decrements were only apparent when considering responses to mismatching vs. matching expression pairs. An additional facet in the conditional accuracy analysis suggested that participants instead showed a reversed trend in response to matching expression pairs; participants were less accurate in response to this type of trial following 3 (vs. 30) repetitions.

Although the proportional and conditional accuracy results show that massed repetition decrements were tied to performance in expression mismatch trials, they do not suggest that decrements selectively followed 30 (vs 3) repetitions of an emotion label. There were no interactions observed between repetition and word type in any of the three analyses. Further, only planned comparisons conducted on the proportional accuracy data suggested that repetition-based decrements were stronger following 30 (vs 3) repetitions of an emotion label, than 30 (vs 3) repetitions of a neutral label (a trend not apparent for conditional accuracy). Moreover, the specific ‘satiating comparison’ (Smith & Klein, 1990; Lindquist et al., 2006) failed to reach statistical significance in both analyses that considered the proportion of correct responses that the participant made (proportional and conditional accuracy).

An examination of the timed-out trial analysis however, showed more robust effects of repetition. Specifically, participants were more likely to miss the response window after they had repeated a label 30 (vs 3) times, whether it matched both or one of the emotional expressions displayed at test. Participants showed similar relative decrements after 30 vs 3 repetitions of either emotion or neutral labels. When considering only trials preceded by the massed repetition of emotion labels, the timed-out trial analysis showed performance decrements whether pairs matched (‘satiating comparison’, Lindquist et al., 2006) or mismatched in expression.

Replicating experiment 1, both the proportional and conditional accuracy analyses revealed that participants performed best in emotion, match trials, where all three stimuli matched each other (e.g. repeat the emotion label ‘sad’, and then see two matching ‘sad’

expressions). In both sets of analyses, facilitation now occurred relative to emotion, mismatch trials (in which the repeated label matched only one face; see similar trends reported in experiment 1), and neutral, match trials. Facilitated performance followed both 3 and 30 repetitions of the relevant label and could not be explained by the adoption of a systematic response bias (e.g. participants did not automatically respond 'match' after repeating an emotion label, Smith & Klein, 1990). In contrast to experiment 1, we did not find corresponding facilitation in neutral, mismatch trials, where all three stimuli mismatched each other.

### 3.3 Experiment 2b: Blocked presentation of emotion and neutral labels

#### 3.3.1 Method

##### 3.3.1.1 Participants

Participants were 60 undergraduate students and staff members from the University of Leeds. Twelve participants were excluded; nine participants were timed out in 90%+ of their responses, one participant frequently used an unspecified key when responding, one participant was not a native English speaker and one further participant spontaneously reported use of a response strategy (selecting 'match' after repeating an emotion label, and 'mismatch' after repeating a neutral label). The remaining sample (N = 48, 35 females) had a mean age of 26.48 years (SD = 9.16, Range: 18-54). All remaining participants were native English speakers and 44 participants were right-handed. Students participated for course-related credits and staff for a monetary incentive.

##### 3.3.1.2 Materials

###### 3.3.1.2.1 Words

The Emotion labels ('Anger', 'Sad', 'Fear') and Neutral labels ('Pen', 'Desk', 'Phone') were the same as those used in experiment 2a.

###### 3.3.1.2.2 Faces

Again, participants responded to three emotional expressions, equally distributed across trial types and trial blocks ('Anger', 'Fear' and 'Sad').

##### 3.3.1.3 Design

Experiment 2b differed from experiment 2a in one important way. Participants encountered emotion and neutral label trials in two separate blocks. Block order was counterbalanced and 24 participants were randomly assigned to each condition.

Participants in the two block-order conditions did not differ significantly in age [ $M_{\text{emotion-first}}$ : 26.63, SD = 9.45;  $M_{\text{neutral-first}}$ : 25.67, SD = 7.80;  $t(46) = 0.38$ ,  $p = .70$ ]. Two left-handed participants were included in each of the block order conditions.

Participants completed 96 trials in total; 48 in each trial block. As in experiment 2a, there were eight conditions, and 12 trials per condition (see trial types in Table 3.1, page 60). Three dependent variables were used: proportional accuracy, conditional accuracy and timed-out trial rates (specific calculations, per variable, are reported in Chapter 2, pages 37–38). Timed out trials accounted for 32.64% of total responses.

#### **3.3.1.4 Procedure**

In contrast to previous experiments (1 and 2a), all participants encountered a five minute break between trial blocks. Here they completed several health and safety exercises, designed for computer users (Software Educational Resources Ltd. New Zealand; see Appendix B). Each exercise focused on manipulation of the shoulders, hands and wrists and the experimenter verbally described and demonstrated each exercise for the participant. Replicating experiment 2a, participants were required to submit a spacebar press to progress to the next trial after the short response window had elapsed. This meant that inter-trial intervals were self-paced, and potentially variable, allowing the participant to adequately prepare for the next trial.

### 3.3.2 Results

#### 3.3.2.1 Main Analysis

Separate 2 (Block Order: Emotion-first, then neutral vs Neutral-first, then emotion)  $\times$  2 (Word Type: Emotion vs. Neutral)  $\times$  2 (Repetition: 3 vs. 30)  $\times$  2 (Expression Matching: Match vs. Mismatch) mixed factorial ANOVAs were conducted on proportional accuracy, conditional accuracy and timed-out trial rates. Block order was included as a between-participants factor, but did not qualify effects with any variables of interest. For completeness we report the effects of block order in footnote 3<sup>9</sup>. Bonferroni corrections were applied for each analysis. Per analysis, effects and planned comparisons that address specific predictions are first reported, followed by additional significant and marginal effects. Effects are considered significant if  $p < .05$ , and marginal if  $p \leq .06$  and  $> .05$ .

Mean proportional accuracy, conditional accuracy and timed-out trial rates, by condition, are displayed in Table 3.3, below.

---

<sup>9</sup> Across analyses, the block order factor did not qualify any hypothesis-specific effects (a main effect of repetition, an interaction between repetition and word type, or an interaction between word type and expression matching). Therefore, we collapse across this factor when presenting the descriptive statistics for this experiment. For completeness, significant effects including block order are reported here. There was a significant interaction between Word Type and Block Order in the proportional accuracy analysis,  $F(1, 46) = 8.22$ ,  $p = .006$ ,  $MSE = 0.044$ ,  $\eta_p^2 = 0.15$ , and the conditional accuracy analysis,  $F(1, 46) = 6.52$ ,  $MSE = 0.066$ ,  $p = .014$ ,  $\eta_p^2 = .13$ . For both analyses, t-tests revealed that participants were significantly less accurate when responding to neutral vs. emotion label trials, but only if they encountered neutral label trials in their first trial block. There was also an interaction between Word Type and Block Order for the analysis conducted on the proportion of timed-out trials,  $F(1, 46) = 7.79$ ,  $p = .008$ ,  $MSE = 0.053$ ,  $\eta_p^2 = 0.15$ . Here t-tests showed that participants were always more likely to be timed out when responding to trials in the first block that they encountered. Replicating the proportional and conditional accuracy analyses, these trends were strongest when the first block consisted of neutral vs. emotion label trials.



**Table 3.3 Mean, proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis).<sup>10</sup>**

Dependent Variable	Repetitions	Emotion Label		Neutral Label	
		Match	Mismatch	Match	Mismatch
Proportional Accuracy	3	0.53 (0.035)	0.41 (0.031)	0.44 (0.032)	0.43 (0.031)
	30	0.50 (0.040)	0.39 (0.032)	0.43 (0.029)	0.41 (0.034)
Conditional Accuracy	3	0.74 (0.033)	0.63 (0.040)	0.60 (0.038)	0.63 (0.036)
	30	0.70 (0.036)	0.62 (0.040)	0.62 (0.036)	0.61 (0.043)
Timed out responses	3	0.28 (0.040)	0.36 (0.042)	0.30 (0.040)	0.34 (0.039)
	30	0.33 (0.040)	0.38 (0.044)	0.33 (0.042)	0.38 (0.040)

### 3.3.2.2 Proportional Accuracy Analysis

#### 3.3.2.2.1 Repetition-based effects (predicted)

There was no main effect of repetition,  $F(1, 46) = 2.85$ ,  $p = .098$ ,  $MSE = 0.017$ ,  $\eta_p^2 = 0.058$ . Participants were similarly likely to give a correct response following 3 repetitions ( $M = 0.45$ ,  $SEM = 0.033$ ) and 30 repetitions of a word ( $M = 0.43$ ,  $SEM = 0.033$ ).

There was no interaction observed between repetition and word type,  $F(1, 46) = 0.30$ ,  $p = .59$ ,  $MSE = 0.019$ ,  $\eta_p^2 = 0.007$ . Planned comparisons showed that participants were similarly likely to give a correct response following 3 repetitions of an emotion label ( $M = 0.47$ ,  $SEM = 0.026$ ) and 30 repetitions of an emotion label ( $M = 0.44$ ,  $SEM = 0.032$ ),  $t(47) = 1.66$ ,  $p = 0.11$ . Participants were also equally likely to submit a correct response after 3 ( $M = 0.43$ ,

<sup>10</sup> Single-samples t-tests revealed that two condition means for proportional accuracy did not differ significantly from chance (0.50): the emotion label, 3 repetition, expression match condition total ( $p = .35$ ), and the emotion label, 30 repetition, expression match condition total ( $p = .90$ ). In addition, five of the proportional accuracy means were found to be significantly lower than chance level (0.50): the emotion label, 3 repetition, expression mismatch condition total ( $p = .005$ ); the emotion label, 30 repetition, expression mismatch condition total ( $p = .001$ ); the neutral label, 3 repetition, expression mismatch condition total ( $p = .030$ ); the neutral label, 30 repetition, expression match condition total ( $p = .032$ ), and the neutral label, 30 repetition, expression mismatch condition ( $p = .012$ ).

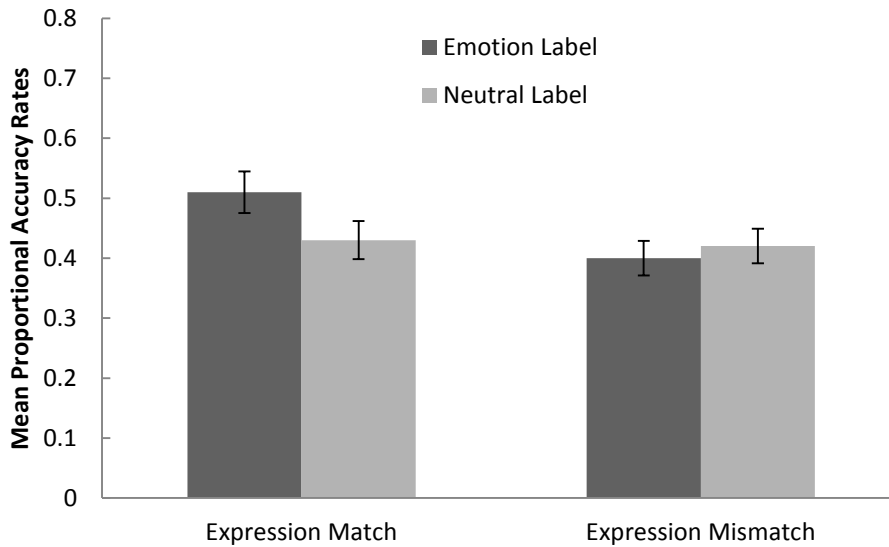
SEM = 0.027) vs. 30 repetitions of a neutral label (M = 0.42, SEM = 0.029),  $t(47) = 0.73$ ,  $p = .47$ .

Specific analysis of the 'satiation comparison' conditions revealed no significant difference in the proportion of correct responses given when participants encountered matching expression pairs, after repeating a relevant emotion label 3 vs. 30 times,  $t(47) = 1.30$ ,  $p = .20$  (see condition means presented in Table 3.3, rows 1-2 of column 1). Nor did participants differ in the proportion of correct responses given when they repeated an emotion label which matched only one of the expressions, 3 vs. 30 times,  $t(47) = 0.80$ ,  $p = .43$  (see condition means presented in Table 3.3, rows 1-2 of column 2).

#### 3.3.2.2.2 *Other significant and marginal effects*

There was a significant main effect of expression matching,  $F(1, 46) = 7.75$ ,  $p = .008$ , MSE = 0.054,  $\eta_p^2 = 0.14$ . Participants were significantly less likely to submit a correct response when they encountered mismatching expression pairs (M = 0.41, SEM = 0.032) vs. matching expression pairs (M = 0.47, SEM = 0.035).

This was also a significant interaction observed between expression matching and word type,  $F(1, 46) = 6.18$ ,  $p = .017$ , MSE = .041,  $\eta_p^2 = 0.12$  (see marginal means presented in Figure 3.5).



**Figure 3.5 Mean, proportional accuracy rates by Word Type and Expression Matching (error bars represent SEM).**

Paired samples t-tests revealed that, after repeating an emotion label, participants were significantly more likely to correctly judge pairs of matching expressions ( $M = 0.51$ ,  $SEM = 0.029$ ) vs. mismatching expressions ( $M = 0.40$ ,  $SEM = 0.035$ ),  $t(47) = 3.70$ ,  $p = .001$ ,  $d = 0.55$ . After repeating a neutral label, participants gave a similar number of correct responses when judging matching ( $M = 0.43$ ,  $SEM = 0.032$ ) and mismatching expression pairs ( $M = 0.42$ ,  $SEM = 0.029$ );  $t(47) = 0.48$ ,  $p = .64$ ,  $d = 0.05$ . Additionally, participants gave significantly more correct responses when judging matching expression pairs after they had repeated an emotion label, as opposed to a neutral label,  $t(47) = 2.27$ ,  $p = .028$ ,  $d = 0.36$ . Participants gave a similar number of correct responses when judging mismatching expression pairs, whether they had repeated an emotion or neutral label,  $t(47) = 0.87$ ,  $p = .39$ .

These trends did not result from the adoption of a heuristic form of response bias (e.g. Smith & Klein, 1990; Black, 2001; Black et al., 2013). When repeating emotion labels, participants were significantly more likely to correctly respond 'match' to matching expression pairs ( $M = 0.51$ ,  $SEM = 0.035$ ) than to incorrectly respond 'match' to mismatching face pairs ( $M = 0.27$ ,  $SEM = 0.030$ ),  $t(47) = 4.31$ ,  $p < .001$ ,  $d = 1.08$ . This suggests that participants did not automatically respond 'match' following the repetition of

an emotion label due to the shared emotional content between labels and facial expressions.

Conversely, when repeating neutral labels, participants were no more likely to correctly respond 'mismatch' to mismatching expression pairs ( $M = 0.42$ ,  $SEM = 0.029$ ) than to incorrectly respond 'mismatch' to matching expression pairs ( $M = 0.37$ ,  $SEM = 0.027$ ),  $t(47) = 1.38$ ,  $p = .17$ .

### 3.3.2.3 Conditional Accuracy Analysis

#### 3.3.2.3.1 Repetition-based effects (predicted)

The main effect of repetition was non-significant,  $F(1, 46) = 0.27$ ,  $p = .60$ ,  $MSE = 0.053$ ,  $\eta_p^2 = 0.006$ . Participants were similarly accurate when their responses followed 3 repetitions ( $M = 0.65$ ,  $SEM = 0.036$ ) vs. 30 repetitions ( $M = 0.64$ ,  $SEM = 0.040$ ).

There was no interaction observed between repetition and word type,  $F(1, 46) = 0.45$ ,  $p = .51$ ,  $MSE = 0.041$ ,  $\eta_p^2 = 0.010$ . Planned comparisons showed that participants did not differ in accuracy when their responses followed 3 ( $M = 0.68$ ,  $SEM = 0.022$ ) and 30 repetitions of an emotion label ( $M = 0.66$ ,  $SEM = 0.027$ ),  $t(47) = 0.84$ ,  $p = .41$ . In addition, participants did not differ in accuracy when their responses followed 3 ( $M = 0.61$ ,  $SEM = 0.026$ ) and 30 repetitions of a neutral label ( $M = 0.62$ ,  $SEM = 0.027$ ),  $t(47) = 0.05$ ,  $p = .96$ .

Specific analysis of the 'satiation comparison' conditions revealed no significant difference in accuracy rates when participants encountered matching expression pairs, after repeating an emotion label 3 vs. 30 times,  $t(47) = 1.07$ ,  $p = .29$  (see condition means presented in Table 3.3, rows 3-4 of column 1). Nor did participants differ in accuracy when they repeated an emotion label which matched only one of the expressions, 3 vs. 30 times,  $t(47) = 0.22$ ,  $p = .83$  (see condition means presented in Table 3.3, rows 3-4 of column 2).

#### 3.3.2.3.2 Other significant and marginal effects

There was a significant main effect of Word Type,  $F(1, 46) = 4.30$ ,  $p = .044$ ,  $MSE = 0.066$ ,  $\eta_p^2 = 0.086$ . Participants were less accurate in trials preceded by repetition of neutral labels ( $M = 0.62$ ,  $SEM = 0.039$ ) vs. emotion labels ( $M = 0.67$ ,  $SEM = 0.038$ ).

Although no interaction was observed between word type and expression matching [ $F(1, 46) = 2.86, p = .098, \text{MSE} = 0.099, \eta_p^2 = 0.058$ ], a pattern of findings emerged that were similar to those observed in the proportional accuracy analysis, and experiments 1 and 2a. There was facilitated performance in expression match trials, but only when preceded by an emotion label ( $M = 0.72, \text{SEM} = 0.032$ ) vs. neutral label ( $M = 0.61, \text{SEM} = 0.029$ ). No differences were shown in response rates for expression mismatch trials, whether preceded by repetition of an emotion label ( $M = 0.62, \text{SEM} = 0.033$ ) or neutral label ( $M = 0.62, \text{SEM} = 0.033$ ).

### 3.3.2.4 Timed-out trial analysis

#### 3.3.2.4.1 Repetition-based effects (predicted)

There was a main effect of repetition,  $F(1, 46) = 7.13, p = .010, \text{MSE} = 0.017, \eta_p^2 = 0.013$ . Participants were significantly more likely to be timed out when trials were preceded by 30 label repetitions ( $M = 0.35, \text{SEM} = 0.040$ ), than 3 label repetitions ( $M = 0.32, \text{SEM} = 0.038$ ).

There was no interaction observed between repetition and word type,  $F(1, 46) = 0.012, p = .92, \text{MSE} = 0.014, \eta_p^2 < 0.001$ . Planned comparisons showed that participants were significantly more likely to be timed out in their responses when they followed 30 repetitions of an emotion label ( $M = 0.35, \text{SEM} = 0.042$ ), rather than 3 repetitions ( $M = 0.32, \text{SEM} = 0.036$ ),  $t(47) = 2.15, p = .037, d = 0.11$ . In contrast, participants were similarly likely to be timed out in responses made following 30 repetitions of a neutral label ( $M = 0.35, \text{SEM} = 0.039$ ) and 3 repetitions ( $M = 0.32, \text{SEM} = 0.038$ ),  $t(47) = 1.84, p = .072, d = 0.11$ .

Specific analysis of the ‘satiation comparison’ conditions revealed that participants were significantly more likely to be timed out in their responses following 30 repetitions vs. 3 repetitions of an emotion label that matched both expressions,  $t(47) = 2.03, p = .048, d = 0.15$  (see condition means presented in Table 3.3, rows 5-6 of column 1). In contrast, participants were similarly likely to be timed out in their responses following 3 and 30 repetitions of an emotion label that matched just one expression in the pair,  $t(47) = 1.20, p = .28$  (see condition means presented Table 3.3, rows 5-6 of column 2).

#### 3.3.2.4.2 *Other significant and marginal effects*

There was a main effect of Expression Matching,  $F(1, 46) = 16.70$ ,  $p < .001$ ,  $MSE = 0.019$ ,  $\eta_p^2 = 0.27$ . Participants were significantly more likely to be timed out in their attempts to discriminate between mismatching expression pairs ( $M = 0.37$ ,  $SEM = 0.042$ ) than matching expression pairs ( $M = 0.31$ ,  $SEM = 0.040$ ).

### 3.3.3 Results summary

Replicating experiment 2a, participants were significantly more likely to be timed-out in attempts to respond following 30 (vs. 3) repetitions of a label. In addition, the proportional accuracy analysis showed trends consistent with a massed repetition decrement; specifically, participants were less likely to give a correct response following 30 vs. 3 word repetitions, though not significantly so. In contrast, these trends were again absent in the conditional accuracy analysis. Where repetition-based decrements were present they were common to performance in both expression match and mismatch trials. This contrasts with the pattern of findings in experiment 2a, where massed-repetition decrements were specifically linked to performance in expression mismatch trials.

As in experiment 2a, findings did not suggest that repetition-based decrements were selectively tied to 30 (vs. 3) repetitions of an emotion label. Across the three analyses, no interactions were observed between repetition and word type. Further, planned comparisons conducted on the proportional and conditional accuracy data revealed neither selective decrements after massed repetition of emotion labels, nor strong decrements following the massed repetition of neutral labels. When considering the timed-out trial data, there were indications that participants were more likely to be timed out in their attempts to respond following 30(vs. 3) repetitions of an emotion label, than a neutral label, particularly when that emotion label matched both of the emotional expressions displayed at test (the ‘semantic satiation’ comparison). However, these effects were not robust enough to drive the overall interaction between repetition and word type.

Replicating experiments 1 and 2a, the proportional accuracy data showed that participants performed best in the subset of trials in which all three stimuli matched one another (repeat the word ‘sad’ then see two sad faces). Performance was enhanced relative to expression match trials, when preceded by a neutral label and expression mismatch trials,

when preceded by an emotion label. Facilitated performance was observed following 3 and 30 repetitions of a judgment-relevant, matching label and could not be explained by the adoption of a systematic response bias (i.e. responding 'match' to all trials preceded by an emotion label, Smith & Klein, 1990). Although the data indicated a small bias to respond 'mismatch' to expression pairs preceded by a neutral label, this pattern of responding cannot explain the overall pattern of facilitative effects. Similar trends were found in the means for the conditional accuracy data but here no interaction was observed between expression matching and word type.

Replicating experiment 2a (but not experiment 1), neither the proportional or conditional accuracy data showed concurrent facilitation in 'parallel' trial types, in which all three stimuli mismatched each other (neutral mismatch trials e.g. repeat the word 'Desk' then see an 'Angry' and 'Sad' face).

### **3.4 Integrated Discussion (intermixed and blocked experiments)**

Lindquist et al's. (2006) design was modified in an important way in experiments 2a and 2b. Adding neutral/unrelated baseline trials (e.g. Smith & Klein, 1990) enabled an assessment of whether general massed decrements, found in both Lindquist et al. (2006) and experiment 1, arise as a result of semantic (spreading activation, Collins & Loftus, 1975) or non-semantic mechanisms (e.g. cognitive fatigue, Tian & Huber, 2010). In sum, a semantic account would be favoured if repetition decrements followed only 30 (vs. 3) repetitions of emotion labels, whereas a non-semantic account would be favoured if repetition decrements followed both 30 (vs. 3) repetitions of emotion and neutral labels.

#### **3.4.1 Carry-over effects and trial presentation**

The way in which baseline trials were presented differed across experiments; in experiment 2a neutral label trials were randomly intermixed alongside emotion label trials and, in experiment 2b, emotion and neutral labels were presented in separate experimental blocks. Arguably, adoption of a blocked testing format would reduce possible carry-over effects, possible when participants completed neutral and emotion label trials in close temporal proximity to one another (e.g. Blaxton & Neely, 1983; Smith & Klein, 1990; Holle et al., 1997; Pilotti et al., 1997; Black, 2001; McKenna & Sharma, 2004). Importantly none

of the effects of interest in experiment 2b (predicted or otherwise) were qualified by an interaction with block-order (i.e. whether the participant first encountered the block of emotion-label trials, or neutral-label trials). This lessens the likelihood that the results of experiment 2a (common performance across neutral and emotion label trials) are a methodological artefact, stemming from the decision to randomly intermix the two types of trial.

### **3.4.2 Repetition-based decrements: construction vs. non-semantic accounts**

In addition, three common trends emerged in the data from experiments 2a and 2b. First, participants across both experiments were more likely to be timed out in decisions that followed 30 (vs 3) repetitions of any label (emotion or neutral). There were also indications that participants were less likely to submit correct responses following 30 (vs 3) repetitions of a label (proportional accuracy analysis; statistically significant in E2a with weaker, congruent trends shown in E2b). In neither experiment were massed-repetition decrements apparent in the conditional accuracy analysis. That repetition-based decrements were only statistically significant in the proportional accuracy analysis (E2a, but not E2b) and timed-out trial analyses may suggest that massed verbal repetition decreases the ability to make a timely response, rather than increasing the likelihood to make an incorrect response.

Second, results from neither experiment 2a nor 2b suggested that repetition-based decrements were explicitly tied to the massed repetition of emotion labels. Across analyses from both experiments, there were no interactions observed between repetition and word type. Planned comparisons showed that participants were significantly more likely to be timed out in their attempts to respond after 30 (vs. 3) repetitions of both emotion and neutral labels. In both experiments 2a and 2b, participants were less likely to submit a timely response when they repeated 30 (vs. 3) times an emotion label that matched both expressions (the 'semantic satiation' comparison, Smith & Klein, 1990). That the semantic satiation comparison was only significant for one accuracy measure provides weak support for the construction position.

Only in experiment 2a did participants also show a tendency to be less accurate (proportional accuracy analysis), and fail to submit a larger proportion of responses, after



repeating an emotion label that matched just one expression 30 (vs. 3) times. This highlights an important inconsistency in the results from the two baseline experiments. While the massed repetition decrements in experiment 2a were exclusively driven by performance in expression mismatch trials, decrements were common to performance in both expression mismatch and match trials in experiment 2b. Additionally, in experiment 2a, the conditional accuracy analysis revealed that participants performed worse after 3 vs 30 repetitions of a label when they responded to matching expression pairs. Although Lindquist et al. (2006, experiment 2) demonstrated that participants were both slower and less accurate to respond to mismatching expression pairs, performance decrements were common to 3 and 30 repetitions of an emotion label (whether it matched one or neither of the expressions displayed). Further, Lindquist et al. (2006, experiment 3) did not find an interaction between repetition and expression matching (see Table 2.2).

Inconsistencies in the results from both baseline experiments may reflect a relative lack of statistical power in experiment 2b; here block order was added to the analysis as an additional between-participants factor, perhaps reducing the possibility to find a massed repetition decrement that was qualified by expression matching. Alternatively, when there is a degree of unpredictability inherent in the task, such as not knowing whether to expect to encounter a neutral or emotion label in the intermixed (vs blocked) experiment, it is possible that participants form a different, or less 'flexible' set of response strategies towards the two trial types (e.g. Pilling et al., 2003; Roberson et al., 2007). Perhaps a difference in strategy adoption somehow interacts with the way in which participants approach mismatching vs. matching expression pairs.

Overall, the results from experiments 2a and 2b provide stronger evidence for a 'generalised' massed repetition decrement, rather than a decrement tied explicitly to the repetition of emotion labels. As such a 'non-semantic' vs. 'semantic' account is favoured; participants likely show repetition-based decrements because 30 (vs. 3) repetitions of any linguistic label increase levels of cognitive fatigue and/or response uncertainty (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013), rather than reducing access to emotion-related semantic knowledge (e.g. Lindquist et al., 2006; Barrett et al., 2007; Gendron et al., 2012).

### **3.4.3 Facilitative effects: exploring the interaction between word type and expression matching**

It is also important to consider a third, common trend. When observing proportional and, to a lesser extent, conditional accuracy rates, results from both experiments 2a and 2b show that participants were most accurate in response to trials where all three stimuli matched (e.g. repeat the emotion label 'sad', then respond to a pair of matching 'sad' expressions). This partially replicates the facilitation found in experiment 1 and may provide further evidence for the suggestion that participants attend to overtly repeated emotion labels to aid expression discrimination (e.g. Lupyan, 2007; Roberson et al., 2007). Proportional accuracy data from experiments 2a and 2b confirm that facilitation was relative to performance in expression match trials, when preceded by repetition of a neutral label, and expression mismatch trials, when preceded by repetition of an emotion label. The same trends were present in the conditional accuracy data for experiment 2a, but did not reach statistical significance in experiment 2b, which perhaps supports the notion that this study lacked statistical power. In the proportional accuracy data, and conditional accuracy data for experiment 2a, facilitation was present after both 3 and 30 repetitions of a judgment-relevant label and could not be explained by a systematic bias to respond 'match' to expression pairs preceded by an emotion label (e.g. Smith & Klein, 1990).

Discussion of a related finding is also important. Planned comparisons conducted for experiments 2a and 2b showed no evidence of a corresponding repetition-based decrement in this subset of trials (the 'semantic satiation' comparison conditions), for either proportional or conditional accuracy rates. In both experiments, when participants repeated an emotion label that matched both expressions they were only more likely to be timed out (rather than less accurate) when repeating that label 30 (vs 3) times. As such, it is unclear whether this pattern of results (facilitation at both levels of repetition, without a corresponding decrement) reflects the 'satiation anomaly', used to explain similar findings in experiment 1 (e.g. Smith & Klein, 1990; Black, 2001; Tian & Huber, 2010; Kuhl & Anderson, 2011; Black et al., 2013, see pages 47–48).

In contrast, neither baseline experiment showed evidence of corresponding facilitation in 'parallel' trial types, where all three stimuli mismatched each other (neutral mismatch trials

e.g. repeat the word 'Desk' then see an 'Angry' and 'Sad' face). This differs from the pattern of results obtained in experiment 1.

There is an important difference between the judgment-irrelevant, mismatch trials used in experiment 1, and the neutral, mismatch trials used in experiments 2a and 2b; the emotional connotations of the repeated label. For experiment 1 it was briefly suggested that facilitation may arise in trial types where all three possible 'comparisons' between the presented stimuli yield a congruent result ('match' in judgment-relevant, match trials and 'mismatch' in judgment-irrelevant, mismatch trials). Arguably, label-to-face comparisons can only be conducted, or regarded to deliver useful information, when emotion labels are used, as they are in experiment 1. The implications of this difference in trial type are further discussed in Chapter 5.

## 3.5 Experiment 3: Intermixed presentation of neutral and non-words

### 3.5.1 Introduction

Following previous semantic satiation work, neutral, or ‘unrelated’ trials, were added as a baseline condition in experiments 2a and 2b (e.g. Smith & Klein, 1990). Arguably, massed repetition of neutral labels should not directly inhibit the semantic nodes and connections important for the representation of emotion-specific knowledge. A large proportion of the semantic satiation (e.g. Smith, 1984; Smith & Klein, 1990; Balota & Black, 1997) and semantic interference literature (e.g. Brown et al., 2005; Damian & Als, 2005; Belke & Steilow, 2013; Higgins & Johnson, 2013) support this claim, suggesting that inhibition or interference will predominantly occur after repeated exposure to judgment-relevant semantic material.

However, other researchers argue that semantic inhibition may be possible after repeated exposure to any lexical item, provided that it has semantic grounding. Using a semantic satiation paradigm, Pynte (1991; experiment 1) showed an increase in decision latencies after 30 (vs. 3) repetitions of category names which were both related and unrelated to a single target word, the ‘satiation decrement’ being marginally more pronounced after repetition of unrelated items. Satiation effects were tied to category decision latencies (judging whether the prime and target words came from the same category) and not to a second dependent variable; the time taken for participants to request presentation of the target word after the repetition period. As a result Pynte (1991, see also Frenck-Mestre et al., 1997) suggested that massed repetition decrements result from general inaccessibility to a semantic system and not from decreased attentiveness or increased cognitive fatigue after 30 (vs. 3) repetitions of a word. Arguably participants still need to access the meaning of a repeated category name to make a relatedness decision, even if that category name is not related to the to-be-judged exemplar.

As such, a massed repetition decrement common to emotion and neutral labels may still have a semantic locus. If so, the results from experiments 2a and 2b may still be broadly consistent with a construction account, based on a variant of the spreading activation/inhibition hypothesis (e.g. Lindquist et al., 2006; Gendron et al., 2012). The decision to use a set of neutral labels with a common semantic theme (objects commonly found in an office) may increase this likelihood. Inter-trial repetition of related, neutral

labels may create a specific, yet cumulative hub of semantic interference (Brown et al., 2005; Belke & Steilow, 2013). Across trials, inhibition of two specific areas of the semantic network (emotion-related and object-related) may have additive effects (e.g. Damian & Als, 2005), initiating a type of spreading activation, that stems from various focal points and indirectly limits access to the network as a whole.

This possibility will be explored in the present study by including trials in which participants repeated a set of pronounceable non-words. These non-words will have no semantic basis, thus massed repetition of these items should not inhibit the semantic network, nor create satiation effects. A similar rationale was proposed and supported when words and non-words were compared in previous semantic satiation work (Sekiguchi, 1999). Here standard satiation decrements were observed when participants made an identity matching decision about words they had fixated upon for 30 (vs. 3) seconds, but showed no difference in performance when they fixated a non-word for 30, or 3, seconds. Further support comes from semantic satiation paradigms that include a lexical decision component (e.g. Neely, 1977; Cohene, Smith & Klein, 1978; Smith, 1984). No satiation effects were observed when participants repeated a category label before making a lexical decision about a related word, likely suggesting that lexical decisions do not require access to the semantic system (e.g. Smith & Klein, 1990; Black et al., 2013, but see Dilkina, McClelland & Plaut, 2010; Shivde & Anderson, 2011 ). In other words, when either the repetition or response phase of a task does not require engagement of the semantic network, satiation and/or semantic-based mechanisms should not be used to explain any massed repetition decrements.

Participants encountered neutral labels and non-word trials in an intermixed, random order. The choice to use an intermixed vs. blocked design would increase statistical power and was further justified as experiments 2a and 2b delivered a similar pattern of results, despite differing in testing format. The following predictions are made: If participants only show reduced accuracy after 30 (vs. 3) repetitions of a neutral label, and not a non-word, this would suggest that massed repetition of any label with semantic grounding indirectly inhibits access to a wider network of semantic knowledge (including emotion-specific knowledge). This would be evidenced by an interaction between word type and repetition, or through targeted comparisons to separately assess repetition effects by word type (neutral vs. non-word). The general massed repetition decrements in experiments 2a and 2b may then be interpreted with a semantic locus and viewed as consistent with the

construction position (a variant of the ‘spreading activation’ account e.g. Collins & Loftus, 1975; Smith, 1984). In contrast, a main effect of repetition, common to 30 (vs. 3) repetitions of neutral and non-words, would further favour a non-semantic interpretation, such as increased cognitive fatigue (e.g. Tian & Huber, 2010). This would have implications both for the present results and those obtained for experiments 2a and 2b. No predictions are made regarding facilitation in expression match trials, as the present study does not require the repetition of emotion labels (judgment-relevant or irrelevant).

### **3.5.2 Method**

#### **3.5.2.1 Participants**

Participants were 48 undergraduate students from the University of Leeds (40 female), who received course-related credits for their participation. The sample had a mean age of 20.17 years ( $SD = 2.43$ , Range: 18-30). All participants were native English speakers and 40 participants were right-handed.

#### **3.5.2.2 Materials**

##### *3.5.2.2.1 Words*

A mixture of old and new neutral labels were presented. All neutral labels had a common semantic theme (items commonly found in an office) and were selected from the MRC Psycholinguistic database (Coltheart, 1981; Wilson, 1988): ‘Pen’, ‘Lamp’, ‘Paper’ and ‘Printer’. The label ‘Desk’ (used in experiments 2a and 2b) was replaced with ‘Lamp’ as independent piloting suggested this word was easier to pronounce over consecutive repetitions. In addition, the previously used label ‘Phone’ was replaced with ‘Paper’ as this new label shared both the same number of letters and syllables as its’ emotion label counterpart (‘Anger’). We also added ‘Printer’ as a corresponding neutral label for the emotion label ‘Disgust’ (matched both for number of letters and syllables).

Although emotion labels were not included in the present experiment, we sought to ensure that differences in frequency for the emotion and neutral label sets could not provide an alternative explanation for any divergent effects. Independent t-tests showed that the new set of neutral labels did not differ significantly from the set of four previously used emotion labels (experiment 1) in frequency, either when based on written frequency scores (Kucera & Francis, 1982):  $M_{\text{emotion}} = 52.75$ ,  $SD = 53.32$ ;  $M_{\text{neutral}} = 49.00$ ,  $SD = 72.35$ ;  $t(6) = 0.083$ ,  $p = .94$ ; or spoken frequency scores (Van Heuven et al., 2014):  $M_{\text{emotion}} = 7769.75$ ,  $SD = 6686.17$ ;

$M_{\text{neutral}} = 6280.75$ ,  $SD = 7639.22$ ;  $t(6) = 0.29$ ,  $p = .78$ . Familiarity and concreteness comparisons were conducted for three emotion and neutral labels apiece, as the MRC Psycholinguistic database did not include information for two labels ('Disgust' and 'Printer'). Emotion labels ( $M = 461.67$ ,  $SD = 63.72$ ) and Neutral labels ( $M = 589.0$ ,  $SD = 41.61$ ) did not differ significantly in familiarity,  $t(4) = 1.60$ ,  $p = .19$ . As expected, neutral labels had significantly higher concreteness ratings ( $M = 595.0$ ,  $SD = 22.27$ ) than emotion labels ( $M = 414.67$ ,  $SD = 23.46$ );  $t(4) = 2.27$ ,  $p = .042$ .

Four non-words were selected from the ARC database ('Gup', 'Foap', 'Borbi' 'Besloor'; Rastle, Harrington & Coltheart, 2002). Non-words matched neutral labels for number of letters and syllables, but were judged to be suitably dissimilar in orthography and phonology. Stimuli were pre-judged to be pronounceable over consecutive repetitions. Participants were encouraged to pronounce each non-word as they saw fit.

#### 3.5.2.2.2 *Faces*

Participants encountered face pairs consisting of four different emotional expressions ('Anger', 'Disgust', 'Fear' and 'Sad'). As experiment 3 did not include the repetition of emotion labels it was appropriate to reintegrate expressions of 'Disgust'. First, it was no longer a concern that satiation effects would be diluted, and second, this meant that the pool of available, unique face pairings was re-expanded, enabling an increase in the number of trials, commensurate with that used in experiment 1. Expression pairs were selected from previously derived face sets and the four different expressions were equally distributed across trial types, and trial blocks.

#### 3.5.2.3 **Design**

A 2 (Word Type: Neutral vs. Non-word) by 2 (Repetition: 3 vs. 30) by 2 (Expression Matching: match vs. mismatch) within-participants design was used. Participants completed 128 trials in total, 64 in each trial block. There were eight separate conditions and 16 trials per condition (see trial types in Table 3.4). Participants encountered neutral and non-word trials in a random, intermixed order. Three accuracy measures were used as dependent variables, per condition: proportional accuracy, conditional accuracy and the proportion of timed-out trials (see specific calculations in Chapter 2, page 37–38). Timed out trials accounted for 16.98% of total responses.

### 3.5.2.4 Procedure

As in experiment 1, participants no longer submitted a spacebar press between trials, as they had done in experiments 2a and 2b. The next trial began immediately after the 854ms response window had elapsed for the current trial.

**Table 3.4 Examples of trial types from each of the eight conditions included in experiment 3 (adapted from Lindquist et al., 2006).**

<i>Word</i>	<i>Type of Word</i>	<i>Expression Match</i>	<i>Trial Example</i>
<b>3</b>	<b>Neutral</b>	<b>Match</b>	3 repetitions of the word 'Pen'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Pen'; compare a 'Sad' and 'Angry' expression pair
	<b>Non-word</b>	<b>Match</b>	3 repetitions of the word 'Gup'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 repetitions of the word 'Gup'; compare an 'Angry' and 'Sad' expression pair
<b>30</b>	<b>Neutral</b>	<b>Match</b>	30 repetitions of the word 'Paper'; compare two 'Angry' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Paper'; compare an 'Angry' and 'Disgusted' expression pair
	<b>Non-word</b>	<b>Match</b>	30 repetitions of the word 'Borbi'; compare two 'Fearful' expressions
		<b>Mismatch</b>	30 repetitions of the word 'Borbi'; compare an 'Angry' and 'Fearful' expression pair



### 3.5.3 Results

#### 3.5.3.1 Main Analysis

Separate 2 (Word Type: Neutral vs. Non-word)  $\times$  2 (Repetition: 3 vs. 30)  $\times$  2 (Expression Matching: Match vs. Mismatch) within-participant ANOVAs were conducted on proportional accuracy, conditional accuracy and timed-out trial rates. Bonferroni corrections were applied, per analysis. Following previous experiments (1, 2a and 2b), we first report effects and planned comparisons that address specific predictions, per analysis. Additional significant and marginal effects are then reported. Effects are considered significant if  $p < .05$ , and marginal if  $p \leq .06$  and  $> .05$ . Mean proportional accuracy, conditional accuracy and timed-out trial rates are shown, per condition, in Table 3.5.

**Table 3.5 Mean proportional accuracy, conditional accuracy and timed-out trial rates by Repetition, Word Type and Expression Matching (SEM presented in parenthesis).**  
<sup>11,12</sup>

Dependent Variable	Repetitions	Neutral Label		Non-word	
		Match	Mismatch	Match	Mismatch
<b>Proportional Accuracy</b>	<b>3</b>	0.52 (0.020)	0.49 (0.029)	0.51 (0.022)	0.50 (0.029)
	<b>30</b>	0.50 (0.027)	0.45 (0.029)	0.49 (0.027)	0.43 (0.030)
<b>Conditional Accuracy</b>	<b>3</b>	0.60 (0.025)	0.57 (0.032)	0.58 (0.026)	0.59 (0.027)
	<b>30</b>	0.58 (0.036)	0.55 (0.032)	0.61 (0.033)	0.51 (0.035)
<b>Timed out responses</b>	<b>3</b>	0.12 (0.021)	0.16 (0.028)	0.14 (0.018)	0.18 (0.029)
	<b>30</b>	0.18 (0.027)	0.19 (0.028)	0.18 (0.029)	0.20 (0.029)

<sup>11</sup> Single-samples t-tests revealed that mean proportional accuracy rates for six of the eight condition totals did not significantly differ from chance level (0.50): the neutral word, 3 repetition, expression match condition total ( $p = .41$ ); the neutral word, 3 repetition, expression mismatch condition total ( $p = .53$ ); the neutral word, 30 repetition, expression match condition total ( $p = .92$ ), the non-word, 3 repetition, expression match condition total ( $p = .77$ ), the non-word, 3 repetition, expression mismatch condition total ( $p = .61$ ), and the non-word, 30 repetition, expression match condition total ( $p = .93$ ). Single-samples t-tests also showed that mean proportional accuracy totals fell significantly below chance levels (0.50) for one condition: the non-word, 30 repetition, expression mismatch condition ( $p = .022$ ).

<sup>12</sup> Single-samples t-tests revealed that mean conditional accuracy totals for two conditions did not significantly differ from chance level (0.50): means for the 30 repetition, neutral word, expression mismatch trials ( $p = .15$ ) and the 30 repetition, non-word, expression mismatch trials ( $p = .73$ ).

### 3.5.3.2 Proportional Accuracy Analysis

#### 3.5.3.2.1 Repetition-based effects (predicted)

There was a main effect of repetition,  $F(1,47) = 9.13$ ,  $p = .004$ ,  $MSE = 0.008$ ,  $\eta_p^2 = 0.16$ . Participants were significantly less accurate to respond after 30 repetitions of word ( $M = 0.47$ ,  $SEM = 0.016$ ), than after 3 repetitions ( $M = 0.50$ ,  $SEM = 0.014$ ).

There was no interaction observed between word type and repetition,  $F(1,47) = 0.031$ ,  $p = .86$ ,  $MSE = 0.012$ ,  $\eta_p^2 = 0.002$ . Planned comparisons revealed that participants were significantly less accurate to respond after 30 repetitions of a non-word ( $M = 0.47$ ,  $SEM = 0.017$ ) than 3 repetitions ( $M = 0.50$ ,  $SEM = 0.017$ ),  $t(47) = 2.03$ ,  $p = .048$ ,  $d = 0.25$ . In contrast, participants were similarly likely to provide a correct response after 30 repetitions of a neutral word ( $M = 0.47$ ,  $SEM = 0.016$ ), and 3 repetitions ( $M = 0.50$ ,  $SEM = 0.017$ ),  $t(47) = 1.84$ ,  $p = .072$ .

There were no other significant or marginal effects.

### 3.5.3.3 Conditional Accuracy Analysis

#### 3.5.3.3.1 Repetition-based effects (predicted)

The main effect of repetition was non-significant,  $F(1, 47) = 2.55$ ,  $MSE = 0.022$ ,  $p = .12$ ,  $\eta_p^2 = 0.052$ . Participants were similarly accurate whether they had repeated a word 3 times ( $M = 0.57$ ,  $SEM = 0.028$ ) or 30 times ( $M = 0.56$ ,  $SEM = 0.034$ ).

There was no interaction observed between repetition and word type,  $F(1, 47) < 0.001$ ,  $p = .99$ ,  $MSE = 0.023$ ,  $\eta_p^2 < 0.001$ . Planned comparisons showed that participants were similarly accurate after 3 repetitions ( $M = 0.58$ ,  $SEM = 0.019$ ) and 30 repetitions of a neutral label ( $M = 0.56$ ,  $SEM = 0.012$ ),  $t(47) = 0.94$ ,  $p = .35$ . Participants were also similarly accurate after 3 repetitions ( $M = 0.58$ ,  $SEM = 0.019$ ) and 30 repetitions of a non-word ( $M = 0.56$ ,  $SEM = 0.020$ ),  $t(47) = 1.45$ ,  $p = .16$ .

There were no other significant or marginal effects.

### 3.5.3.4 Timed out trial Analysis

#### 3.5.3.4.1 Repetition-based effects (predicted)

There was a significant main effect of repetition,  $F(1,47) = 12.31$ ,  $p = .001$ ,  $MSE = 0.010$ ,  $\eta_p^2 = 0.21$ . Participants were significantly more likely to be timed out in their responses following 30 repetitions of a word ( $M = 0.19$ ,  $SEM = 0.029$ ) than 3 repetitions ( $M = 0.15$ ,  $SEM = 0.026$ ).

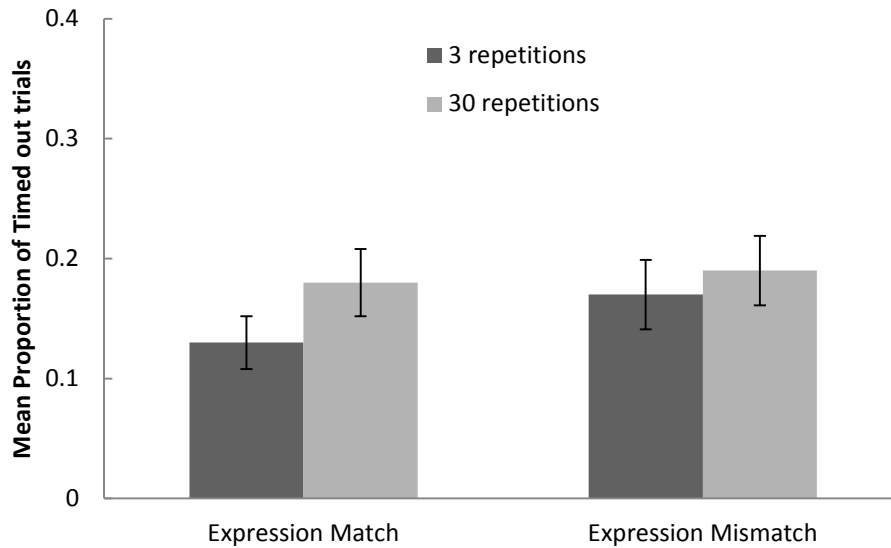
There was no interaction observed between word type and repetition,  $F(1,47) = 1.87$ ,  $p = .18$ ,  $MSE = 0.006$ ,  $\eta_p^2 = 0.038$ . Planned comparisons showed that participants were significantly more likely to be timed out in their attempts to respond after 30 repetitions of a neutral word ( $M = 0.19$ ,  $SEM = 0.019$ ), than 3 repetitions ( $M = 0.14$ ,  $SEM = 0.024$ ),  $t(47) = 3.49$ ,  $p = .001$ ,  $d = 0.29$ . Similarly, participants were significantly more likely to be timed out in responses that followed 30 repetitions of a non-word ( $M = 0.19$ ,  $SEM = 0.028$ ), rather than 3 repetitions ( $M = 0.16$ ,  $SEM = 0.024$ ),  $t(47) = 2.04$ ,  $p = .047$ ,  $d = 0.17$ .

#### 3.5.3.4.2 Other significant and marginal effects

There was a marginal effect of word type,  $F(1,47) = 3.78$ ,  $p = .058$ ,  $MSE = 0.005$ ,  $\eta_p^2 = 0.075$ . Participants were more likely to be timed out in their attempts to respond following repetition of non-words ( $M = 0.18$ ,  $SEM = 0.027$ ) than neutral words ( $M = 0.16$ ,  $SEM = 0.027$ ).

There was a significant main effect of expression matching,  $F(1,47) = 7.28$ ,  $p = .010$ ,  $MSE = 0.009$ ,  $\eta_p^2 = 0.13$ . Participants were significantly more likely to be timed out when responding to mismatching expression pairs ( $M = 0.18$ ,  $SEM = 0.029$ ), than matching expression pairs ( $M = 0.16$ ,  $SEM = 0.025$ ).

The main effect of expression matching was qualified by a significant interaction with repetition,  $F(1, 47) = 4.11$ ,  $p = .048$ ,  $MSE = 0.005$ ,  $\eta_p^2 = 0.080$ . Pertinent marginal means are displayed in Figure 3.6, below.



**Figure 3.6 Mean proportion of timed out trials, by Repetition and Expression Matching (error bars represent SEM).**

Paired samples *t*-tests showed that, when participants encountered trials in which pairs of expressions matched, they were significantly less likely to be timed out in their attempts to respond following 3 repetitions of a word ( $M = 0.13$ ,  $SEM = 0.022$ ) than 30 repetitions of a word ( $M = 0.18$ ,  $SEM = 0.028$ ),  $t(47) = 3.87$ ,  $p < .001$ ,  $d = 0.31$ . However, when responding to mismatching expression pairs, participants were similarly likely to be timed out when responding after 3 repetitions of a word ( $M = 0.17$ ,  $SEM = 0.029$ ) and 30 repetitions of a word ( $M = 0.19$ ,  $SEM = 0.029$ ),  $t(47) = 1.79$ ,  $p = .081$ . After 3 repetitions of a word participants were significantly more likely to be timed out when responding to pairs of expressions that mismatched vs. matched,  $t(47) = 3.50$ ,  $p = .004$ ,  $d = 0.24$ . After 30 repetitions, participants were equally likely to be timed out when responding to pairs of expressions that mismatched vs. matched,  $t(47) = 1.05$ ,  $p = .30$ .

### 3.5.4 Discussion

The present study allowed assessment of whether the general massed repetition decrement, found in experiments 2a and 2b, was still indicative of a form of semantic inaccessibility. Massed repetition of any ‘meaningful’ label, whether emotion-relevant or not, may reduce discrimination performance by ‘globally’ inhibiting access to the semantic network (e.g. Pynte, 1991, Frenck-Mestre et al., 1997). This would perhaps occur via a generalised form of spreading activation/inhibition, sequentially initiated and

simultaneously active at various focal points in the network (e.g. Damian & Als, 2005). This account would be supported if massed repetition decrements were only observed after 30 repetitions of a neutral label, but not a non-word. While non-words can be processed at a lexical, orthographic and phonological level, they lack a corresponding semantic representation. As repeating a non-word fails to initially activate nodes in the semantic network, massed repetition cannot lead to inhibition of that network (e.g. Sekiguchi, 1999, see also Neely, 1977; Cohene et al., 1978; Smith, 1984). However, 30 (vs. 3) repetitions of both neutral labels and non-words may similarly produce a state of enhanced cognitive fatigue, or response uncertainty (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013). These alternative, non-semantic, explanations would be favoured if decrements were common to 30 repetitions of neutral labels and non-words.

#### **3.5.4.1 Repetition-based decrements: construction vs. non-semantic accounts**

Participants were significantly more likely to be timed-out in their attempts to respond following 30 (vs. 3) repetitions of both neutral or non-words. The proportional accuracy analysis showed that participants were significantly less likely to give a correct respond following 30 (vs. 3) repetitions of a non-word, with similar trends shown following the repetition of neutral words. Replicating previous results, the massed-repetition decrement was absent from the conditional accuracy analysis. This may suggest that 30 (vs. 3) repetitions of both neutral and non-words reduced performance by increasing the proportion of trials in which the participant failed to make a timely response, rather than those in which they made an incorrect response.

There was no evidence to support the notion that the massed-repetition decrement was selectively tied to 30 (vs. 3) repetitions of a neutral word, but not a non-word. In all three analyses, there were no interactions observed between word type and repetition. Of interest, while planned comparisons conducted on the proportional accuracy and timed-out trial data showed similar repetition based-decrements following the repetition of neutral and non-words, comparative trends varied. Specifically, the proportional accuracy analysis showed stronger effects following the repetition of non-words (vs. neutral words), while the opposite was true for the timed-out trial analysis. Despite variations in the strength of effects, these results are broadly consistent with a non-semantic proposal; 30 (vs. 3) verbal repetitions of any word may have caused the participant to experience a heightened state of cognitive fatigue and/or response uncertainty, which indirectly

reduced their ability to discriminate between emotional expressions (e.g. Black, 2001; Tian & Huber, 2010). Importantly, the present results cast doubt on the suggestion that general semantic inaccessibility accounts for the general massed repetition decrements observed in the previous experiments reported in this thesis (2a and 2b).

The timed-out trial analysis showed a further interaction between repetition and expression matching. Participants were less likely to be timed out in their responses when they followed 3 (vs 30) repetitions of a word, but only when repetition preceded the presentation of a matching expression pair. This differs from the nature of the interaction between repetition and expression matching, found in the proportional and conditional accuracy data for experiment 2a (see integrated discussion on pages 81–84). Importantly, both experiments 2a and 3 used a design in which different types of word were presented to participants in an intermixed, random format. This perhaps suggests that participants developed different strategies when responding to matching vs. mismatching expressions pairs when they could not predict the type of label that they would repeat from trial-to-trial (e.g. see Pilling et al., 2003 and Roberson et al., 2007 for discussion of how label-based strategies vary when participants cannot predict the type of interference they will encounter from trial-to-trial). Again, it is important to note that Lindquist et al. (2006, experiment 3) did not find an interaction between repetition and expression matching (see Table 2.2). As Lindquist et al. (2006) used only one type of label (emotion words) this may support the notion that this effect is driven by the unpredictably, stemming from the intermixed presentation of different word types.

#### **3.5.4.2 Non-replication of facilitative effects**

As the present work did not include emotion labels it was perhaps unsurprising that no relative facilitation was found in either type of expression match condition (preceded by the repetition of neutral or non-words). As previously discussed, facilitation in the subset of trials in which all three stimuli match (experiments 1, 2a and 2b) may result from the exact match in emotional content between the repeated label and the expressions displayed by both faces within a pair. In Chapter 5 we describe how receiving congruent feedback ('match') from two initial label-to-face comparisons might facilitate a final discrimination between expressions, when that discrimination provides a third, congruent 'match' response. Crucially, neither neutral labels, nor non-words, can inform these initial face-to-label comparisons.





## Chapter 4

### Using 'non-linguistic' baseline measures to explore non-semantic mechanisms for the massed repetition decrement (experiment 4)

#### 4.1 Introduction

The trends observed in four previous experiments suggest that participants show performance decrements after 30 (vs. 3) repetitions of any word, whether an emotion label (experiments 1, 2a and 2b), a neutral label (experiments 2a, 2b and 3) or a non-word, with no semantic basis (experiment 3). Repetition-based decrements appeared to be more robust in the timed-out trial analysis, suggesting that participants found it more difficult to provide a timely response after massed verbal repetition.

Throughout it has been suggested that finding an undifferentiated massed repetition decrement may support a non-semantic and/or non-linguistic interpretation of the data. When using an immediate repetition paradigm, various non-semantic effects may be more common following 30 (vs. 3) verbal repetitions, for example, intra-trial cognitive fatigue or response uncertainty (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013). Focusing on fatigue, participants may show selective decrements after 30 (vs 3) repetitions of a word because massed repetition is cognitively demanding, particularly when required at a consistent and rapid rate (Saito, 1997), and may have a greater impact on attentional resources (e.g. Black, 2001; Tian & Huber, 2010). This may be particularly likely when a 'blinking technique' is used; a methodological aspect carried forward from Lindquist et al's. (2006) work. Here words are flashed on the screen per repetition, creating a new orienting response on each occasion (Smith, 1984). Importantly, this kind of intra-trial fatigue would be equally likely after massed repetition of *any* word.

The potential role of heightened intra-trial cognitive fatigue is specifically addressed in the present study (see also Gendron et al., 2012). The verbal repetition phase will be replaced with an activity that would initially engage the participant's attention, but would likely lead to a similar degree of fatigue after massed exposure. Importantly this activity should not activate, nor later inhibit, the semantic network. Arguably the inclusion of non-words in

experiment 3 allowed us to test this proposal; non-words had no semantic basis, but massed repetition of these items might lead to a state of cognitive fatigue. As an important extension then, a non-linguistic activity was chosen that would have similar, targeted effects.

An appropriate manipulation was selected from the working memory literature: steady-rhythm tapping (e.g. Baddeley, 1986). In these tasks manual tapping is used as a control activity, compared against conditions in which participants engage in 'articulatory suppression' (a task similar to the repetition phase in the semantic satiation work, during which participants overtly recite strings of syllables, or words). Across trials, tapping and articulatory suppression activities are performed concurrently with a task that relies on verbal coding. These tasks may require covert generation of a verbal label to encode or represent visually presented stimuli (e.g. pictures or switching cues) and/or use of subvocal rehearsal to maintain verbal information (e.g. word lists; Baddeley, 1986; Larsen & Baddeley, 2003). Crucially, findings from a number of different tasks suggest that participants show similar 'general' decrements in performance on the verbal task when both types of activity (tapping and articulatory suppression) are performed as a dual-task (e.g. Baddeley, Chincotta & Adlam, 2001; Larsen & Baddeley, 2003; Miyake & Emerson, 2003; Miyake et al., 2004; but see Baddeley, Eldridge & Lewis, 1981; Logie & Baddeley, 1987; Saeki & Saito, 2004- experiment 3). Participants only show 'specific' decrements, indicative of a temporary inability to use verbal coding strategies, following the articulatory suppression manipulation (e.g. loss of the phonological similarity effect when recalling letter strings, Larsen & Baddeley, 2003; failures to verbally translate task cues, Miyake & Emerson, 2003). In other words, both activities (tapping and articulatory suppression) should produce a moderate attentional cost in comparison to baseline conditions, in which the participant simply completes the verbal task without performing an additional activity (Baddeley, 1986; Pashler, Johnson & Ruthruff, 2000). This is at least the case for steady, equal-rhythm tapping, when performed at a rapid rate (see Saito, 1994, 1997; 2001; Saito & Ishio, 1998 and Larsen & Baddeley, 2003 for work with syncopated, complex-rhythm tapping).

Although the present task will include a steady-rhythm tap condition to induce cognitive fatigue, it will differ from past research in two important ways. First, instead of using a dual-task format, in which tapping is performed concurrently with the key task, participants

will perform the tapping activity before they attempt to discriminate between emotional expressions. Despite this modification, it is argued that completion of the tapping activity will still have a transient, negative impact on the participant's cognitive resources, leading to modest decrements in performance. Crucially, repetitive tapping to a steady rhythm demands attention and leads to modest depletions in executive function, even when performed alone (Pashler et al., 2000). As such the manipulation likely creates greater cognitive demands than a baseline condition, where no task is presented in the period prior to expression discrimination (analogous to having no 'dual-task' e.g. Baddeley, 1986; but see Baddeley et al., 1981; Logie & Baddeley, 1987). These demands may translate into residual fatigue, still present during the very restricted time window in which participants perform discrimination. This parallels the rationale posed in experiments 1-3; that massed verbal repetition creates a residual, transient state of cognitive fatigue (and/or semantic interference), which interferes with discrimination in that particular trial. Extending this argument, it is expected that tapping will create greater residual fatigue when performed 30 vs 3 times.

Second, participants in the present work will be asked to repetitively tap with their foot, rather than their hand. This decision was driven by practicality, as correct responses in the present task require an equal number of left and right-hand key presses. Therefore hand tapping may inflate biased response patterns and/or the number of timed out trials observed (e.g. Saito, 1994). Two previous studies have used foot, as opposed to hand tapping, as a control for articulatory suppression conditions (Miyake et al., 2004; Saeki & Saito, 2004). Both studies report similar 'global' performance decrements when comparing conditions in which foot tapping and articulatory suppression activities are performed. These comparisons yield effects of similar magnitude to those reported when comparing hand tapping vs. articulatory suppression in previous work.

It was also important to ensure that foot tapping would be experienced as equally demanding (both physically and cognitively) for all participants (e.g. Miyake et al., 2004). Following Saito and Ishio (1998), we will exclude participants who might find it comparatively easy to retain a steady tapping rhythm (e.g. those with extensive formal music training). Researchers also report that tapping rates are more consistent, and faster, with the dominant vs. non-dominant foot (e.g. Augustyn & Peters, 1986; Fearing, Browning, Corey & Foundas, 2001). As a high proportion of uncrossed lateral dominance is found in

the right-handed, but not left-handed population, only right-handed participants will be recruited (e.g. Peters & Durling, 1979; Augustyn & Peters, 1986; Chapman, Chapman & Allen, 1987). These participants will be asked to repetitively tap with their right (dominant) foot. This will ensure that participants can maintain the tapping rhythm, which must be equivalent to the rapid rate of verbal repetition required in experiments 1, 2a, 2b and 3 (e.g. Baddeley, Lewis & Vallar, 1984; Baddeley, 1986).

In conclusion, participants will be asked to repetitively tap their right foot 3 or 30 times before performing an expression discrimination task. A novel 'watch' baseline was also introduced, during which participants passively watch the screen as 3 or 30 fixation crosses appeared. Crosses are presented at the same intervals as the symbols used to regulate tapping rate. Participants encounter an equal number of 'tap' and 'watch' trials, presented in an intermixed, random order. An intermixed design was chosen to enhance statistical power and to reduce the possibility of extreme physical fatigue, by ensuring that participants do not encounter a long, consecutive sequence of 'tap' trials.

The following predictions are made. If participants only show reduced discrimination accuracy after 30 (vs 3) repetitions of steady-rhythm tapping, but not passive watching, then this would implicate the role of cognitive fatigue in previously observed massed-repetition decrements. Evidence for this effect would primarily be shown by a significant interaction between activity type and repetition. However, planned comparisons will also be used to explore this effect. If these planned comparisons reveal no relative decrements in performance after 30 (vs 3) instances of tapping, or similar decrements after 30 (vs. 3) instances or tapping and watching (a main effect of repetition), then this may indicate that an alternative, non-semantic factor contributes toward the current and previous repetition-based decrements. No additional predictions are made concerning facilitation in expression match vs. mismatch trials. Participants do not repeat emotion labels (relevant or otherwise) in this experiment; therefore, no initial label-to-face comparisons are available to influence critical face-to-face discriminations.

## 4.2 Method

### 4.2.1 Participants

Participants were 40 undergraduate students at the University of Leeds (39 females), who participated for course-related credits. The sample had a mean age of 19.13 years (SD = 0.84, Range = 18-21). The data for two participants were removed (both female). One participant had a high proportion of timed out trials (80%) and the other consistently pressed an unspecified key when making their decisions. The remaining sample all reported to be right-handed, native English speakers. All participants reported to have no extensive, formal music training (e.g. Saito & Ishio, 1998; Saito, 2001), nor existing lower leg/foot injuries on the right side of their body (e.g. tendonitis, shin splints).

### 4.2.2 Materials

#### 4.2.2.1 Face pairs

Following experiments 1 and 3, participants were presented with pairings that utilised all four facial expressions ('Sad', 'Anger', 'Fear' and 'Disgust'). These expressions were equally distributed across trial types and trial blocks.

### 4.2.3 Design

A 2(Activity: Tap vs. Watch) x 2(Repetition: 3 vs. 30) x 2(Expression Matching: Match vs. Mismatch) within-participants design was used. Participants completed 128 trials in total, 64 in each block. There were eight conditions and 16 trials per condition (examples of each trial type are shown in Table 4.1). Trials from each condition were presented in a random, intermixed order. As before, three accuracy measures were calculated for use as dependent variables: proportional accuracy, conditional accuracy and timed-out trial rates (see specific calculations in Chapter 2, pages 37–38). Timed out trials accounted for 20.99% of total responses.

**Table 4.1 Example trials from each of the eight conditions (adapted from Lindquist et al., 2006).**

<i>Repetition</i>	<i>Activity</i>	<i>Expression Match</i>	<i>Trial Example</i>
<b>3</b>	<b>Tap</b>	<b>Match</b>	3 'taps'; compare two 'Sad' expressions
		<b>Mismatch</b>	3 'taps'; compare a 'Sad' and 'Angry' expression pair
	<b>Watch</b>	<b>Match</b>	3 instances of 'watching' the fixation cross; compare two 'Sad' expressions
		<b>Mismatch</b>	3 instances of 'watching' the fixation cross; compare an 'Angry' and 'Sad' expression pair
<b>30</b>	<b>Tap</b>	<b>Match</b>	30 'taps'; compare two 'Angry' expressions
		<b>Mismatch</b>	30 'taps'; compare an 'Angry' and 'Disgusted' expression pair
	<b>Watch</b>	<b>Match</b>	30 instances of 'watching' the fixation cross; compare two 'Fearful' expressions.
		<b>Mismatch</b>	30 instances of 'watching' the fixation cross; compare an 'Angry' and 'Fearful' expression pair

#### 4.2.4 Procedure

Prior to the discrimination task participants were asked to engage in one of two activities ('tap' or 'watch'), cued by on-screen symbols. Participants completed eight practice trials to learn cue-activity mappings. When three dashes (- - -) appeared participants were asked to tap their right foot in rhythm to symbol presentation (one tap, per symbol). Participants were seated in front of the computer screen and were asked to keep their feet shoulder width apart. Participants kept the heel of their right foot placed firmly on the ground and raised and depressed the ball of their foot to complete a single 'tap' (see Peters & Durning, 1979; Augustyn & Peters, 1986 & Fearing et al., 2001 for similar methods). A piece of paper was placed beneath the participant's right foot so that they could keep track of their tapping rate without averting their eyes from the computer screen. The experimenter demonstrated how foot tapping should be performed and instructed the participant to adjust their tapping rate if it did not coincide with symbol presentation. In the remaining 64 trials, participants were presented with a fixation cross (+). When this symbol appeared, participants were asked to passively watch as the symbol repeated, and to refrain from counting the symbols. Throughout the experiment each symbol remained on the screen for 500ms per repetition, followed by a 200ms inter-stimulus interval, in which a blank screen was displayed. This matched the stimulus-display durations used in previous verbal repetition experiments. Following 3 or 30 presentations of either symbol, an asterisk (\*) was presented for 500ms, signalling the onset of the discrimination task. This replicated the procedure used during experiments 1-3. Participants were required to press the spacebar following the end of the 854ms response window to progress from one trial to the next. This ensured that inter-trial intervals were self-paced and potentially variable. Please note that full participant instructions are provided in Appendix C.

The random, intermixed trial order ensured that participants were not required to successively engage in the tapping manipulation for prolonged periods of time. However, participants were also encouraged to take a five minute break between the trial blocks and were told that they could refrain from completing a tapping trial if they experienced any pain or discomfort. Participants were asked not to make a response during the discrimination task if they had skipped the tapping manipulation for that particular trial. No participant skipped more than 5% of tapping trials and no exclusions were made on this basis.

## 4.3 Results

### 4.3.1 Main Analysis

Separate 2 (Activity: Tap vs. Watch)  $\times$  2 (Repetition: 3 vs. 30)  $\times$  2 (Expression Matching: Match vs. Mismatch) within-participant ANOVAs were conducted on proportional accuracy, conditional accuracy and timed-out trial rates. Bonferroni corrections were applied, per analysis.

Per analysis, effects and planned comparisons that address specific predictions are first reported, followed by any additional significant and marginal effects. We consider effects significant if  $p < .05$ , and marginal if  $p \leq .06$  and  $> .05$ . Mean proportional accuracy, conditional accuracy and timed-out trial rates are displayed, by condition, in Table 4.2.



**Table 4.2 Mean proportional accuracy, conditional accuracy and timed-out trial rates, by Repetition, Activity Type and Expression Matching (SEM presented in parenthesis).**

Dependent Variable	Repetitions	Tap Activity		Watch Activity	
		Match	Mismatch	Match	Mismatch
Proportional Accuracy	3	0.47 (0.029)	0.53 (0.028)	0.45 (0.031)	0.46 (0.026)
	30	0.43 (0.029)	0.49 (0.031)	0.46 (0.031)	0.42 (0.026)
Conditional Accuracy	3	0.54 (0.031)	0.62 (0.031)	0.58 (0.032)	0.61 (0.031)
	30	0.56 (0.024)	0.62 (0.037)	0.61 (0.029)	0.56 (0.034)
Timed out responses	3	0.13 (0.020)	0.14 (0.018)	0.20 (0.030)	0.23 (0.030)
	30	0.20 (0.029)	0.19 (0.029)	0.25 (0.036)	0.26 (0.029)

### 4.3.2 Proportional Accuracy Analysis

#### 4.3.2.1 Repetition-based effects (predicted)

There was a main effect of repetition,  $F(1,37) = 7.79$ ,  $p = .008$ ,  $MSE = 0.008$ ,  $\eta_p^2 = 0.18$ .

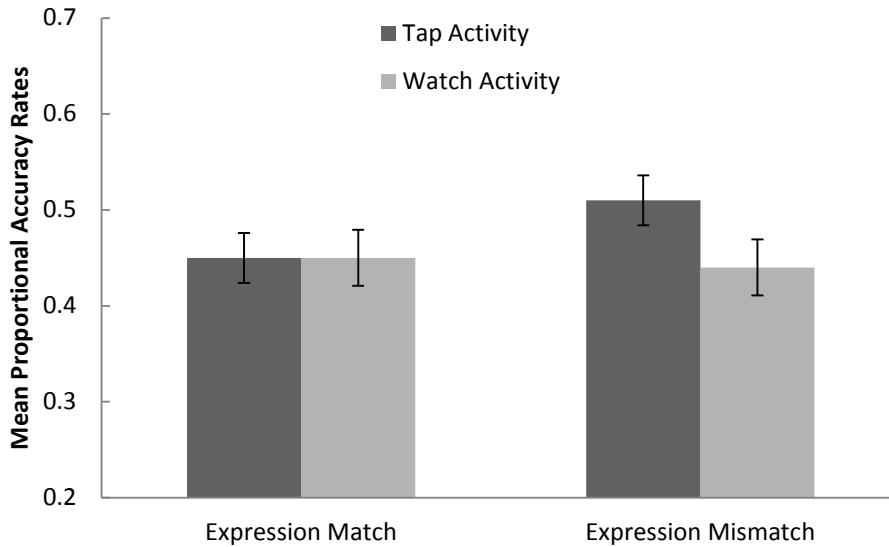
Participants were significantly less likely to give correct responses after engaging in either activity 30 times ( $M = 0.45$ ,  $SEM = 0.029$ ), as opposed to 3 times ( $M = 0.48$ ,  $SEM = 0.029$ ).

There was no interaction observed between activity type and repetition,  $F(1, 37) = 0.43$ ,  $p = .52$ ,  $MSE = 0.017$ ,  $\eta_p^2 = 0.012$ . Planned comparisons revealed that participants were marginally less likely to submit a correct response after tapping 30 times ( $M = 0.46$ ,  $SEM = 0.021$ ) vs. 3 times ( $M = 0.50$ ,  $SEM = 0.019$ ),  $t(37) = 1.94$ ,  $p = .060$ . In contrast, participants submitted a similar number of correct responses after 30 ( $M = 0.44$ ,  $SEM = 0.019$ ) and 3 instances of passive watching ( $M = 0.46$ ,  $SEM = 0.021$ ),  $t(37) = 1.10$ ,  $p = .28$ .

#### 4.3.2.2 Other significant and marginal effects

There was a significant main effect of activity type,  $F(1, 37) = 4.90$ ,  $p = .033$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.17$ . Participants were significantly less likely to respond correctly after engaging in the passive watch activity ( $M = 0.45$ ,  $SEM = 0.029$ ), than the tapping activity ( $M = 0.48$ ,  $SEM = 0.029$ ).

This effect was qualified by interaction between Activity type and Expression Matching,  $F(1, 37) = 7.42, p = .010, MSE = 0.013, \eta_p^2 = 0.17$ . Marginal means are displayed in Figure 4.1, below.



**Figure 4.1 Mean, proportional accuracy rates by Activity and Expression Matching (error Bars represent SEM).**

Paired samples t-tests were conducted to explore the interaction. When responding to mismatching expression pairs, participants were significantly more likely to give a correct response following the tap activity ( $M = 0.51, SEM = 0.026$ ), as opposed to the watch activity ( $M = 0.44, SEM = 0.029$ ),  $t(37) = 3.30, p = .002, d = 0.42$ . All other comparisons were non-significant: Tap Match ( $M = 0.45, SEM = 0.024$ ) vs. Watch Match ( $M = 0.45, SEM = 0.029$ ),  $t(37) = 0.32, p = .75$ ; Tap Match vs. Tap Mismatch,  $t(37) = 1.44, p = .16$ ; Watch Match vs. Watch Mismatch,  $t(37) = 0.40, p = .70$ .

### 4.3.3 Conditional Accuracy Analysis

#### 4.3.3.1 Repetition-based effects (predicted)

The main effect of repetition was non-significant,  $F(1, 35) = 0.003, p = .96, MSE = 0.013, \eta_p^2 < 0.001$ . Participants were similarly accurate to respond following 3 repetitions ( $M = 0.59, SEM = 0.031$ ), and 30 repetitions of an activity ( $M = 0.59, SEM = 0.031$ ).

There was no interaction observed between repetition and activity type,  $F(1, 35) = 0.45, MSE = 0.017, p = 0.51, \eta_p^2 = 0.013$ . Planned comparisons showed that participants did not

differ in accuracy after engaging in the tapping activity 3 times ( $M = 0.58$ ,  $SEM = 0.015$ ) or 30 times ( $M = 0.58$ ,  $SEM = 0.019$ ),  $t(37) = 0.034$ ,  $p = .97$ . Nor did participants differ in accuracy when they engaged in the passive watch activity 3 times ( $M = 0.60$ ,  $SEM = 0.019$ ) or 30 times ( $M = 0.59$ ,  $SEM = 0.019$ ),  $t(37) = 0.44$ ,  $p = .66$ .

#### 4.3.3.2 *Other marginal and significant effects*

There was a significant interaction observed between activity type and expression matching,  $F(1, 35) = 5.49$ ,  $p = .025$ ,  $MSE = 0.020$ ,  $\eta_p^2 = 0.14$ . However, subsequent paired samples t-tests revealed that all comparisons of the pertinent conditions were non-significant: Tap Match ( $M = 0.54$ ,  $SEM = 0.024$ ) vs. Tap Mismatch ( $M = 0.62$ ,  $SEM = 0.029$ ),  $t(37) = 1.72$ ,  $p = .094$ ; Watch Match ( $M = 0.60$ ,  $SEM = 0.028$ ) vs. Watch Mismatch ( $M = 0.59$ ,  $SEM = 0.028$ ),  $t(35) = 0.20$ ,  $p = .85$ ; Tap Match vs. Watch Match,  $t(36) = 1.73$ ,  $p = .092$ ; Tap Mismatch vs. Watch Mismatch,  $t(36) = 1.06$ ,  $p = 0.30$ .

### 4.3.4 Timed-out trial analysis

#### 4.3.4.1 *Repetition-based effects (predicted)*

There was a main effect of repetition,  $F(1, 36) = 11.99$ ,  $p = .001$ ,  $MSE = 0.015$ ,  $\eta_p^2 = 0.25$ . Participants were significantly more likely to be timed out when asked to respond after 30 repetitions of an activity ( $M = 0.23$ ,  $SEM = 0.031$ ) than 3 repetitions ( $M = 0.18$ ,  $SEM = 0.024$ ).

There was no interaction observed between repetition and activity type,  $F(1, 36) = 0.56$ ,  $p = .46$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.015$ . Planned comparisons revealed that participants were significantly more likely to be timed out in their attempts to respond after 30 instances of tapping ( $M = 0.19$ ,  $SEM = 0.026$ ), than 3 instances, ( $M = 0.13$ ,  $SEM = 0.017$ ),  $t(37) = 2.96$ ,  $p = .005$ ,  $d = 0.46$ . Similarly, participants were significantly more likely to be timed out in their attempts to respond after 30 instances of passive watching ( $M = 0.26$ ,  $SEM = 0.031$ ) than 3 instances ( $M = 0.23$ ,  $SEM = 0.029$ ),  $t(37) = 2.03$ ,  $p = .05$ ,  $d = 0.16$ .

#### 4.3.4.2 *Other significant and marginal effects*

There was a significant main effect of Activity Type,  $F(1, 36) = 25.78$ ,  $p < .001$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.42$ . Participants were significantly more likely to be timed out following engagement in the watch activity ( $M = 0.24$ ,  $SEM = 0.031$ ), than the tap activity ( $M = 0.17$ ,  $SEM = 0.024$ ).

## 4.4 Discussion

The present study explicitly addressed whether residual cognitive fatigue could be identified as a non-semantic factor that contributes toward the massed repetition decrements, observed in experiments 1-3 (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013). In order to extend findings from experiment 3, which utilised non-words, a non-linguistic manipulation was used to induce a residual state of cognitive fatigue, while leaving semantic resources intact: steady-rhythm tapping (e.g. Baddeley, 1986). If cognitive fatigue drives repetition-based decrements then participants were expected to be less accurate following 30 vs. 3 repetitions of foot tapping, but not after 30 vs. 3 instances of passively watching, as a fixation cross repeatedly appeared on the screen. However, if participants did not show relative performance decrements after 30 vs 3 foot taps, or showed similar decrements after 30 repetitions of tapping and watching, then this would prompt discussion of other non-semantic/linguistic factors that might contribute toward the massed repetition decrement in this and previous work.

### 4.4.1 Repetition-based decrements: exploring the potential role of residual cognitive fatigue

Results suggest that participants were less likely to give correct responses (proportional accuracy) and more likely to be timed out in their responses following 30 vs. 3 repetitions of either activity (tapping or passive watching). Repetition-based decrements were not observed in the conditional accuracy analysis. Planned comparisons were conducted to assess whether repetition decrements differed, dependent on the activity the participant had engaged in. When considering only the proportional accuracy data, there were weak indications that participants selectively submitted less correct responses after 30 (vs. 3) instances of tapping, as opposed to passive watching. However, there were no repetition-based decrements found in the conditional accuracy data, and the third analysis revealed

that participants were similarly more likely to be timed out in their attempts to respond following 30 (vs. 3) instances of both tapping and passive watching.

Overall, results suggest that participants continued to show a massed repetition decrement even when verbal repetition was not required. There were indications that this decrement was common to both 30 (vs 3) instances of tapping and passive watching, suggesting that residual cognitive fatigue may not be the non-semantic factor responsible for the previously observed massed repetition decrements. As before, 30 repetitions of either activity appeared to decrease the participant's ability to make timely responses, rather than increasing the number of incorrect responses that they submitted.

Despite these findings it may be premature to completely rule out the contributory effects of cognitive fatigue in previous work. First, it is important to note that some researchers document only modest decrements in performance when steady-rhythm tapping is performed (e.g. Saito, 1997; Larsen & Baddeley, 2003). In addition, sometimes performance levels do not differ when comparing conditions in which manual tapping is required as a dual-task, and no dual-task is performed (baseline conditions e.g. Baddeley et al., 1981; Logie & Baddeley, 1987; Saeki & Saito, 2004, experiment 3). If repetitive tapping leads only to modest performance decrements when conducted as a 'dual-task', then it may be unlikely to create a residual state of cognitive fatigue, which affects performance in a subsequent discrimination task. In sum, it is important to question whether the repetitive tapping manipulation represents the best 'test' of cognitive fatigue in our work, given that it may not impose cognitive demands equal to those experienced after the massed verbal repetition, used in experiments 1-3. A further methodological detail elevates this concern. To ensure parity across experiments, participants were asked to tap at the same rate as they had been required to engage in verbal repetition (once every 700ms); a method recommended by Baddeley et al. (1984). However, this meant that manual tapping was conducted at a marginally slower rate than that used in work where general performance decrements have been found e.g. two/three beats per second (Baddeley, 1986; Larsen & Baddeley, 2003).

While it remains unclear whether residual cognitive fatigue contributes towards repetition-decrements in the verbal repetition experiments (1-3), it is important to identify a possible non-semantic factor that can predict decrements after 30 instances of tapping and

watching. Researchers also criticise the immediate repetition paradigm for creating unequal expectancies about when a response is required; participants may experience more uncertainty following 30 (vs 3) repetitions (e.g. Tian & Huber, 2010, but see Balota & Black, 1997; Black, 2001; Black et al., 2013). Crucially, this degree of uncertainty may be present independent of the activity that participants perform. Observational evidence supports this idea. Across all of the semantic satiation experiments, participants appeared to quickly learn that a response was often required after three repetitions of any activity (whether it was word repetition, tapping or passive watching). Behaviourally, participants showed a slight pause in their repetition rate when the fourth repetition was required on a 30 repetition trial. This may reflect automatic processing of number, or ‘subitising’, when a small number of items are presented either simultaneously or sequentially (e.g. Camos & Tillmann, 2008). As such, decrements following 30 repetition trials may reflect the loss of an anticipatory advantage. As response uncertainty may be an inherent consequence of the 30 repetition trial-type, this factor may contribute towards trends found in all five experiments. Arguably, heightened response uncertainty may elevate the proportion of trials in which participants fail to provide a timely response. Potentially supportive then is the observation that, across experiments, the massed-repetition decrement was most apparent in analyses that accounted for the proportion of timed-out trials, rather than those in which participants made an incorrect decision. Of course, without directly comparing performance after 30 verbal repetitions and 30 instances of tapping/watching it cannot be concluded that response uncertainty plays an equal role in the observed repetition decrements, and whether decrements are of a similar magnitude, when different activities prompt uncertainty.

#### **4.4.2 Non-replication of facilitative effects**

In contrast to previous experiments (1, 2a and 2b), participants did not show selective improvements in performance when responding to matching vs. mismatching expression pairs, whether these matching pairs were preceded by the tap or watch activities. The subset of expression match trials were also not associated with facilitated performance in experiment 3, whether trials were preceded by the repetition of neutral or non-words. This supports the suggestion that performance in expression match trials specifically benefits from the label-to-face comparisons that can only be made when emotion labels are presented (experiments 1, 2a and 2b). More specifically, facilitation in performance occurs when the repeated emotion label is congruent with both of the emotional expressions in

the pair; here the congruent feedback from two label-to-face comparisons appears to inform the critical face-to-face discrimination.

However, there were some indications in the present data to suggest that participants responded differently to matching and mismatching expression pairs after they had engaged in the tap, but not the watch activity. Post-hoc tests revealed that these trends were statistically weak and primarily driven by comparative performance in different, single conditions, across accuracy analyses. The proportional accuracy data suggested that participants were more likely to respond correctly when they encountered mismatching expression pairs, but only following engagement in the tap activity. In contrast, the conditional accuracy data suggested that participants were less accurate to respond to matching expression pairs, but only when they were preceded by the tap activity. However, this second trend was not statistically significant. Given the weak and contradictory findings across analyses, no further interpretation of these effects are possible.

#### **4.4.3 Summary of key findings**

In sum, findings from the present study suggest that participants were less likely to correctly respond, and more likely to be timed out, when trials were preceded by 30 (vs. 3) repetitions of either an active tap or passive watch activity. This may weaken the assumption that intra-trial cognitive fatigue is the primary non-semantic factor responsible for the massed-repetition decrement in previous work. Instead, an alternative non-semantic factor, such as response uncertainty, may play a contributory role. This assumption is supported by finding that massed-repetition decrements, in all five experiments, are more robustly shown in analyses that account for the proportion of timed-out trials.





## Chapter 5

### **What can findings from the semantic satiation paradigm tell us about the role of language in emotion perception?**

#### **5.1 Summary of main findings across experiments 1-4**

Five experiments were conducted to investigate the impact of language, and linked semantic knowledge, on the ability to interpret emotional expressions. Following Lindquist et al. (2006) a semantic satiation was employed. In experiments 1-3, participants repeated a label out loud 3 or 30 times before making a simultaneous perceptual discrimination between two emotional faces. By manipulating the type of label that participants encountered we were able to explore whether massed repetition decrements resulted from inaccessibility to semantic knowledge or from alternative, non-semantic factors, inherent within the immediate repetition paradigm (e.g. cognitive fatigue, response uncertainty, Black, 2001; Tian & Huber, 2010; Black et al., 2013). The role of one possible non-semantic factor was explored in a fourth experiment: cognitive fatigue. We extended our previous work by using a non-linguistic manipulation to potentially induce intra-trial fatigue: repetitive foot tapping. Two key trends emerged in the pooled data; one after participants had repeated any label and one after they had repeated emotion labels, only.

First, there were stronger indications that participants showed similar performance decrements after massed repetition of all types of label (emotion, neutral and nonsense), rather than selective decrements, following the repetition of emotion labels. Across the verbal repetition experiments (1-3), massed repetition effects were predominantly found in the proportional accuracy and timed out trial analyses, rather than the conditional accuracy analyses. This may suggest that massed verbal repetition disproportionately led to a decrease in participant's ability to make a timely response, rather than an increase in their commission of incorrect responses. That discrimination errors did not selectively arise following the repetition of emotion labels (relevant and irrelevant) casts doubt on both explanations associated with the construction position; the category-based account and the spreading activation account (e.g. Lindquist et al., 2006; Gendron et al., 2012). Both accounts suggest that massed exposure to emotion labels temporarily reduces access to

semantic knowledge, important for the interpretation of emotional expressions. These findings prompted investigation of a non-semantic/linguistic mechanism in experiment 4. While participants showed decrements in accuracy following 30 (vs 3) repetitions of foot tapping, they also showed similar decrements following 30 (vs 3) instances of passive watching, as a symbol appeared on the screen. This may suggest that the massed repetition decrement arises due to a different, non-semantic mechanism, such as response uncertainty/loss of an anticipatory advantage (e.g. Tian & Huber, 2010). This may explain why the massed repetition decrement was most robust in analyses that considered trials in which participants failed to make a timely response. We further discuss this possibility below.

Second, participants showed facilitated performance when responding to a pair of matching emotional expressions, but only after they had repeated a ‘judgment-relevant’ emotion label (e.g. repeat the word ‘sad’, see two ‘sad’ expressions; experiments 1, 2a and 2b). In all three experiments this relative facilitation was present after both 3 and 30 repetitions of a matching emotion label. Within this subset of trials (the ‘satiation comparison’, Smith & Klein, 1990; Lindquist et al., 2006), there was only evidence of a corresponding, significant decrease in accuracy after 30 (vs 3) repetitions of the judgment-relevant emotion label in experiment 1. As such, it is unclear whether these trends are consistent with the oft-reported semantic satiation anomaly (e.g. Smith & Klein, 1990; Black, 2001; Tian & Huber, 2010; Kuhl & Anderson, 2011; Black et al., 2013; see full discussion of this anomaly in Chapter 2, page 47–48). Below we discuss the possibility that facilitation arises as a result of the three comparisons that the participant can make, per trial (e.g. two label-to-face comparisons and one face-to-face comparison). Accuracy may be heightened when all three comparisons yield a congruent response. As such we discuss the possibility that overtly repeated labels are strategically attended to, and drive expression discrimination in either a direct (e.g. Lupyan, 2007, 2012) or indirect manner (e.g. Pylyshyn, 1999; Roberson & Davidoff, 2000; Kikutani et al., 2008).

## **5.2 Generalised massed repetition effects**

There were strong indications of a massed repetition decrement, unqualified by word type, in two experiments (2a and 3), and, to a lesser extent, in experiment 2b. Specifically, participants appeared worse to perform the expression discrimination task after 30 repetitions, whether the repeated word was an emotion label, neutral label or nonsense

word. In experiments 2a and 3, generalised massed repetition decrements were shown in both the proportional accuracy and timed out trial analyses. Experiment 2b showed evidence of the same decrement in the timed-out trial analysis, while similar trends were observed in the proportional accuracy data. The massed repetition decrement went unobserved in the conditional accuracy analysis for all three experiments (2a, 2b and 3). This may suggest that trends in the proportional accuracy analysis are disproportionately driven by the number of trials in which participants failed to make a response, as opposed to giving an incorrect response. As such, the repetition manipulation may slow discrimination, rather than producing an active bias, whereby participants systematically make incorrect decisions about the expressions they see.

### **5.2.1 Lindquist et al. (2006) and the ‘category-based’ construction account**

Importantly only the methodological replication of Lindquist et al. (2006, experiment 1) provided evidence supportive of the ‘category-based’ construction account. This account suggests that semantic knowledge about ‘discrete’ emotional states is necessary for the interpretation of corresponding emotional expressions. For experiment 1, planned comparisons conducted on the proportional accuracy, conditional accuracy and timed-out trial data revealed that participants were less accurate after satiation of a judgment-relevant vs. irrelevant emotion label. The proportional accuracy and timed-out trial analyses confirmed that these repetition-based differences were stronger in the ‘satiation comparison’ trials, in which the repeated emotion label matched both expressions in a pair (judgment-relevant, expression match trials), rather than just one expression (judgment-relevant, mismatch trials; Smith & Klein, 1990; Lindquist et al., 2006). Previous research suggests that satiation decrements are often stronger in this subset of trials (e.g. Smith & Klein, 1990; Balota & Black, 1997; Tian & Huber, 2010). Here semantic knowledge relevant to both stimuli in a pair is rendered inaccessible, making meaningful discrimination difficult. Interestingly this is the only experiment in which repetition effects were observed in the conditional accuracy analyses; here there were trends suggestive of a massed repetition decrement and a selective decrease in accurate responses following 30 vs 3 repetitions of a judgment-relevant vs. irrelevant emotion label. In experiment 1 then, the repetition manipulation actively influenced the proportion of all three types of response that the participant could make: correct responses, incorrect responses and timed out trials.

Divergent findings may suggest that there is something specific about Lindquist et al.'s (2006) design that produces these selective trends. In comparison to other experiments that included emotion labels (2a and 2b), the replication included an additional 16 trials during which participants repeated a 'judgment-relevant' emotion label, and an additional 80 trials during which participants repeated any emotion label (judgment-relevant and irrelevant). Research suggests that the anticipated satiation decrements, which follow massed repetition of relevant labels, are often modest and statistically weak, requiring a large number of trials to detect (e.g. Smith & Klein, 1990; Black, 2001; Black et al., 2013). In addition, 'spreading activation', resulting from the satiation manipulation, may be cumulative in effect, particularly when a limited number of emotion labels and corresponding expressions are used (e.g. Tian & Huber, 2010; Kuhl & Anderson, 2011; English & Visser, 2014). In the subsequent experiments contained in this thesis, cumulative effects may have been disrupted by (a) interleaving neutral label trials (experiment 2a) or (b) reducing the number of emotion label trials included within the experiment (experiments 2a and 2b).

### **5.2.2 Exploring the interaction between repetition and expression-matching: the impact of trial presentation**

Further, the generalised massed-repetition decrement was found to be less robust in some experiments, than others. Although unqualified by word type, findings from two of the baseline experiments suggest that the massed repetition decrement was qualified by expression matching (experiments 2a and 3). The nature of this interaction was inconsistent across experiments. In experiment 2a, both analyses that considered the proportion of correct responses suggested that participants were only less accurate after 30 (vs. 3) repetitions of a label when they encountered *mismatching* expression pairs. In contrast, in experiment 3, the timed-out trial analysis showed that participants were only significantly more likely to miss the response window after 30 (vs. 3) repetitions when they encountered *matching* expression pairs. In common these experiments (2a and 3) presented two different word types in a random, intermixed order. Arguably participants may adopt different label-driven response strategies when there is an aspect of unpredictability inherent in the task; in this case, the type of to-be-repeated label that would be presented, per trial. Using a categorical perception paradigm both Pilling et al. (2003) and Roberson et al. (2007) report findings consistent with a differential/unstable reliance on covert stimulus labelling when participants could not predict the type of

interference they would encounter prior to discrimination (i.e. visual and verbal interference activities were intermixed across trials). As such, unstable strategy adoption, based on changes in word type, may interact with trial type, causing the participant to respond differently when presented with matching and mismatching expression pairs. Repetition-based decrements went unqualified in experiments 2b and 4. Experiment 2b used a blocked presentation strategy, which may have encouraged adoption of a 'stable' response strategy, which varied according to block rather than trial, or may have reduced the statistical power required to find this interaction. In experiment 4, although activity instructions were intermixed across trials, it is unclear whether participants similarly develop activity-based and label-based strategies for responding.

### **5.2.3 The potential role of non-semantic mechanisms: residual cognitive fatigue, response uncertainty and label inaccessibility**

Nevertheless, despite variations in the robustness of effects across trials that required different decisions (expression match or mismatch), a clear and consistent pattern emerged regarding word type; across experiments 1-3 participants were less accurate after 30 (vs. 3) repetitions of any word. Therefore, an alternative, non-semantic account was put forward to explain undifferentiated massed repetition decrements. Following the criticisms levelled at the immediate repetition paradigm, the potential role of cognitive fatigue was explored in experiment 4; 30 (vs 3) repetitions of a particular activity may reduce accuracy by placing heightened demand on cognitive resources (Black, 2001; Tian & Huber, 2010; Black et al., 2013). Intra-trial fatigue was induced using a repetitive, tapping procedure (e.g. Baddeley, 1986), which should not initially engage, nor later inhibit, the semantic network. No performance differences were found whether participants engaged in 30 repetitions of the steady-rhythm tapping procedure or passively watched as fixation symbols appeared on the screen. However, in comparison to their respective 3-repetition baselines, decrements were found after both 30 instances of tapping and passive watching. These results have several implications.

First, they may suggest that cognitive fatigue does not contribute towards, or account, for repetition-based decrements in the previous verbal repetition experiments (1-3). Though it remains unclear whether repetitive, steady-rhythm tapping and verbal repetition result in the same type and/or degree of cognitive fatigue (e.g. Baddeley et al., 1981; Logie & Baddeley, 1987; Saito & Saeki, 2004, experiment 3) and, therefore, whether tapping

represented the most appropriate manipulation for this purpose. Second, it may be that other non-semantic factors can be implicated in the massed repetition decrement. Feelings of uncertainty about when a response is required and/or losses in anticipatory advantage may also disproportionately occur following 30 (vs. 3) of any activity (Tian & Huber, 2010). The likelihood that these factors contribute to the massed repetition decrement may be supported by the observation that decrements were strongest in the data that accounted for the proportion of trials where participants failed to make a timely response. However, the difficulties of identifying a common, non-semantic factor are acknowledged as experiment 4 did not include a verbal repetition condition for means of comparison. Thus it is unclear whether the massed repetition decrement is of a similar magnitude when it follows both verbal and non-verbal activities.

Second, some researchers have suggested that the general massed repetition decrement in Lindquist et al.'s. (2006) work may be caused by label inaccessibility, rather than semantic inaccessibility (e.g. Gendron et al., 2012; Lindquist et al., 2014). Arguably Lindquist et al.'s. (2006) task requires that participants 'categorically' perceive a pair of emotional expressions to decide whether they do or do not belong to the same semantic category (Harnad, 1987). Categorical perception tasks frequently follow a variant of the 'X-AB' procedure. Here participants are first presented with a single face (X), followed by a simultaneously (Young et al., 1997; Roberson et al., 1999; Roberson & Davidoff, 2000) or sequentially presented pair of faces (AB; Etcoff & Magee, 1992, Calder, Young, Perrett, Etcoff & Rowland, 1996), and must decide whether face X matches face A or B in emotional expression, and identity (e.g. Roberson & Davidoff, 2000). Various researchers suggest that successful discrimination in the X-AB task relies upon an automatic labelling process; participants covertly apply categorical, verbal labels to each face in the triad to supplement suboptimal visual codes when making their comparisons (e.g. Roberson & Davidoff, 2000; Roberson et al., 2007, 2010; Fugate, Gouzoles & Barrett, 2010; though see Goldstone, Lippa & Shiffrin, 2001 and Sauter, LeGuen & Haun, 2011). The important role played by the verbal code is exemplified by the commonly found 'cross-category' advantage (e.g. Roberson et al., 2012): when the faces in the AB pair present a categorical mismatch, participants can quickly reject the incorrect face as it both looks different and can be labelled differently from the initial target face (e.g. when face 'X' displays a predominantly 'sad' expression, face A may display an 80% 'sad'/20% 'fearful' morph, and face B an 80% 'fearful'/20% 'sad' morph). In contrast, a comparison of verbal codes is insufficient for quick rejection when

participants encounter a pair of faces that present a categorical match (e.g. face A displays an 80% 'sad'/'20% 'fearful' morph, and face B, a 70% 'sad'/'30% 'fearful' morph). Under these circumstances an additional, and cognitively demanding, comparison of visual codes is necessary for discrimination (e.g. Roberson & Davidoff, 2000; Roberson et al., 2010). Crucially, the 'cross-category' advantage disappears when participants engage in a verbal interference manipulation either during the initial encoding of the target face or during the retention phase, between seeing the target face and pair of faces. Similar to articulatory suppression, verbal interference may be induced via the overt repetition of letter or word strings (e.g. Hermer, Vasquez, Spelke & Katsnelson, 1999; Roberson & Davidoff, 2000; Roberson et al., 2007; Perry & Lupyan, 2013).

Applying the above to Lindquist et al's. (2006) modified semantic satiation procedure, participants may automatically attempt to covertly label the two faces to perform the discrimination task. This additional step would be unnecessary in a standard semantic satiation paradigm, in which a word pair is commonly presented for judgment, and requires no further representational translation (e.g. Smith & Klein, 1990; Balota & Black, 1997; Black et al., 2013). Arguably, conversion attempts to represent visual stimuli (emotional faces) using lexical category labels may be disrupted following massed repetition of a label; a procedure that closely parallels that used to induce verbal interference and articulatory suppression (e.g. Larsen & Baddeley, 2003; Wickham & Swift, 2006; Lupyan, 2009; Perry & Lupyan, 2013). This may be particularly likely when participants only have a very limited amount time in which to perform discrimination, during which a residual form of verbal interference may still be present. Crucially, we might expect residual verbal interference to occur after the massed repetition of any label, whether emotion-relevant or not (e.g. Roberson & Davidoff, 2000; Winawer et al., 2007; Lupyan, 2009). As such, an account based on label inaccessibility may provide an alternative explanation for the undifferentiated massed repetition decrement, found in experiments 1-3. However this alternate explanation suffers from several limitations:

First, it is unclear how many overt verbal repetitions are required to create verbal interference, and render labels inaccessible (e.g. Perry & Lupyan, 2013). During initial face encoding, some researchers demonstrate that the presentation of a single, incongruent label is sufficient (e.g. Roberson et al., 2007; Suegami & Michimata, 2010), although it is possible that there exists a linear relationship between the length of overt repetition and

the verbal interference effect (but see Roberson & Davidoff, 2000 for caveats). In sum, it is unclear whether 30 repetitions would create a greater degree of initial and residual verbal interference than 3 repetitions.

Second, previous work documents the effects of 'concurrent', rather than 'residual', verbal interference on covert labelling. Verbal interference is commonly applied concurrent with the following stages of the X-AB task: (a) during encoding of face X (e.g. Roberson & Davidoff, 2000; Roberson et al., 2007) (b) during the retention interval between the presentation of face X and faces A and B (e.g. Roberson & Davidoff, 2000), or (c) during the simultaneous discrimination between the face pair, A and B (e.g. Winawer et al., 2007; Suegami & Michimata, 2010). In other words, participants engage in verbal interference exactly when they are required to covertly generate a label during face encoding, or to maintain a label across a retention interval (e.g. Miyake et al., 2004). Evidently this differs from a situation in which 'verbal interference' occurs prior to label generation, as in the semantic satiation task. Overall, these methodological differences reduce the applicability of the label inaccessibility argument for Lindquist et al.'s. (2006) paradigm.

Third, findings from experiment 4 show that the same massed decrements persist in the absence of verbal repetition. Arguably, neither the passive watch nor the manual tapping activities should inhibit systems required for covert label generation. Indeed, the working memory literature uses manual tapping as a control condition for articulatory suppression; while both activities tax attentional resources to a similar extent (e.g. Pashler et al., 2000), only articulatory suppression should interfere with the ability to form, translate or maintain verbal codes (e.g. Baddeley, 1986; Larsen & Baddeley, 2003; Miyake & Emerson, 2003; Miyake et al., 2004). Based on these three limitations, it is unlikely then that label inaccessibility provides a valid, alternative explanation for the pattern of findings obtained across the present semantic satiation experiments.

In conclusion, participants performed consistently worse in the expression discrimination task after they had engaged in 30 (vs. 3) repetitions of an activity, whether that was verbal repetition (experiments 1, 2a, 2b and 3) or a non-verbal activity, such as foot tapping (experiment 4). Failing to replicate Lindquist et al. (2006), performance decrements were not tied to the massed repetition of emotion labels (relevant or irrelevant) when a stronger set of baseline trials were introduced to the paradigm. This limits use of a construction



account to explain behaviour in the semantic satiation paradigm; that participants struggle to perform discrimination when massed repetition of an emotion label renders inaccessible semantic knowledge about emotional states (e.g. Barrett et al., 2007). Instead, non-semantic explanations are favoured to account for the present results, for example, response uncertainty or loss of an anticipatory advantage (e.g. Tian & Huber, 2010).

### **5.3 Facilitation when all three stimuli match: A labelling effect?**

In experiments 1, 2a and 2b participants showed facilitated performance in a particular subset of trials; where all three stimuli matched each other (e.g. repeat the ‘judgment-relevant’ emotion label ‘Sad’, then see two ‘Sad’ faces). Across experiments, facilitative effects were strongest in the analyses that considered the proportion of correct responses made (proportional accuracy) but similar trends were observed in the conditional accuracy data. Participants performed better in expression match trials preceded by a judgment-relevant emotion label in comparison to (a) expression match trials, preceded by a judgment-irrelevant emotion label; (b) expression mismatch trials, preceded by a relevant emotion label (experiments 2a and 2b, only), and (c) expression match trials, preceded by a neutral label. The latter comparison suggests that facilitative effects are emotion-specific; participants only showed better performance when responding to expression match trials when they were preceded by a relevant emotion label, and not by neutral words, non-words, or non-linguistic activities (experiments 3 and 4). Unlike the observed repetition decrements, facilitative effects appear to be driven by the relative proportion of correct and incorrect responses, rather than by fluctuations in the proportion of timed out responses.

Interestingly, facilitation in these ‘satiation comparison’ trials (e.g. Smith & Klein, 1990; Lindquist et al., 2006) was observed at both levels of repetition. In comparison to baseline conditions (‘judgment-irrelevant’ or ‘neutral’ label trials) researchers frequently find a degree of priming following both 3 and 30 repetitions of a judgment-relevant/related word (the semantic satiation anomaly e.g. Tian & Huber, 2010, Kuhl & Anderson, 2011; see full description in Chapter 2, page 47–48). Yet within this subset of trials, researchers also commonly report a corresponding decrease in performance when selectively comparing responses after 30 vs. 3 repetitions of the judgment-relevant label (e.g. Smith & Klein, 1990; Balota & Black, 1997). In other words, although 30 repetitions of a judgment-relevant label should reverse the priming found after 3 repetitions of that word, label-satiation

should not reduce performance to a ‘baseline’ level, found after repetition of judgment-irrelevant/unrelated labels (Tian & Huber, 2010). Across experiments, strong evidence for this corresponding repetition-based decrement was only found in the planned comparisons conducted for experiment 1, and not in experiments 2a and 2b. An alternative explanation is required for this selective facilitation.

Facilitation may be the product of the decision-level comparisons that participants make before they reach the critical face-to-face discrimination. Per trial, two preceding comparisons are available: between the repeated label and face ‘A’, and that same label and face ‘B’. This parallels the decision process available in the X-AB task (e.g. Roberson & Davidoff, 2000). This argument relies on the assumption that the presented emotion label is both attended to and remains accessible following massed repetition (Pynte, 1991; Kuhl & Anderson, 2011). In support, the inhibitory effects of the satiation manipulation are commonly thought to influence the links between the nodes representative of the repeated label itself and semantic knowledge, leaving both of the latter components intact (e.g. Smith, 1984; Tian & Huber, 2010). Crucially trials that include repetition of judgment-relevant emotion labels, prior to presentation of a matching expression pair, are relatively unique in that, if computed correctly, all three comparisons yield the same response (‘match’). We discuss how two initial, matching ‘label-to-face’ comparisons may facilitate the critical ‘face-to-face’ decision. A case is presented both for a ‘direct’ label-feedback account (Lupyan, 2007, 2012) and an ‘indirect’ label-based, strategic account (e.g. Roberson & Davidoff, 2000).

### **5.3.1 Assessing the utility of a ‘direct’, label-feedback account (Lupyan, 2007, 2012)**

A ‘direct’ account is informed by Lupyan’s (2007, 2012) label-feedback model (see section 1.4 for a detailed description). Predominantly explored in relation to object recognition, this model suggests that the frequent pairing of category labels with their perceptual referents means that the simple act of either hearing or saying a category label may have automatic, top-down effects on perception (e.g. Lupyan, 2007; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013). It is important to repeat here the distinction between the construction position (Barrett et al., 2007) and Lupyan’s (2012) label-feedback hypothesis (initially described in the General Introduction, pages 15–20). Although both frameworks suggest that labels are helpful because they activate ‘conceptual’ knowledge about a

referent, constructionists focus on links with an ‘amodal’ repository of meaningful semantic information, whereas the label-feedback hypothesis emphasises specific links with ‘perceptual’ knowledge (e.g. a diagnostic, perceptual template of what a typical example of that object looks like). The word ‘link’ is used tentatively and identifies a second distinction. Construction theorists suggest that words are only indirectly important for emotion perception; they serve their role by activating semantic knowledge, and it is this knowledge that constrains or shapes emotion perception. In contrast, Lupyan (2007, 2012) argue that labels are an intrinsic part of an object’s conceptual knowledge base and can play a direct role in perception, simultaneously accentuating the category-diagnostic features of a display and directing our attention towards them. These effects are possible as category labels lead to transient activation in category-specific cortical areas, providing a head-start to the perceptual system (e.g. Puri & Wojciulik, 2008; Lupyan & Spivey, 2010a; Esterman & Yantis, 2010; Lupyan & Ward, 2013; but see Eger et al., 2007). As such, while satiation of emotion labels should impair access to ‘semantic’ knowledge, it should not impair access to conceptual forms of perceptual knowledge (Lupyan & Thompson-Schill, 2012), which rely only on label accessibility (Lupyan, 2009; Lupyan & Ward, 2013).

This direct account may be used to explain the selective facilitation found in judgment-relevant, match trials. Repeating the emotion label ‘fear’ should automatically prime the participant to identify the perceptual, category-diagnostic features of a fearful expression (e.g. Lupyan, 2007, 2012). When the first expression in a pair shows label-accentuated ‘fearful’ features, feedback loops send strong confirmatory evidence back to the representational, or ‘label’ level. With the relevant label re-activated, participants should again find it easy to identify category-diagnostic features in the second expression. Perceptual warping may take place, abstracting across the idiosyncratic features in each face, and enhancing categorical similarities between them (Lupyan, 2008b; Lupyan & Spivey, 2010b). This supports the decision that the two faces display the same expression. These effects may be particularly likely in a study that uses a small number of emotional expressions (four), as the same category-specific cortical regions are repeatedly activated across trials (Esterman & Yantis, 2010).

However, by the label-feedback account we might expect to find similar facilitation in both expression match and mismatch trials, when preceded by a judgment-relevant label. This was not the case; participants performed significantly better in judgment-relevant,

expression match vs. mismatch trials both in experiments 2a and 2b. Lupyan and Spivey's (2010b) findings provide support for this prediction. These researchers assessed the potentially facilitative effects of category label exposure when participants viewed a simultaneous visual search display, containing a heterogeneous set of items (e.g. a single '2', and a series of '5's). Participants always showed facilitated detection of the target numeral ('2') when a relevant label was played aurally alongside the visual search display. Lupyan and Spivey (2010b) implicated two, simultaneous mechanisms. Label exposure both eased identification of stimuli that possessed category-diagnostic features (e.g. '2's) and enhanced the ability to correctly reject stimuli that did not display category-diagnostic features (e.g. '5's). The perceptual warping mechanism ensures that these effects can occur concurrently; exemplars from the same category appear to increase in perceptual similarity and exemplars from different categories become more disparate (e.g. Gilbert, Reiger, Kay & Ivry, 2008; Lupyan, 2008b). Translated for the present work, expression mismatch trials, preceded by a judgment-relevant emotion label, should provide participants with the opportunity to assess that one face possesses relevant, category-diagnostic features, whereas the other does not. However, Lupyan and Spivey's (2010b) results have not always been replicated. Labels did not lead to facilitated performance in a detection task when the label matched only one of the items in a presented triad (Lupyan, Thompson-Schill and Swingley, 2010). Similarly, when participants assessed whether a label matched a subsequently presented object, performance was only better than a no-label baseline when a relevant, but not irrelevant label, was heard (Lupyan & Thompson-Schill, 2012). In other words, labels only helped participants to confirm a 'match', rather than to reject a 'mismatch'. These inconsistent results may explain why participants did not show equal facilitation in expression-match and mismatch trial types, when preceded by a judgment-relevant emotion label (experiments 2a and 2b).

Experiment 1 delivered a related finding that deserves comment. A cross-over interaction in the proportional and conditional accuracy analysis suggested that participants showed heightened accuracy in both the judgment-relevant, expression match trials and judgment-irrelevant, expression mismatch trials (e.g. repeat the emotion label 'Sad', see a 'Fearful' and 'Angry' face). If conducted correctly, the three comparisons available in both types of trial would result in a congruent, universal outcome; 'match' or 'mismatch', respectively. Importantly, relative, consistent facilitation was not found when mismatch trials were preceded by neutral labels (experiments 2a, 2b and 3), non-words (experiment 3) or non-

verbal activities (experiment 4), which all present a decision-level analogue (i.e. three mismatch responses). This may suggest that participants make strategic use of overtly repeated labels when they are viewed as congruent with task goals, and potentially informative (i.e. emotion labels, Faulkner, Rhodes, Palermo, Pellicano & Ferguson, 2002).

Again, it is unlikely that a 'direct' labelling account can equally accommodate both of the facilitative effects that drive the cross-over interaction. According to Lupyan's (2007, 2012) framework, labels should help us to reject label-incongruent stimuli by facilitating quick recognition that they lack appropriate, category-diagnostic features (e.g. Lupyan, 2008b; Lupyan & Spivey, 2010b, but see Lupyan, Thompson-Schill & Swingley, 2012). In a perceptual discrimination task, participants may need to simultaneously 'reject' or 'accept' two faces, based on their relationship to the repeated label; if both faces 'mismatch' the repeated label they should be rejected, and if they both faces 'match' they should be accepted. The need to make these simultaneous decisions quickly may necessitate a perceptual warping process. As a consequence it should be easier to 'reject' both faces if they both mismatch the repeated label, but match each other (e.g. repeat the label 'sad', then see two 'fearful' expressions; Lupyan, 2008b; Lupyan & Spivey, 2010b). Here, both expressions will lack the labelled category-diagnostic features to an equal extent and will be highly compatible for perceptual warping; the faces are already similar on a categorical level and may present a less 'demanding' warp. As such, a 'direct' labelling account would predict a more effective 'rejection' process, and therefore higher accuracy rates, in response to judgment-irrelevant, match vs. judgment-irrelevant mismatch trials.

### **5.3.2 Assessing the utility of an 'indirect', strategic labelling account**

Predictions made by the label-feedback account rely on a central assumption: that labels actively change how we visually process referent objects (e.g. Lupyan, 2007, 2012). This assumption is not shared by 'indirect' accounts (e.g. Pylyshyn, 1999), which instead argue that the information derived from verbal codes (whether faces receive the same categorical label) simply supplement suboptimal visual information (whether faces 'look' the same; e.g. Roberson & Davidoff, 2000; Pilling et al., 2003; Kikutani et al., 2008). Applied to the present work, 'verbal codes' may be disproportionately used to conduct the two 'label-to-face' comparisons, where 'visual codes' may be disproportionately considered during the single, critical, 'face-to-face' comparison. Assuming that the two 'verbal' label-to-face comparisons can be made, they may not offer any advantage to the visual 'face-to-

face' comparison, as labels do not actively shape visual percepts (e.g. Pylyshyn, 1999). Perhaps then, facilitation in judgment-relevant, match trials and judgment-irrelevant, mismatch trials emerge as the result of a biased decision process. In these trials, participants give 'match' and 'mismatch' responses, respectively, because the two initial label-to-face comparisons converge to give weight to these conclusions. Participants may not actually attempt to make the further, effortful visual comparison between the expressions displayed on the two faces. In other words, in both the judgment-relevant, match and judgment-irrelevant, mismatch trials (experiment 1), correctly making the two initial label-to-face comparisons provides the information necessary to make the correct 'face-to-face' comparison.

In conclusion, participants were better at discriminating in two types of trial; when they encountered a pair of expressions that matched each other, and the repeated emotion label (experiments 1, 2a and 2b) and when they encountered a pair of expressions that mismatched each other, and the repeated emotion label (experiment 1). Arguably both effects suggest that participants attend to the emotion labels that they overtly repeat early in a trial, and that two initial label-to-face comparisons influence the critical discrimination conducted between emotional faces. It currently remains unclear whether 'direct' or 'indirect' labelling explanations provide a better fit for the data. Conducting label-to-face comparisons may 'change' how those faces are perceived, actively influencing discrimination, or may simply provide an additional form of evidence, removing reliance on a visual comparison of those faces. Acceptance of either account may suggest that language (labels) influence emotion perception in ways alternative to those proposed by constructionists (i.e. in and of themselves).

## **5.4 Methodological limitations and future directions**

### **5.4.1 The impact of restricted response windows on accuracy rates**

The present work replicated various methodological aspects of Lindquist et al. (2006, experiment 3). Based on the finding that semantic satiation effects are transient (Frenck-Mestre et al., 1997; Kounios et al., 2000; Gendron et al., 2012), participants were required to make their discriminations within a short response window, while satiation effects were still active (854ms, Lindquist et al., 2006, experiment 3). This allowed assessment of whether participants use semantic information early in the time course of expression

discrimination (e.g. Gendron et al., 2012). However, researchers also note that adoption of a restricted time window may lead participants to favour speed over accuracy in their responding, perhaps forcing sufficient errors to reduce accuracy to, or below, chance levels (50%, Draine & Greenwald, 1998). Across experiments 1-4, single-samples t-tests were used to assess whether this was the case in the present work. When considering the proportional accuracy data for all experiments (1-4), 87.5% of the condition means either did not differ significantly from chance, or fell significantly below chance. Per experiment, at least 75% of the proportional accuracy means were at or below chance levels. Importantly, using different variables to measure accuracy meant that a comparison could be drawn between proportional and conditional accuracy rates. When considering the conditional accuracy data for all experiments (1-4), only 12.5% of the condition means either did not differ significantly from chance, or fell significantly below chance. Per experiment, no more than 25% of the condition means failed to meet or exceed chance levels.

The proportional and conditional accuracy variables differ in their consideration of timed-out trials. Where the calculation for proportional accuracy considers correct responses as a proportion of a total error score, conflating the contribution of timed-out trials and incorrect responses, the conditional accuracy calculation considers only the relative proportion of correct and incorrect responses. Therefore, differences in chance and below level chance performance, across the two dependent variables, suggest that the restricted response window increased the number of instances whereby participants failed to make a timely response, rather than increasing the number of incorrect responses made.

Observing chance and floor level accuracy rates are considerably problematic in the semantic satiation work given that key predictions focus on relative decrements in performance, across the 3 and 30 repetition conditions (see Lindquist et al., 2006). Crucially we may fail to observe some of the anticipated massed-repetition decrements as 69% of the floor level proportional accuracy rates were observed in the 30-repetition conditions. In these conditions then, it may not be plausible for manipulations to drive performance levels any lower. These criticisms also apply to Lindquist et al. (2006; experiment 3), who reported accuracy rates below 50% in every condition of their experiment (see descriptive statistics reported in Table 2.1). However, as an example, we consider one crucial comparison in the present work, which assessed performance across 3 and 30 repetition

conditions (the 'satiation comparison' e.g. Smith & Klein, 1990; Lindquist et al., 2006). A construction account would predict a massed repetition decrement to arise in this subset of trials, given that semantic knowledge relevant to both emotional expressions should be rendered inaccessible (Lindquist et al., 2006). In experiment 1 this comparison revealed a significant difference in proportional accuracy rates, despite chance levels of performance being observed in both the 3 and 30-repetition conditions. The same comparison was not significant when considering the proportional accuracy rates for other experiments that included emotion labels (experiments 2a and 2b). This despite observing accuracy rates at chance levels in experiment 2b and significantly above chance levels in experiment 2a. If satiation differences were apparent in this subset of trials in experiments 2a and 2b, proportional accuracy rates were at an adequate level for these effects to be observed.

Arguably the type of response identified as problematic (timed-out trials) are less pertinent to the research questions posed; without a response we cannot tell whether participants can or cannot correctly discriminate between emotional expressions following satiation. Instead, they may just be slower to do so. Due to the high number of timed-out trials recorded across experiments (between 16.98% - 32.64% of all trials), it may be useful to recommend adoption of an open-ended response window in subsequent semantic satiation work, and use of reaction times as the primary dependent variable (see Lindquist et al., 2006, experiment 1). Adopting these measures will result in fewer losses of data. In addition, although the adoption of a short response window was justified as semantic satiation is a transient effect (e.g. Frenck-Mestre et al., 1997; Kounios et al., 2000; Gendron et al., 2012), it is still plausible to observe satiation decrements when using reaction times (e.g. Smith & Klein, 1990; Balota & Black, 1997; Black, 2001; Black et al., 2013); participants may be unable to make their discriminations while semantic knowledge remains inaccessible, slowing their responses. In particular, although Gendron et al. (2012) used a different decision in their task (repetition priming vs. expression discrimination), they used reaction times as their dependent variable, despite their aim to assess the early influence of language on emotion perception.

#### **5.4.2 The immediate repetition paradigm: limitations and solutions**

As previously described, immediate repetition paradigms are frequently criticised for introducing unequal demands across the 3 and 30 repetition conditions. In addition to semantic satiation effects then, relative decrements in performance following massed verbal repetition may be indicative of a number of other non-semantic factors, for



example, cognitive fatigue, response uncertainty and/or loss of an anticipatory advantage (e.g. Black, 2001; Tian & Huber, 2010; Black et al., 2013). In the present work, these considerations necessitated the introduction of several baseline conditions; most notably, the steady-rhythm tapping manipulation in experiment 4. This non-linguistic manipulation was intended to pose cognitive demands similar to those experienced in the verbal repetition experiments (1-3), however, it is unclear whether this was the case (see Chapter 4, and section 5.2 above; Baddeley et al., 1981; Logie & Baddeley, 1987; Saeki & Saito, 2004).

Previous research may offer a more elegant solution, by equalising the demands posed by the control (3 repetition) and satiation (30 repetition) conditions. For example, Tian and Huber (2010) spread the repetition of to-be-satiated items across a number of trials. Per trial, participants made a speeded word matching, or relatedness decision, between a cue word and target word (e.g. 'apple' and 'pear'). 'Satiation' trials involved the repeated presentation of a cue word; this cue word could be re-presented a total of 10 times during a single trial block. 'Control' trials were intermixed within the block and incorporated 10 unique cue words. Tian and Huber's (2010) findings were indicative of 'associative satiation'; participants showed increased decision latencies when relatedness judgments contained a category exemplar that had been repeated six or more times, and a related target. However, it is unlikely that this procedure would be appropriate for the present work; explicitly deciding whether an emotion label matches a single emotional face is unlikely to test the assumption that language shapes the 'perception' of expression. Conversely, Gendron et al. (2012) present a more appropriate solution in their extension of Lindquist et al's. (2006) work. Participants repeated a word 30 times on every trial. In 'satiation' trials they would repeat the same emotion label 30 times, but in 'control' trials, they would repeat an unrelated/neutral word 27 times, followed by 3 repetitions of an emotion label. This repetition phase preceded implicit decisions about a single, emotional face (see further discussion of Gendron et al., 2012 below).

### **5.4.3 Delineating between the perceptual vs. post-perceptual influence of language on emotion perception: future directions**

The present work also allows for two observations regarding possible post-perceptual influences of language during face discrimination. Although Lindquist et al. (2006) suggested that their discrimination task was purely perceptual in nature, requiring that

participants quickly interpret the emotional expressions displayed on both faces, subsequent researchers suggest that there may be post-perceptual, or decision-level, processes involved (e.g. Gendron et al., 2012; Black et al., 2013; Lindquist et al., 2014). For example, researchers often suggest that participants engage in an automatic, covert labelling process when asked to discriminate between two categorical stimuli, like emotional expressions (e.g. Roberson & Davidoff, 2000; Pilling et al., 2003; Roberson et al., 2007; Gilbert et al., 2008; Winawer et al., 2007). As argued by subsequent researchers, this covert labelling process may have been disrupted under conditions requiring massed verbal repetition; as such the massed repetition decrement may result from residual inaccessibility to labels, rather than semantic knowledge (e.g. Gendron et al., 2012; Lindquist et al., 2014). The present work made two important contributions in this respect.

First, findings from experiment 4 show that participants continued to show massed repetition decrements after 30 (vs. 3) repetitions of two non-verbal activities, which should not prevent the ability to covertly generate labels (e.g. Baddeley, 1986; Wickham & Swift, 2006). This lessens the likelihood that label inaccessibility drives present or past results when using a modified semantic satiation paradigm. Second, there were robust indications in the proportional and conditional accuracy data that participants did employ some form of decision-level strategy when performing discrimination. Across experiments 1, 2a and 2b participants showed facilitated performance in trials where they responded to two matching facial expressions, but only when discrimination was preceded by repetition of a matching emotion label. In conclusion then, rather than covertly generating labels to facilitate discrimination at a decision-level, participants may strategically pay attention to overtly repeated labels. This allows the participant to make two initial label-to-face comparisons before engaging in the critical comparison between the two expressions. Facilitation may arise when the two initial label-to-face comparisons provide a response congruent with an accurate critical decision ('match' or 'mismatch', respectively).

These observations present two considerations for future research. First, researchers must be careful to select a task that can effectively delineate between the perceptual and post-perceptual effects that language can have on the processing of stimuli. Gendron et al.'s (2012) extension of Lindquist et al.'s (2006) work highlights some important parameters for such a task. These researchers used a repetition priming technique in conjunction with a semantic satiation manipulation. Here participants repeated a relevant emotion label 3 or

30 times (see above) before they briefly encountered an emotional face (50ms). After backwards masking, the same face was presented again and participants were asked to make an implicit decision; either 'is this the face you saw before?' (experiment 1) or 'are the actor's eyes close together or far apart?' (experiment 2). In sum, when faces are presented briefly, researchers can monitor the effects of label and semantic accessibility on the initial perception of an emotional face. Indeed, in Gendron et al. (2012), faces were presented during the temporal period in which the satiation manipulation should render semantic knowledge inaccessible. In addition, use of an implicit, or orthogonal decision, meant that participants could not make explicit use of repeated labels to boost accuracy in the task. Variants of these two methodological components are adopted in subsequent work, presented in chapter 6.

Gendron et al.'s. (2012) results provided strong support for the construction perspective. Repetition priming (indexed by decision speed) was only found after 3, and not 30 repetitions, of a relevant label, suggesting that emotional faces are literally perceived differently when the satiation manipulation removes access to emotion-related semantic knowledge during initial encoding (e.g. Lindquist et al., 2006, 2014). However, facilitative results obtained in the present series of experiments (1, 2a and 2b) may suggest that language labels drive the interpretation of emotional expression in ways independent of semantic knowledge. Similar ideas are expressed in Lupyan's (2007, 2012) label-feedback model; simple exposure to a category label may create a strong perceptual expectancy, accentuating the category-diagnostic features of a referent object and directing our attention towards them. By this account, labels themselves may improve the perception of congruent objects. Although the label-feedback hypothesis has been almost exclusively tested in the object recognition literature, the present findings suggest a possible extension of theory to more complex stimuli (emotional faces). Relevant findings are presented in Chapter 6.



## Chapter 6

### Using a threshold method to assess whether language directly shapes initial perception of emotional expressions (experiment 5)

#### 6.1 Introduction

##### 6.1.1 Assessing the *Direct vs. Indirect* influence of language: the importance of methodology

When considering the language-mediated perception literature, it is often difficult to interpret facilitative findings. While they may indicate that labels *directly* shape our perception of an object, they may simply suggest that language *indirectly* supports an inferential process, by which we try to make sense of the ambiguous visual information that we receive (e.g. Barrett et al., 2007; Lupyan & Spivey, 2010b; Gendron et al., 2012; Lupyan & Ward, 2013). In other words, is the object we see in the presence of language any ‘different’ from that we see its absence?

Previously described work suggests that the ability to delineate between *direct* and *indirect* interpretations depends heavily on the methodology employed (Lindquist et al., 2006; Gendron et al., 2012; Chapter 5). Across both the emotional expression and object perception literature there appear to be two key methodological components required to assess ‘direct’ or ‘active’ influences of language on perception. The first methodological consideration is that the visual stimuli may need to be perceptually degraded in some way, through techniques such as threshold viewing durations and immediate backward masking (Lupyan & Spivey, 2010a; Gendron et al., 2012), binocular rivalry (e.g. Anderson et al., 2011) or continuous flash suppression (e.g. Lupyan & Ward, 2013). The second methodological consideration requires that participants make a decision about the visual stimuli that is orthogonal, or only implicitly tied to the category label they have encountered or repeated (e.g. ‘Sad’), prior to visual inspection. For objects this might be a target absent/target present decision (e.g. Lupyan & Spivey, 2010a; Chen & Spence, 2011; Lupyan & Ward, 2013). For emotional faces this will commonly involve a decision about facial attributes, perhaps processed in parallel or close temporal proximity to emotional

information, for example, gender (Roberts & Bruce, 1988; Roesch, Sander, Mumenthaler, Kerzel & Scherer, 2010) or featural information (e.g. ‘are the actor’s eyes close together or far apart?’ (Gendron et al., 2012). Alternatively, a simple emotionality decision may be used, as is common in emotion perception work that does not investigate the role of language; ‘Does the face display an emotional expression?’ (e.g. Szczepanowski & Pessoa, 2007; Bornemann, Winkielman & van der Meer, 2012). Arguably, participants can correctly make an emotionality decision without reliance on an *overtly* encountered label.

Researchers also commonly suggest that valence and arousal judgments (whether an expression is positive or negative, and how intensely that expression is shown) often precede discrete categorisation. This further suggests that emotionality decisions should not rely on *covert* labelling (e.g. Esteves & Ohman, 1993; Russell & Barrett, 1999; Russell, 2003; Lindquist et al., 2014, but see Calvo et al., 2012), which may be considered a post-perceptual process (Lupyan & Ward, 2013). When these two conditions are met, any facilitation from exposure to a language label may be interpreted to mean that that label (and linked conceptual knowledge) has directly ‘boosted’ perception of the referent object, making it recognisable (e.g. Lupyan & Spivey, 2010a; Lupyan & Ward, 2013). Commonly, facilitation will result from exposure to a label which matches the presented stimuli (hear the label ‘chair’, then see a picture of a ‘chair’, Lupyan, 2008a; Lupyan & Spivey, 2010a; Gendron et al., 2012; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013).

The present work aims to meet both of these methodological requirements. In brief, emotional faces will be perceptually degraded, and participants will make an implicit decision about each emotional face, which does not require explicit retrieval of an emotion label. Per trial, participants will encounter a language label before they are presented with a face at ‘threshold’ duration (i.e. the minimum duration required for the participant to detect that the face displays an emotional expression, at above chance level). This face will be immediately backwards masked using a random pattern square, increasing perceptual ambiguity (e.g. Sweeny, Grabowecky, Suzuki & Paller, 2009). Of theoretical importance, the perception of stimuli presented at threshold can still benefit from the top-down influence of semantic forms of conceptual knowledge, which may be reduced when using alternative methods (e.g. continuous flash suppression, Bar et al., 2006; Lupyan & Ward, 2013). Participants will make an implicit emotionality decision about each face, without categorisation; ‘does the face display an emotional expression?’... ‘yes or no?’ (e.g. Szczepanowski & Pessoa, 2007; Bornemann et al., 2012). To further inhibit categorisation

attempts, based on covert labelling, participants will not be made aware of the types of emotional expression that could be displayed across the task (e.g. Szczepanowski & Pessoa, 2007; Szczepanowski, Traczyk, Wierzchoń & Cleeremans, 2013).

### **6.1.2 Manipulating the ‘type’ and ‘relevance’ of linguistic cues**

Following previous research, the ‘relevance’ of the repeated word will be varied, within-participants. Per trial, participants will either repeat a relevant emotion word (e.g. repeat the word ‘Sad’ before seeing a ‘Sad’ face), an irrelevant emotion word (repeat the word ‘Surprise’ before seeing a ‘Sad’ face) or a neutral word, with no emotional connotations (e.g. repeat the word ‘Pen’ before seeing a ‘Sad’ face). Word repetition will be completed before participants are briefly exposed to face stimuli. Various researchers show that implicit decisions about the stimuli are made more quickly and accurately when they follow exposure to a label or characteristic sound that matches that stimulus. This facilitation is relative to conditions in which participants hear an irrelevant label or uncharacteristic sound. This cue-relevance effect occurs both when the label precedes a single, perceptually ambiguous target (e.g. Lupyan & Spivey, 2010; Chen & Spence, 2011; Lupyan & Ward, 2013) or a visual search array, in which the target is present (Schmidt & Zelinsky, 2009, 2011; Iordanescu, Guzman-Martinez, Grabowecky & Suzuki, 2008; Iordanescu, Grabowecky & Suzuki, 2011). Participants also typically show enhanced performance in cue-relevant vs. baseline/neutral trials, in which either a neutral noise (e.g. a beep, a segment of white noise, a ‘ready’ signal; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013) or no cue is presented (Lupyan & Spivey, 2010a; Chen & Spence, 2011).

These cue-relevance effects are also typically more robust when a moderately long inter-stimulus interval (approx. 1000ms-1500ms) separates label/sound exposure and presentation of the visual stimulus (Schneider, Engel & Debener, 2008; Chen & Spence, 2011; Lupyan & Thompson-Schill, 2012; Wilschut, Theeuwes & Olivers, 2014). Lengthened inter-stimulus intervals (hereafter referred to as ISIs) are also associated with worse performance in trials preceded by an irrelevant vs. baseline/neutral cue (e.g. Lupyan & Thompson-Schill, 2012; but see Lupyan & Spivey, 2010a; Chen & Spence, 2011). Lupyan and Ward (2013) reported similar effects when they increased the length of their continuous flash suppression display. This perhaps gave participants longer to ‘commit’ to an irrelevant-category label and the perceptual expectancies associated with it (i.e. expecting

the stimulus to possess category-diagnostic features of an irrelevant referent; Puri & Wojciuluk, 2008; Lupyan & Thompson-Schill, 2012).

The type of to-be-repeated word will also be manipulated, between-participants. Participants will either repeat a set of labels ('Sad', 'Surprise' and 'Pen/Printer') or verbs which, when emotion-related, have semantic connections both with the selected emotion labels and expressions presented ('Sob', 'Gasp', 'Bus/Bike'; see Chapter 7 for details of verb generation). It will be of interest to compare whether repetition of a relevant emotion label, and a relevant emotion verb (discretely associated with the emotional expression shown on the test face) would lead to similar patterns of perceptual facilitation. This comparison will have theoretical implications. If facilitation is only found after repetition of relevant emotion labels (relative to neutral/baseline and/or irrelevant word trials), then this may suggest that labels are special in the way that they create a set of online 'perceptual expectancies'. These predictions stem from Lupyan's (2007, 2012) label-feedback model; that simple exposure to a label may both accentuate the category-diagnostic features of a referent object and direct our attention toward them (see also Puri & Wojciuluk, 2008; Schmidt & Zelinsky, 2009). However, if relative facilitation is found following both repetition of relevant emotion labels and emotional verbs then this may suggest that both cues have the potential to activate the same conceptual knowledge base, useful for perception (Lupyan & Thompson-Schill, 2012).

Predictions concerning word type primarily stem from work conducted by Lupyan and Thompson-Schill (2012). Across a series of experiments participants either heard a category label or a semantically-related cue before being asked to make explicit or implicit decisions regarding a picture of that same object (e.g. deciding whether the picture presented a match for the cue, or selecting which of the label-relevant objects in a pair was inverted in orientation). Semantically-related cues included characteristic sounds made by the referent object (a 'barking' sound for a dog), verbal translations of those sounds ('arf-arf'), or verbs related to characteristic actions made by that referent (the word 'bark'). Lupyan and Thompson-Schill (2012) proposed that if all types of relevant cue (labels, verbs and sounds) led to facilitated decisions about a referent object then this would indicate that all cues mapped onto, and activated, the same conceptual knowledge base. However, two alternative interpretations were available if only some types of relevant cue elicited behavioural facilitation. The first interpretation is that cues activate the same forms of



conceptual knowledge, but across a different time course. The second interpretation is that different cues preferentially activate distinct forms of conceptual knowledge, unequally important for aiding the perception of referent objects (e.g. Casasanto & Lupyan, 2015).

Across experiments, Lupyan and Thompson-Schill (2012) found consistent facilitation after presentation of relevant category labels, but not other forms of semantically-related cue (e.g. characteristic sounds and verbs). Label facilitation was maintained after several manipulations of ISI length, between label offset and presentation of the visual stimulus (400ms; 1000ms; 1500ms). At longer ISIs (1000ms+), Lupyan and Thompson-Schill (2012) reasoned that even 'weaker' (non-label) cues might 'catch-up' with stronger cues, providing a degree of facilitation when relevant. Under these conditions there is both stronger potential for the participant to commit to a particular category, and for feedback from the category-level to influence visual processing (Lupyan, Thompson-Schill & Swingley, 2010; Schmidt & Zelinsky, 2011). Although some facilitation was reported when a long ISI was adopted, and relevant sound cues were compared to a no-cue baseline, label cues still led to a larger degree of facilitation when submitted to the same comparison. Therefore, Lupyan and Thompson-Schill (2012) concluded that different cues likely activate different types of conceptual knowledge related to referent objects, unequally useful for the decision required. In particular, relevant labels may activate a more 'perceptually precise' form of conceptual knowledge than that activated by other cues (e.g. Friedman et al., 2003; Schmidt & Zelinsky, 2009).

However, other manipulations of cue type have resulted in a different pattern of results. For example, when the referent object was very briefly presented, and a target present/absent decision required, Chen and Spence (2011) found consistent facilitation after presentation of characteristic sounds, but not relevant category labels. It is important to note that Chen and Spence (2011) manipulated ISIs differently to Lupyan and Thompson-Schill (2012), perhaps reducing the ability to compare findings across these experiments. Indeed, Chen and Spence (2011) only found cue-relevance effects for relevant labels when there was a 'long' ISI of 346ms between cue presentation and the target present/absent decision. In contrast, characteristic sounds led to facilitation across both a 'short' (173ms) and 'long' ISI (346ms, see also Schneider et al., 2008). Chen and Spence's (2011) work differs in another important way. Unlike Lupyan and Thompson-Schill (2012), Chen and Spence (2011) do not conceptualise different cue types as part of a 'hierarchy'. By a

hierarchical view, different cues would be 'ranked' according to their relative potential to activate the same conceptual knowledge base; labels may sit atop the hierarchy because they hold stronger associative connections to that knowledge base (e.g. Lupyan, 2007, 2012). Alternatively, Chen and Spence (2011) suggest that it is likely that words, sounds and pictures all map onto the same conceptual representation, but that the speed and strength of concept activation will depend on the possible translations required for different types of cue. In particular, Chen and Spence (2011) suggest that they found stronger cue-relevance effects for characteristic sounds because sounds, unlike words, did not require internal translation from their lexical form before they became useful and/or diagnostic (Glaser & Glaser, 1989; Roelofs, 2005; Wilschut et al., 2014). In other words, characteristic sounds have the potential to activate conceptual knowledge faster than category labels. As Chen and Spence (2011) used substantially shorter ISIs and object presentation durations (13ms) than other researchers (e.g. Lupyan & Thompson-Schill, 2012), it is perhaps unsurprising that they report stronger cue-relevance effects for characteristic sounds vs. relevant labels.

### **6.1.3 Specific predictions for the present work**

In sum, participants in the present task will repeat a single label or verb, out loud, before viewing a briefly presented, and immediately backwards masked, face. They will assess whether the face does or does not display an emotional expression. The face will display an emotional expression on half of the trials encountered. When an emotional expression is shown the repeated label or verb may (a) match, or be congruent with the expression shown (relevant trials), (b) mismatch, or be incongruent with the expression shown (irrelevant trials), or (c) have no emotional connotations (neutral trials). Following previous research, ISIs between word-cue offset and face presentation will also be manipulated. Longer ISIs are commonly used when researchers present cues that they consider to have weaker conceptual mappings (i.e. verbs); longer intervals may allow more time for conceptual activation, or category-commitment, which may concurrently influence visual processing of relevant stimuli (Lupyan et al., 2010; Schmidt & Zelinsky, 2011; Lupyan & Thompson-Schill, 2012). When using an ISI of 1000ms, researchers commonly find some facilitation following weaker cues, or equal facilitation after both strong and weak cues (e.g. Schneider et al., 2008; Schmidt & Zelinsky, 2011; Lupyan & Thompson-Schill, 2012). In contrast, strong cues may lead to behavioural facilitation at shorter ISIs, between 200-500ms (e.g. Chen & Spence, 2010; Salverda & Altmann, 2011; Lupyan & Thompson-Schill, 2012). Following the format of a similar previous study, two ISIs were chosen for the

present work; 400ms would represent a short ISI, and 1000ms a long ISI (Lupyan & Thompson-Schill, 2012).

It is predicted that participants will show facilitated performance after repetition of relevant labels (vs neutral and irrelevant labels). This may suggest that category labels hold a special ability to create 'perceptual expectancies'; accentuating and directing attention towards category-diagnostic features of a visual stimulus to facilitate perception (e.g. Lupyan, 2007, 2012; Puri & Wojciulik, 2008; Wilschut et al., 2014). For labels, the effect of word relevance should be present at both a long and short ISI (Lupyan & Thompson-Schill, 2012). The interaction between word relevance and ISI may have a second facet. Participants may show decrements in performance following repetition of irrelevant emotion labels, in comparison to neutral labels (baseline), but this difference may be more pronounced following a long vs. short ISI. However, the literature is less consistent on this point (e.g. Lupyan & Spivey, 2010a; Chen & Spence, 2011; Lupyan & Ward, 2013). Based on Lupyan and Thompson-Schill's (2012) initial predictions, participants in the verb condition may also show a cue-relevance effect (better performance following repetition of relevant verbs vs. irrelevant and neutral verbs). However, this effect should only emerge when verb repetition is followed by a long ISI, which would allow longer for these weak cues to activate the same conceptual knowledge base (e.g. Lupyan & Thompson-Schill, 2012; Wilschut et al., 2014). According to Lupyan and Thompson-Schill (2012) slower or weaker facilitation following verbs vs. labels may indicate that the two types of cue activate the same conceptual knowledge base, but across a different time course; stronger cues (labels) activate the concept faster, or more strongly, than weak cues (verbs). However, if only repetition of relevant labels, and not verbs, lead to relative facilitation then this may suggest that the two types of cue activate different kinds of conceptual knowledge, unequally useful for driving the perception of emotional expression.

## 6.2 Method

### 6.2.1 Participants

Seventy-Two participants completed this experiment (14 male) and no replacements were made. The sample had a mean age of 27.31 (SD = 11.03, Range: 18-62). Thirty-five participants were randomly assigned to the label condition ( $M_{\text{age}} = 25.29$ , SD = 7.71; 4 left-handed) and thirty-seven to the verb condition ( $M_{\text{age}} = 29.22$ , SD = 13.27; 4 left-handed). There was no significant difference in age across word type conditions,  $t(58.39) = 1.55$ ,  $p = .13$ . Collapsing across this between-subjects factor, 35 participants ( $M_{\text{age}} = 25.86$ , SD = 10.0) were assigned to complete the emotion word blocks first, followed by the neutral block, and 37 participants ( $M_{\text{age}} = 28.68$ , SD = 11.89) completed the neutral word block first, followed by the emotion blocks. Participant age did not differ significantly by block-order condition,  $t(70) = 1.09$ ,  $p = .28$ . Altogether, 48 participants had a pre-assessed threshold of 17ms, 20 a threshold of 32ms and four a threshold of 50ms. For subsequent analyses participants with a threshold of 32ms and 50ms were grouped together ( $n = 24$ ). Participants with the shorter threshold of 17ms were significantly younger ( $M_{\text{age}} = 22.77$ , SD = 4.27) than those with a threshold of 32/50ms ( $M_{\text{age}} = 36.38$ , SD = 14.47),  $t(25.02) = 4.51$ ,  $p < .001$ .

### 6.2.2 Materials

#### 6.2.2.1 Word stimuli

During experimental blocks, participants were required to repeat a single word aloud, per trial (an emotion label, emotion verb or neutral label). First, two emotion labels were selected that were representative of universal, emotional states; 'Sad' and 'Surprise' (e.g. Ekman, 1992, 2003; Ekman & Cordaro, 2011) In contrast to other negatively-valenced emotions (e.g. 'Anger', 'Fear' and 'Disgust'), expressions of 'Surprise' should not share any of the category-diagnostic features associated with expressions of 'Sad(ness)' (Smith et al., 2005; Schurgin et al., 2014; though see Calvo & Lundqvist, 2008). Therefore, it was unlikely that exposure to the irrelevant emotion label 'Surprise' would lead to any perceptual facilitation when participants encountered 'Sad' expressions. Second, emotion-related words, or verbs, were then chosen to match each emotion label. Verbs were selected from our own feature-generation data (Portch, Havelka, Brown & Giner-Sorolla, 2015; see Chapter 7 and Appendix E). As defined by Pavlenko (2008; see also Altarriba & Basnight-Brown, 2010), emotion-related words, or verbs, are those that describe an overt, emotional

behaviour but do not name the emotional state itself e.g. the word 'Cry' might describe the behaviour of someone feeling 'Sad'.

Ideal selection would rely on the following criteria: first, that the verb had been generated in response to either the 'Sad' or 'Surprise' emotion label by at least 3 of the 25 participants, representing a 'modal' exemplar. Second, that the verb was a non-overlapping exemplar, meaning that it must not have been generated in response to another emotion label by two or more participants. Verbs would be excluded from selection if they could be classed as a synonym for any of the modal exemplars given in response to any other emotion label. Synonyms were judged using an online thesaurus (<http://www.Thesaurus.com>). This would ensure that selected verbs were unambiguously associated with the discrete emotional states, denoted by the emotion labels 'Sad' and 'Surprise'<sup>13</sup>.

Ideal criteria were met when selecting an emotion-related verb to appropriately represent the emotional state of 'Sad(ness)'. Here the verb 'Sob' was chosen. However, it was not possible to meet both criteria when selecting a verb to represent the emotional state of 'Surprise'. The verb 'Gasp' (and related synonyms) were given by 13/25 participants in response to the label 'Surprise', meeting the first criteria. However, the verb 'Gasp' was also given by two participants in response to the label 'Fear', thus failing to meet the second criteria. Nevertheless, 'Gasp' was considered to present the best option available in the feature-generation data (see all modal exemplars, per emotion label, listed in Appendix E). Although the selected emotion-related verbs matched each other for number of syllables, the restrictions imposed during selection meant that it was not possible to match (a) emotion labels to emotion-related verbs, nor (b) emotion-related verbs to themselves, for number of letters. Written frequency scores, taken from SUBTLEX-UK (Van Heuven et al., 2014), suggested that emotion labels were significantly more frequent ( $M = 15464$ ,  $SD = 3582.20$ ) than emotion verbs ( $M = 314$ ,  $SD = 152.74$ ),  $t(2) = 5.98$ ,  $p = .027$ .

---

<sup>13</sup> It was particularly important to ensure that emotion-related verbs would strongly activate semantic knowledge discretely related to one emotional state, only. When using verbs to activate semantic knowledge, inconsistent cue-relevance effects have been associated with sub-standard selection procedures (the verb may not unambiguously activate the appropriate knowledge base e.g. Cummings et al., 2006; Niedenthal et al., 2009; Lupyan & Thompson-Schill, 2012).

Two sets of two neutral words were then chosen to represent a ‘baseline’; please note that all neutral words could be classified as ‘nouns’ rather than adhering to the ‘label’ and ‘verb’ distinction, imposed for emotion words. Neutral words matched their paired emotion words (labels or verbs, respectively) for number of syllables. Each set of neutral words also had a separate semantic theme: office objects for neutral ‘labels’ (‘Pen’ and ‘Printer’), and vehicles for neutral ‘verbs’ (‘Bus’ and ‘Bike’). Written frequency scores, taken from SUBTLEX-UK (van Heuven et al., 2014), suggested that neutral labels ( $M = 2612$ ,  $SD = 3057.53$ ) and neutral verbs ( $M = 4353.50$ ,  $SD = 4577.10$ ) did not differ significantly in frequency,  $t(2) = 0.45$ ,  $p = .70$ . Taken together, there was no significant difference in frequency for all labels (emotional and neutral;  $M = 9038$ ,  $SD = 7902.63$ ) and all verbs (emotional and neutral;  $M = 2333.75$ ,  $SD = 3235.66$ ),  $t(6) = 1.55$ ,  $p = .17$ .

#### 6.2.2.2 *Face stimuli*

Front-facing stimuli were selected from ‘Set A’ of the Karolinska Directed Emotional Faces database (hereafter referred to as the KDEF, Lundqvist, Flykt, Ohman, 1998)<sup>14</sup>. Only male identities were used; 20 faces were selected which displayed a ‘Sad’ expression and 20 which displayed a ‘Neutral’ expression (40 unique faces in total). Normative data were used to inform face selection (Goeleven, DeRaedt, Leyman & Verschuere, 2008). The top twenty most accurately categorised faces were chosen for each expression, based on Goeleven et al.’s. (2008) biased hit rate data. Replacements were made if the face had been mistakenly categorised as another of our key expressions by more than 25% of participants (e.g. if a ‘Sad’ expression had been mistakenly categorised as ‘Neutral’, if a ‘Neutral’ expression had been mistakenly categorised as ‘Sad’, or if either type of expression had been mistakenly categorised as ‘Surprised’). In total, four replacements were made, with the next most accurately categorised face being selected. All 20 ‘Sad’ expressions had been correctly

---

<sup>14</sup> The KDEF database (Lundqvist et al., 1998) offers two main advantages over the POFA (Ekman & Friesen, 1976), which was used in experiments 1-4 of this thesis. First, the KDEF includes 70 unique identities, allowing us to choose a wider stimulus set and reducing the likely negative effects (e.g. priming) when participants are consistently re-exposed to the same faces across trials. Second, each identity within the database been photographed twice, per emotional expression, allowing us to choose the best quality photograph, with the highest normative ratings, per identity.

categorised by 75+% of Goeleven et al's. (2008) participants. Thirteen 'Neutral' expressions had been correctly categorised by 75+% of participants, and seven by over 50% of participants. The mean correct and mistaken categorisation percentages for 'Neutral' and 'Sad' expressions are shown in Table 6.1. Repeated identities were unavoidable when using this method; both 'Sad' and 'Neutral' expressions were selected for 10 of the KDEF identities. This meant that 10 unique identities were selected from the KDEF which displayed either 'Sad' or 'Neutral' expressions, respectively.

We also selected the next four most accurately categorised 'Sad' and 'Neutral' expressions from 'Set A' of the KDEF. These faces would be shown in addition to the 40 core stimuli, but only during the preliminary, threshold trials. These additional selections enabled an increase in the number of trials, per presentation duration, without increasing the presentation of repeated stimuli (16 unique faces shown at 17ms, 32ms and 50ms, durations, respectively).

**Table 6.1 Mean correct and incorrect categorisation percentages for selected KDEF faces, as presented in Goeleven et al. 2008 (SD presented in parenthesis).**

	<b>Mean Correct Categorisation Rate (SD)</b>	<b>Mean mistaken categorisation rate: 'Sad' (SD)</b>	<b>Mean mistaken categorisation rate: 'Neutral' (SD)</b>	<b>Mean mistaken categorisation rate: 'Surprise' (SD)</b>
<b>Experimental Faces: 'Sad' (n=20)</b>	88.51% (7.07%)	-	2.23% (4.19)	4.19% (1.52)
<b>Experimental faces: 'Neutral' (n=20)</b>	78.07%(10.65%)	12.75% (3.90%)	-	0%
<b>Threshold faces: 'Sad' (n=4)</b>	71.39% (2.50%)	-	0%	0%
<b>Threshold faces: 'Neutral' (n=4)</b>	62.50% (6.25%)	25.00% (4.23%)	-	0%



All faces were formatted using the GNU Image Manipulation Platform ([www.gimp.org](http://www.gimp.org)). An elliptical cutting tool was used to select an oval-shaped region of the face that included internal features (eyebrows, eyes, nose and mouth) and excluded external features (the hair, external face shape and ears), as well as the neck and shoulder region. Similar methods have previously been used to ensure that participants extract emotion-relevant information from internal facial features, rather than other external cues (e.g. Calvo & Lundqvist, 2008; Sanchez & Vazquez, 2013). All faces were converted to greyscale and sized at 6cm x 8cm (306 x 397 pixels). Faces were superimposed and centrally positioned over the same random pattern square (e.g. Sweeny et al., 2009), sized at 11.46cm x 11.46cm (433 x 433 pixels). Face stimuli and background squares were both grey, but differed in contrast and luminance so that the central faces could be easily discriminated (e.g. Smith, Grabowecky & Suzuki, 2007).

### *6.2.2.3 Masking stimuli*

A similar method was used to create masking stimuli, which immediately followed presentation of the test face (e.g. a backwards masking procedure). Three previously unused male faces were selected from the KDEF, each of which displayed a neutral expression (see Esteves & Ohman, 1993 and Bornemann et al., 2012 for similar methods). Faces were re-sized to the same dimensions as the test faces, and superimposed over the same random pattern square. Using the GNU Image Manipulation platform, blur and smudge tools were used to obscure facial features and partially match luminance and contrast across areas that had and had not been occupied by features (see example in Figure 6.1). This resulted in a grey face shape that occupied the same space as test faces but lacked facial features, minimising the chance that masking stimuli would compete for attention or interfere with residual featural processing of the test face (e.g. Milders, Sahraie & Logan, 2008; Roesch et al., 2010; Pell & Richards, 2011; Bornemann et al., 2012). The three masks were randomly assigned across trials.

### *6.2.2.4 Apparatus*

Participants completed all parts of the experiment on a Toshiba Satellite Pro laptop with an Intel Core i5-3230M processor and a refresh rate of 100HZ. Stimuli were presented on a 15.6 inch LCD screen, using DMDX experimental software (version 5.1.1.4; Forster & Forster, 2003).

## 6.2.3 Design

### 6.2.3.1 Pre-experimental, threshold trials

First, participants encountered 48 trials in which they were required to make an emotionality decision about a briefly presented face (“does this face display an emotional expression?” ‘yes’ or ‘no’).

Three different face presentation durations were chosen; 17ms, 32ms and 50ms. Across participants, these durations were used to approximate threshold viewing conditions, defined as the minimum amount of time required for participants to visually assess that a face does or does not display an emotional expression, at or above chance levels of performance<sup>15</sup>. Participants encountered 16 trials at each duration interval, which were randomly intermixed. Threshold performance was indicated by accuracy rates between 50% and 75% (e.g. Watson & Pelli, 1983; Milders et al., 2008; Roesch et al., 2010). The presentation duration at which that particular participant was closest to threshold performance would be used for subsequent, experimental trials. The faster presentation duration was selected when participants responded with an accuracy rate at threshold for two possible durations. During threshold trials participants encountered the same 40 faces that they would encounter in later experimental blocks, as well as eight additional faces (four ‘Sad’ faces and four ‘Neutral’ faces) that were not later repeated in the experimental trials. Each face was presented once, only. Faces displaying ‘Sad’ and ‘Neutral’ expressions were equally sampled across trials at each presentation duration.

### 6.2.3.2 Experimental Trials

A 2 (Word Type: Label vs. Verb) × 3 (Word Relevance: Relevant vs. Irrelevant vs. Neutral) × 2 (ISI: 400ms vs. 1000ms) × 2 (Face Type: Emotional, Neutral) mixed design was employed. Word Type was a between-participants factor. A similar number of participants were randomly assigned to encounter either labels (n = 35; ‘Sad’, ‘Surprise’, ‘Pen’ and ‘Printer’) or verbs (n = 37; ‘Sob’, ‘Gasp’,

---

<sup>15</sup> In previous emotion detection work participants ordinarily exhibit perceptual thresholds between 30-50ms. However, some findings suggest that participants show a high of degree of variability in emotion detection, and can often perform the task very quickly, justifying inclusion of a shorter presentation duration (e.g. Martin, Berry, Dobranski, Horne & Dodgson, 1996; Maxwell & Davidson, 2004; Szczepanowski & Pessoa, 2007; Roesch et al., 2010).

'Bus' and 'Bike'). All participants completed two emotion word blocks and one neutral word block. Each block included 80 trials.

Block order was counterbalanced in two ways. First, roughly half of participants completed two emotion word blocks first, followed by the neutral block ('emotion first' participants;  $n = 35$ ); or one neutral block first, followed by two emotion word blocks ('neutral first' participants;  $n = 37$ ). Second, the order of the two emotion word blocks were counterbalanced, and the two possible orders were encountered equally by participants, independent of whether they completed emotion trial blocks first or second.

During the experiment participants only encountered one type of emotional expression ('Sad'). Independent of whether participants saw a 'Sad' or 'Neutral' expression, trials preceded by the words 'Sad' or 'Sob' always constituted 'relevant' trials (the word would always match the face if it displayed an emotional expression), and trials preceded by the words 'Surprise' or 'Gasp' always constituted 'irrelevant' trials (the word would always mismatch the face if it displayed an emotional expression). In addition, trials preceded by a neutral verb or label constituted a different type of trial ('Neutral'). Participants encountered an equal number of trials in which 'Sad' and 'Neutral' expressions were displayed, whether they followed repetition of relevant, irrelevant or neutral labels/verbs. Participants were always given the same amount of time to read and verbally repeat each word from the screen (1000ms). Following word offset, participants either encountered a short (400ms) or long (1000ms), unfilled ISI before face presentation. An equal number of trials with a short and long ISI were sampled.

The same 40 faces (20 displaying a 'Sad' expression and 20 displaying a 'Neutral' expression) were shown once in each of the six conditions: relevant word, short ISI; relevant word, long ISI; irrelevant word, short ISI; irrelevant word, long ISI; neutral word, short ISI and neutral word, long ISI. It was ensured that participants only ever saw the same face twice within the same block of trials. Trials from each condition were equally sampled during each block (four types of trial in the emotion word blocks and two types of trial in the neutral word block). Faces displaying 'Neutral' and 'Sad' expressions were equally sampled across trials in each condition. Within each block, participants encountered respective trial types in a random, intermixed order.

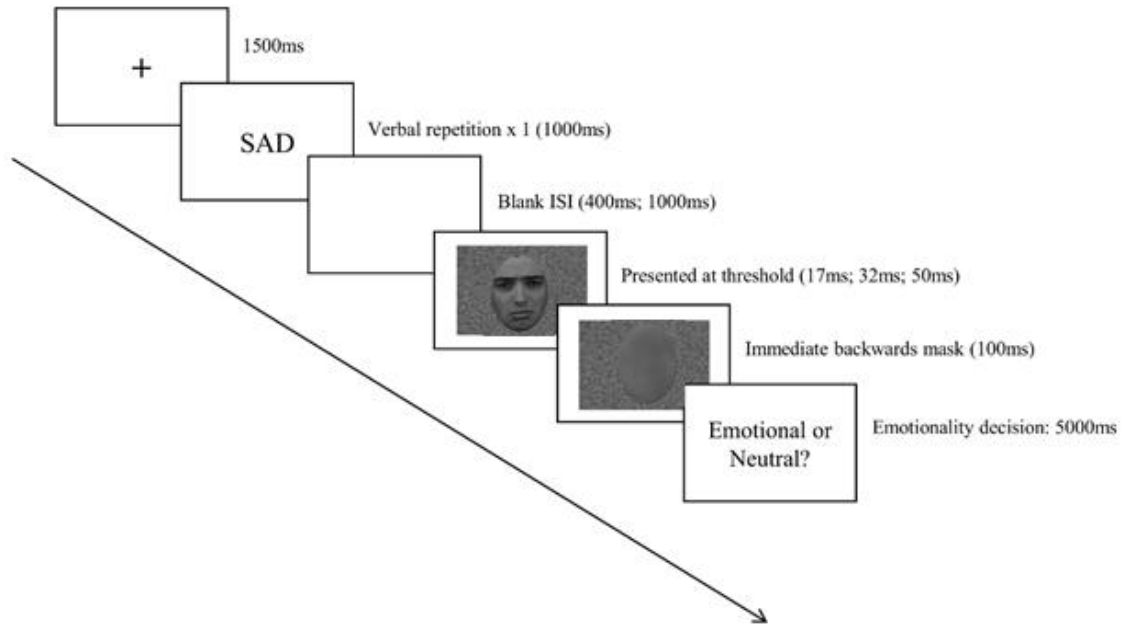
## 6.2.4 Procedure

### 6.2.4.1 *Pre-experimental trials*

Across 48 trials, participants were required to make an emotionality decision about a briefly presented face (“does this face display an emotional expression?” ‘yes’ or ‘no’). At the beginning of each trial participants saw a fixation cross for 1500ms, followed by a blank screen for 100ms. A face was then briefly presented for 17ms, 32ms or 50ms, superimposed over a random pattern square. The mask stimuli were presented immediately afterward and remained visible for 100ms. Following mask presentation, participants were given five seconds to make their emotionality decision before the next trial automatically began. They were encouraged to make their decisions as quickly and as accurately as possible. Participants pressed the left shift key to respond ‘yes’ (indicating that an emotional face had been shown) or the right shift key to respond ‘no’ (indicating that a neutral face had been shown). Participants were not required to repeat words during these trials. Accuracy rates were recorded, per presentation duration.

### 6.2.4.2 *Experimental trials*

Participants completed 240 trials in total, split across three blocks (two emotional and one neutral). Participants were permitted to take a short break after each trial block (80 trials). Trial structure followed a similar format to that used in the pre-experimental block (see Figure 6.1). Each trial began with the presentation of a central fixation cross for 1500ms. A single word was presented next, in black font (size 36) in the middle of a white screen. This could be an emotion verb (e.g. ‘Sob’), emotion label (e.g. ‘Cry’) or a neutral label/verb (e.g. ‘Pen’). Participants were instructed to read the word aloud, once, as soon as it appeared and were given 1000ms to do so. A blank slide was presented for a short (400ms) or long (1000ms) duration before a face was briefly presented; always at the individual’s pre-determined threshold (17ms, 32ms or 50ms). Faces were immediately backwards masked for 100ms. Per trial participants were asked to make the same emotionality decision that they had for the threshold trials, and key mapping remained the same. Again, they were encouraged to make their decisions as quickly and as accurately as possible and five seconds were permitted for responses, per trial (see full participant instructions, presented in Appendix D). The next trial either began immediately after a response had been submitted, or after five seconds, in the event that the participant did not provide a response. Participant accuracy and reaction time were recorded. Timer onset coincided with mask presentation (100ms + time to make a response). For subsequent analyses, then, mask duration was subtracted from reaction times.



**Figure 6.1** The temporal sequence for a single experimental trial.

*This figure uses identity AM01 from 'Set A' of the KDEF (Lundqvist et al., 1998). In this trial participants are required to respond 'Yes', to indicate that the face displays an emotional expression.*

## 6.3 Results

### 6.3.1 Data preparation

Five dependent variables were calculated per condition, for each participant. Four variables were related to accuracy. Hit rates reflect when participants had correctly identified an 'emotional face' as being 'emotional' and were expressed as a proportion, using the following calculation:  $[\text{hits}/(\text{hits} + \text{misses})]$ . False alarms reflect when participants incorrectly responded that a 'neutral face' was 'emotional' and were expressed as a proportion, using the following calculation  $[\text{false alarms}/(\text{false alarms} + \text{correct rejections})]$ . Sensitivity rates were expressed using  $d'$ , following the conventional calculation (Macmillan, 1993):  $z(\text{Hits}) - z(\text{False Alarms})$ . Response bias was represented using  $c$  (criterion) and calculated using the following formulae (Macmillan, 1993):  $-0.5 * z(\text{Hits}) - z(\text{False Alarms})$ . Following Macmillan and Creelman (1991), transformations were applied to hit and false

alarm rates of 0 and 1. For hit rates, 0.5 was added to the number of hits before being divided by the sum of hits and misses. For false alarms, the sum of false alarms and correct rejections were divided by 0.5.

Reaction times comprised a fifth dependent variable. When calculating average reaction times we considered only accurate responses made to emotional faces (responding 'yes' when an emotional face had been shown). Before calculations, mask duration (100ms) was subtracted from each response time and any trials recorded below 300ms or above 3000ms were removed (e.g. Balota & Black, 1997). This resulted in the removal of less than 1% of total responses.

For each of the six conditions, Z-scores were also used to identify and remove accuracy and reaction time totals that were 3 standard deviations above or below the sample mean. This technique resulted in the following data exclusions: For the hit rate analysis, one condition total was removed for two participants each, and two condition totals were removed for one further participant. For the false alarm analysis one condition total was removed for two participants, each. For the  $d'$  analysis, one condition total was removed for two participants, each. For the reaction time analysis one participant was completely omitted from further analysis due to having reaction times 3 standard deviations faster than the sample mean for four of the six conditions. Two conditions totals were removed for two further participants and one condition total was removed for a final two participants, each.

### **6.3.2 Accuracy Analysis (Hit rates, False alarms, $d'$ and $c$ )**

We report the results of separate 2 (Word Type: Label vs. Verb)  $\times$  3 (Word Relevance: Relevant vs. Irrelevant vs. Neutral)  $\times$  2 (ISI length: Short vs. Long) mixed factorial ANOVAs, conducted for each of the four dependent variables. Word Type was a between-participants factor. An additional between-participant factor was included to increase the statistical power of the analysis: block order (emotion-block-first vs. neutral-block-first). There was a significant interaction between block order and word relevance for hits, false alarms and response bias. However, each interaction appeared to be driven by performance in one condition – the neutral word block, when completed by neutral-block-first participants. In this trial block participants exhibited more hits and more false alarms when compared to other block order combinations (e.g. the neutral block when encountered by emotion-block-first participants), though these differences were not significant. Together, these trends suggest that participants adopted a more liberal response bias when responding to neutral word trials when this type of trial was encountered in the first vs. second trial block. However, these effects may not be driven by the type/relevance of word that participants encountered (e.g. emotion

vs. neutral) but by the difference in performance across the first and last trial blocks that participants completed. It was not possible to test this proposal for participants who first encountered emotion word trials, as there were two emotion word blocks, which were counterbalanced. Critically though, the interpretation of results does not change when block order was omitted from the analysis, therefore, no further effects involving block order are reported.

Participant threshold was excluded as a factor from the main analysis; there was no reason to expect differences in performance as a function of threshold as pre-experimental procedures were used to ensure that participants completed the task at a level equated to their own individual abilities (a presentation duration that resulted in accuracy rates between 50-75% e.g. Roesch et al., 2010). These allocation methods also meant that there were an uneven number of participant assigned to each threshold condition (48 participants in the 17ms condition and 24 participants in the 32ms condition).

Mean totals for each of the four accuracy variables are shown in Table 6.2, per condition.

**Table 6.2 Mean hit rates, false alarms,  $d'$  and  $c$  (criterion) measures as a function of Word Type, Word Relevance and ISI length ( $N \leq 72$ , SEM presented in parenthesis).**

Word Type	ISI Length	Word Relevance	Hit Rates (proportion)	False Alarms (proportion)	$d'$ (sensitivity)	$c$ (criterion)
Label (N=35)	Short	Relevant	0.73 (0.017)	0.29 (0.027)	1.25 (0.11)	-0.029 (0.051)
		Irrelevant	0.69 (0.027)	0.24 (0.024)	1.32 (0.12)	0.13 (0.061)
		Neutral	0.71 (0.025)	0.26 (0.023)	1.29 (0.11)	0.048 (0.071)
	Long	Relevant	0.73 (0.025)	0.25 (0.022)	1.39 (0.11)	0.046 (0.068)
		Irrelevant	0.71 (0.020)	0.20 (0.019)	1.49 (0.093)	0.18 (0.063)
		Neutral	0.73 (0.025)	0.26 (0.027)	1.34 (0.13)	0.0033 (0.063)
Verb (N=37)	Short	Relevant	0.69 (0.026)	0.22 (0.021)	1.33 (0.12)	0.15 (0.062)
		Irrelevant	0.67 (0.028)	0.23 (0.025)	1.26 (0.13)	0.19 (0.064)
		Neutral	0.70 (0.026)	0.25 (0.028)	1.25 (0.12)	0.10 (0.077)
	Long	Relevant	0.73 (0.026)	0.21 (0.021)	1.49 (0.12)	0.12 (0.074)
		Irrelevant	0.71 (0.026)	0.22 (0.020)	1.36 (0.12)	0.15 (0.062)
		Neutral	0.72 (0.026)	0.23 (0.025)	1.36 (0.12)	0.15 (0.081)

Results of the 2 (Word Type: Label vs. Verb; between)  $\times$  3 (Word Relevance: Relevant vs. Irrelevant vs. Neutral; within)  $\times$  2 (ISI length: Short vs. Long; within) ANOVA are shown in Table 6.3, below. Results are parsed according to dependent variable. Effects are emboldened and later discussed if they meet the criterion for significance ( $p < .05$ ) or can be considered marginal ( $p \leq .06$  and  $> .05$ ).



**Table 6.3 Results of the Word Type × Word Relevance × ISI length mixed factorial ANOVA, for hit rates, false alarms,  $d'$  and  $c$  (criteria).***Significant ( $p < .05$ ) and marginal effects ( $p \leq .06$  and  $> .05$ ) are presented in bold*

Dependent Variable	Effect	DF	MSE	F	p	$\eta_p^2$
Hit Rates	Word Type	(1, 65)	0.088	0.21	.65	0.003
	Word Relevance	(1.74, 113.16)	0.014	2.00	.15	0.030
	<b>ISI</b>	<b>(1, 65)</b>	<b>0.008</b>	<b>4.19</b>	<b>.045</b>	<b>0.060</b>
	Word Relevance × Word Type	(1.74, 113.16)	0.014	0.12	.86	0.002
	Word Relevance × ISI	(2, 130)	0.008	0.35	.71	0.003
	Word Type × ISI	(1, 65)	0.008	0.87	.36	0.013
	Word Relevance × Word Type × ISI	(2, 130)	0.008	0.49	.62	0.007
False Alarms	Word Type	(1, 66)	0.078	0.84	.36	0.013
	Word Relevance	(1.77, 116.74)	0.011	2.45	.097	0.036
	<b>ISI</b>	<b>(1, 66)</b>	<b>0.007</b>	<b>8.37</b>	<b>.005</b>	<b>0.11</b>
	Word Relevance × Word Type	(1.77, 116.74)	0.011	2.93	.064	0.043
	Word Relevance × ISI	(1.74, 114.56)	0.009	0.55	.56	0.008
	Word Type × ISI	(1, 66)	0.007	0.18	.67	0.003
	Word Relevance × Word Type × ISI	(1.74, 114.56)	0.009	1.22	.29	0.018
$d'$	Word Type	(1, 66)	1.90	0.012	.91	<0.001
	Word Relevance	(2, 132)	0.24	0.45	.64	0.007
	<b>ISI</b>	<b>(1, 66)</b>	<b>0.19</b>	<b>7.88</b>	<b>.007</b>	<b>0.11</b>
	Word Relevance × Word Type	(2, 132)	0.24	1.24	.29	0.018
	Word Relevance × ISI	(2, 132)	0.18	0.26	.77	0.004
	Word Type × ISI	(1, 66)	0.19	0.001	.98	<0.001
	Word Relevance × Word Type × ISI	(2, 132)	0.18	0.25	.78	0.004
$c$	Word Type	(1, 68)	0.62	1.02	.32	0.015
	<b>Word Relevance</b>	<b>(1.60, 108.69)</b>	<b>0.10</b>	<b>3.98</b>	<b>.030</b>	<b>0.055</b>
	ISI	(1, 68)	0.057	0.042	.84	0.001
	Word Relevance × Word Type	(1.60, 108.69)	0.10	1.21	.30	0.017
	Word Relevance × ISI	(1.66, 113.15)	0.066	0.010	.98	<0.001
	Word Type × ISI	(1, 68)	0.057	0.27	.61	0.004
	Word Relevance × Word Type × ISI	(1.66, 113.15)	0.066	1.66	.20	0.024

\*Greenhouse-Geisser corrections were applied where sphericity assumptions were violated: The 'Word Relevance' effect for hit rates, false alarms and  $c$ (criteria), and the 'Word Relevance × ISI' effects for false alarms and  $c$ (criteria).

### 6.3.2.1 Hit Rates

There was a significant main effect of ISI length for hit rates. Participants were better to identify an emotional face as ‘emotional’ following a long ISI ( $M = 0.72$ ,  $SEM = 0.018$ ) than a short ISI ( $M = 0.70$ ,  $SEM = 0.018$ ).

### 6.3.2.2 False alarms

There was a significant main effect of ISI length for false alarms. Participants were significantly less likely to categorise a neutral expression as ‘emotional’ following a long ISI ( $M = 0.23$ ,  $SEM = 0.015$ ), than a short ISI ( $M = 0.25$ ,  $SEM = 0.018$ ).

### 6.3.2.3 Sensitivity ( $d'$ )

There was a main effect of ISI length. Participants showed significantly higher sensitivity following a long ISI ( $M = 1.40$ ,  $SEM = 0.081$ ) than a short ISI ( $M = 1.28$ ,  $SEM = 0.082$ ).

### 6.3.2.4 Response Bias ( $c$ )

There was a significant main effect of Word Relevance for response bias rates. Pairwise comparisons revealed that participants were more likely to adopt a ‘conservative’ criterion (classifying both neutral and emotional faces as neutral in expression) following repetition of irrelevant emotion words ( $M = 0.16$ ,  $SEM = 0.041$ ) vs. relevant emotion words ( $M = 0.080$ ,  $SEM = 0.042$ ;  $p = .003$ ), but not neutral words ( $M = 0.079$ ,  $SEM = 0.043$ ;  $p = .11$ ). Participants adopted a similar level of conservative bias following repetition of relevant and neutral words ( $p = 1.00$ ).

## 6.3.3 Reaction time analysis

A 2 (Word Type: Label vs. Verb; between)  $\times$  3 (Word Relevance: Relevant vs. Irrelevant vs. Neutral; within)  $\times$  2 (ISI length: Short vs. Long; within) mixed factorial ANOVA was also conducted on mean reaction times, for trials in which participants had correctly responded that an emotional face had been shown. Mean correct reaction times, per condition, are shown in Table 6.4.

**Table 6.4 Mean reaction times as a function of Word Type, Word Relevance and ISI length (SEM presented in parenthesis).**

Word Type	ISI Length	Relevant	Irrelevant	Neutral
<b>Label (33)</b>	<b>Short</b>	553.63 (24.45)	591.32 (27.55)	547.13 (27.66)
	<b>Long</b>	567.04 (26.83)	590.56 (30.42)	558.14 (27.61)
<b>Verb (34)</b>	<b>Short</b>	647.92 (31.20)	660.03 (28.88)	668.92 (32.41)
	<b>Long</b>	643.53 (28.28)	635.28 (25.04)	660.44 (27.97)

In Table 6.5 below all effects from the mixed factorial ANOVA are presented. Following treatment of the accuracy data, effects are further explored when significant ( $p < .05$ ) or marginal ( $p \leq .06$  and  $> .05$ ).

**Table 6.5 Mean correct reaction times as a function of Word Type, Word Relevance and ISI length.**

*Significant ( $p < .05$ ) and marginal effects ( $p \leq .06$  and  $> .05$ ) are emboldened.*

Effect	DF	MSE	F	p	$\eta_p^2$
<b>Word Type</b>	<b>(1, 66)</b>	<b>135736.55</b>	<b>5.44</b>	<b>.023</b>	<b>0.076</b>
Word Relevance	(1.64, 108.26)	8560.04	1.33	.27	0.020
ISI	(1, 66)	5495.22	0.091	.76	0.001
<b>Relevance × Word Type</b>	<b>(1.64, 108.26)</b>	<b>8560.04</b>	<b>3.71</b>	<b>.036</b>	<b>0.053</b>
ISI × Word Type	(1, 66)	5495.22	1.98	.16	0.029
Relevance × ISI	(2, 132)	4324.49	0.67	.52	0.010
Relevance × ISI × Word Type	(2, 132)	4324.49	0.016	.98	<0.001

*\*Greenhouse-Geisser corrections were applied for the 'Word Relevance' effect*

There was a significant main effect of Word Type. Participants were faster to correctly categorise a face following repetition of labels ( $M = 568.48$ ,  $SEM = 27.42$ ) than verbs ( $M = 652.39$ ,  $SEM = 28.96$ ).

This effect was qualified by an interaction with Word Relevance. Marginal means are displayed in Table 6.6, below.

**Table 6.6 Mean, correct reactions times by Word Type and Word Relevance (SEM presented in parenthesis).**

	Relevant Word	Irrelevant Word	Neutral Word
<b>Label (n = 34)</b>	574.61 (21.91)	597.95 (24.20)	570.47 (22.11)
<b>Verb (n = 36)</b>	633.39 (25.17)	637.52 (23.66)	650.24 (25.13)

There was a significant main effect of Word Relevance for participants in the Label condition,  $F(2, 64) = 3.04$ ,  $p = .050$ ,  $MSE = 2384.79$ ,  $\eta_p^2 = 0.087$ , but not for those in the Verb condition,  $F(1.54, 50.77) = 0.68$ ,  $p = .48$ ,  $MSE = 5030.92$ ,  $\eta_p^2 = 0.020$ .

Pairwise comparisons revealed that participants in the Label condition were significantly slower to correctly identify that an emotional face was ‘emotional’ following repetition of an irrelevant label vs. a neutral label ( $p = .047$ ), but not after an irrelevant vs. relevant label ( $p = .13$ ). In addition, participants made correct decisions at a similar speed when responding after repetition of relevant and neutral labels ( $p = 1.00$ ).

Further, independent samples t-tests showed that participants in the Label and Verb conditions made correct decisions at a similar speed following repetition of relevant words [ $t(65) = 1.72$ ,  $p = .091$ ] and irrelevant words [ $t(65) = 1.14$ ,  $p = .26$ ]. However, participants were significantly slower to make correct decisions following repetition of neutral verbs in comparison to neutral labels, [ $t(65) = 2.32$ ,  $p = .023$ ,  $d = 0.58$ .]

## 6.4 Discussion

### 6.4.1 Summary of key predictions and findings

The present work examined whether language *directly* shapes our perception of emotional expression. It was predicted that participants would be faster and/or more accurate to judge that a face showed an emotional expression when preceded by repetition of a word ‘relevant’ to that expression (e.g. repeat the label ‘Sad’ prior to seeing a ‘Sad’ expression; Lupyan & Spivey, 2010; Chen & Spence, 2011; Schmidt & Zelinsky, 2011; Iordanescu et al., 2011; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013). Performance would be facilitated in comparison to trials in which participants repeated a word that was ‘irrelevant’ to the emotional expression shown (e.g. ‘Surprise’), or was neutral in

connotation (e.g. 'Pen'). Two further design aspects were manipulated: the type of word that was repeated and the ISI between word repetition and face presentation. It was predicted that the effect of word relevance would either be (a) restricted to conditions in which participants repeated category labels (e.g. 'Sad') vs. verbs (e.g. 'Sob'), or (b) would be found to proceed along a different time course for the two types of word; at both a short and long ISI for labels, and at a long ISI, only, for verbs (e.g. Lupyan & Thompson-Schill, 2012).

In contrast to predictions, there were no indications in the pertinent accuracy data (hit rates or sensitivity) to suggest that participants were better to make an emotionality decision following repetition of a relevant (vs. irrelevant or neutral) word. There were also no interactions observed between word relevance and word type, suggesting that these trends were common to the repetition of both labels and verbs. Further, neither of the main variables of interest (word relevance and word type) interacted with levels of the ISI manipulation for either dependent variable, suggesting that neither relevant cue increased in usefulness when a longer time was provided for conceptual activation, or category commitment, to occur (Lupyan et al., 2010; Schmidt & Zelinsky, 2011; Lupyan & Thompson-Schill, 2012). Taken together, these findings may suggest that language does not 'directly' influence the perception of emotional expressions.

There was a main effect of word relevance found for response bias rates. Post-hoc comparisons suggested that participants were more likely to categorise all faces as 'neutral' in expression (a conservative bias) following the repetition of irrelevant cues (both labels and verbs), than relevant cues. However, it is important to note that participants showed a similar direction and degree of response bias following repetition of relevant and neutral cues, and that participants showed bias in a conservative direction following the repetition of all cue types.

Although there were no interactions observed when ISI length was included as a factor, a significant main effect was shown across three analyses. Participants gave significantly more hits and significantly fewer false alarms after a long vs. short ISI. These trends were reflected in the sensitivity data; participants were more sensitive to the emotionality of a face after a long vs. short ISI.

For the reaction time data there was a significant main effect of word relevance and an interaction between word relevance and word type. In contrast to predictions, participants were no faster to correctly infer that an emotional expression had been shown following repetition of a relevant (vs. irrelevant or neutral) label. Participants showed similar reaction times following repetition of neutral and relevant labels. Although participants were slowest to correctly respond following repetition of an irrelevant label, the difference was only significant when compared to neutral trial performance. In addition, participants only showed a tendency to respond more slowly after the repetition of verbs vs. labels when they repeated neutral words. The reaction time findings support the notion that language may not directly influence emotion perception.

#### **6.4.2 Methodological limitations: cue type and presentation duration**

Several methodological limitations may have reduced the opportunity to find cue-relevance effects. Two temporal considerations are important when attempting to use language cues to activate conceptual knowledge. First, there must be a sufficient ISI between cue and target presentation for conceptual activation, or 'category-commitment' to develop (e.g. Lupyan et al., 2010; Lupyan & Thompson-Schill, 2012). Both the short and long ISIs chosen were consistent with those used in previous work, where cue-relevant facilitation has been found for 'strong' cues and, to a lesser extent, 'weak' cues (e.g. Schneider et al., 2008; Chen & Spence, 2010; Salverda & Altmann, 2011; Schmidt & Zelinsky, 2011; Lupyan & Thompson-Schill, 2012). Second, the cue itself must be presented for a sufficient duration to initiate the process of conceptual activation (Wolfe, Horowitz, Hyle & Vasan, 2004; Schmidt & Zelinsky, 2011; Lupyan & Thompson-Schill, 2012). In the present task each cue word was presented on the screen for 1000ms, during which time the participant would repeat the word aloud. Pilot testing suggested that this was ample time for the participant to finish reading the word before the blank ISI began. Various researchers suggest that cue-relevance effects emerge when cues are presented for shorter durations than 1000ms (e.g. Friedman, Cywocicz & Dziobek, 2003). In particular, Mahr and Wentura (2014) continued to find facilitation in a cross-modal stroop task when spoken colour word primes were compressed from 400ms to 40ms, but the ISI maintained at 100ms. Mahr and Wentura (2014) suggested that these effects may be constrained to tasks that use a small number of cues and corresponding stimuli, as repetition across trials may build a strong perceptual template for responding to the stimulus set, reducing the need to attend to a particular

cue. However, this argument could be applied to the present work, as each participant would only ever encounter four types of cue word and two types of facial expression.

Arguably, cue duration may only be problematic when combined with an atypical method of cue presentation, as it was in the present work. In contrast to past research, participants were required to read the cue-word aloud, rather than hearing a recording of the word being spoken (e.g. Chen & Spence, 2010, 2011; Lupyan & Thompson-Schill, 2012; Iordanescu et al., 2008, 2011; though see Lupyan & Swingley, 2011<sup>16</sup>). This may provide an explanation for the accuracy findings concerning ISI; participants showed heightened sensitivity (more hits and fewer false alarms) following a long vs. short ISI. Perhaps then, verbal repetition created an additional demand that transiently disrupted the participant's ability to prepare to make the emotionality decision. Critically, the effects of ISI did not interact with word relevance, nor word type. Based on arguments introduced in work comparing sound vs. lexical/label cues, it is possible that aurally-presented cues lead to faster conceptual activation than label cues, which may require further translation (e.g. Chen & Spence, 2011; Iordanescu et al., 2011). In future work then, it may be useful to present spoken word cues (relevant, irrelevant and neutral) to ensure that conceptual activation is appropriately initiated, and has time to unfold prior to the presentation of visual stimuli. Alternatively, Iordanescu et al. (2008, 2010, 2011) ensured that cues had been sufficiently processed using a dual-cueing procedure. Prior to the presentation of a target absent-or-present visual array, participants were both aurally instructed about target identity before each trial, and also heard a cue-related characteristic sound prior to, or simultaneous, with the visual array.

---

<sup>16</sup> Lupyan and Swingley (2011) found facilitation when 'self-directed' speech was required in a visual search task. Participants performed best when they overtly vs. covertly read a (relevant) target label before selecting that target from a visual array. In contrast to the present work, Lupyan and Swingley (2011) inserted a relatively long ISI, of 2200ms, between label and target presentation. Under these circumstances there may be greater potential for translation and conceptual activation, despite atypical presentation of label cues (overtly repeated by the participant).

### 6.4.3 Methodological limitations: face stimuli

The present task also made use of a restricted set of emotional stimuli; male faces that displayed a sad expression. Approaches which suggest that language has a *direct* influence on perception emphasise that labels create perceptual expectancies for a referent object, accentuating category-diagnostic features in the display, and drawing our attention toward them (e.g. Lupyan 2007, 2012). Indeed, cue-relevance effects are commonly stronger for more ‘typical’ or distinctive objects that clearly exhibit these features (e.g. Lupyan, 2008a; Lupyan & Spivey, 2010b; Lupyan & Thompson-Schill, 2012). Some researchers suggest that all six of the ‘basic’ emotional expressions possess a non-overlapping set of category-diagnostic features, meaning that any expression would be suitable for the present work (e.g. Smith et al., 2005; Schurgin et al., 2014). However, the literature presents an unbalanced picture; facial expressions of happiness and, to a lesser extent, anger, are considered to be the most easily identifiable, which may be linked to their possession of stronger, or a larger number of, category-diagnostic features (e.g. Esteves & Ohman, 1993; Calvo & Lundqvist, 2008; Calvo & Nummenmaa, 2008; Gosselin & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002). In contrast, expressions of sadness sometimes receive lower categorisation rates than other emotional expressions, perhaps because they are commonly mistakenly categorised as neutral in expression (e.g. Calvo & Lundqvist, 2008; Goeleven et al., 2008; Sánchez & Vázquez, 2013). Although care was taken to select ‘prototypical’ stimuli from a validated database (Lundqvist et al., 1998), sad expressions may not possess appropriately distinctive ‘category-diagnostic’ features to detect the direct effects of language in emotion perception.

However, the choice to use only male, sad expressions was empirically motivated. In previous pilot work participants were presented with sad, fearful and angry expressions, displayed by both female and male actors within the KDEF (Lundqvist et al., 1998). Similar to the present work, participants repeated relevant, irrelevant or neutral words (labels or verbs) before viewing faces at threshold, but then were required to make an implicit gender decision about each face (see Roesch et al., 2010). Again, it was predicted that participants would show facilitation in their implicit gender decisions when they had repeated a word that matched/was congruent with the emotional expression displayed; if relevant words actively support perception of emotional expression they may also support parallel processing of other facial attributes (e.g. Roesch et al., 2010; Gendron et al., 2012). However, findings for fearful and angry expressions robustly supported stereotypical



gender/expression conjunctions (e.g. Hess, Blairy & Kleck, 2000; Hess, Adams & Kleck, 2004; Hess, Adams, Grammer & Kleck, 2009). Specifically, participants were faster to correctly identify faces as female (vs. male) when they displayed a fearful expression, and to correctly identify faces as male (vs. female) when they displayed an angry expression. Stereotypical conjunction effects have been associated with perceptual overlaps in facial attributes; for example, angry expressions typically accentuate the features of a face that we associate with masculinity (e.g. lowered brows and squared jawlines; Hess et al., 2000). Importantly, these effects were present independent of the type, or relevance, of the word repeated prior to face presentation, perhaps suggesting that conjunctions are processed automatically, even in the presence of additional conceptual knowledge. Despite the present task using an emotionality, rather than gender decision, it was still possible that stereotypical conjunctions would impact upon the pattern of results. For example, when 'Anger' represents the relevant label, participants might show heightened accuracy when judging congruent male faces and lowered accuracy when judging female faces because anger is an emotion respectively consistent and inconsistent with the actor's gender.

Therefore, it was a pragmatic decision to use only one gender and expression conjunction. Sad expressions were chosen because the pilot work showed no clear evidence of stereotypical gender bias in participants' responses to these stimuli. This is perhaps because there do not appear to be featural overlaps in the few category-diagnostic features associated with sad expressions and typical female or male facial structures (e.g. Fabes & Martin, 1991; Hess et al., 2000, 2009). This assumption is supported by a study that paired androgynous, emotional faces with apparent male or female hairstyles (Hess et al., 2004, 2009). Here participants gave higher anger ratings to apparent male faces, higher fear ratings to apparent female faces, but similar 'sad' and 'neutral' ratings to both male and female faces, particularly when displayed at low emotional intensities.

#### **6.4.4 Theoretical implications**

The present findings also have theoretical relevance. Only some researchers claim that language labels have a 'direct' influence on visual perception (e.g. Lupyan, 2007, 2012). Others argue that language labels, and linked conceptual knowledge, simply constrain our interpretation of the low-level visual information, presented in an expression (e.g. Barrett et al., 2007). In other words, emotion perception is an inferential rather than diagnostic process, which benefits from a language-based boost (e.g. Lindquist et al., 2006; Gendron

et al., 2012). A similar study, conducted by Nook, Lindquist & Zaki (2015), may provide support for this view. In their task participants viewed a briefly presented emotional face, either before seeing an emotion label (e.g. 'anger') or a second emotional face. For each pair, participants were required to decide whether the two stimuli presented an emotional match or mismatch. Participants were faster and more accurate to judge face-label pairs, than face-face pairs, particularly when the face and label were congruent with one another. This suggests that emotion labels, and linked conceptual knowledge, help to constrain our interpretation of previously viewed emotional stimuli by clearly communicating information consistent with a particular emotional concept (e.g. Carroll & Russell, 1996; Widen & Russell, 2003). In this sense, labels may be more informative for interpretation when they follow, rather than precede, the visual stimulus. Comparing Nook et al.'s. (2015) findings to the present work suggests that the exact temporal placement of word and face stimuli are important. Chen and Spence (2010) investigated this relationship in a target absent/present task using objects; object labels or characteristic sounds were aurally presented either before, simultaneous with, or after the object (or mask) was displayed (see also Spence, Shore & Klein, 2001; Roelofs, 2005 for use of similar manipulations). Participants were best to correctly judge that an object was present when the relevant label was presented 300ms after target offset. Similar to Nook et al. (2015), Chen and Spence (2010) suggested that cues may facilitate post-perceptual processes, or 'memory' for the category-diagnostic features shown.

There may be indications in the present data that participants used cue words 'heuristically', or strategically, to inform their decisions. Response bias rates indicated that participants were more likely to suggest that all faces displayed a 'neutral' expression (a conservative bias) following the repetition of irrelevant cues (both labels and verbs), than relevant cues, but not neutral cues. Given that participants only encountered one type of emotional expression in the present task ('sadness') they may have begun to use repeated labels as a form of heuristic (rather than veridical) cue. Arguably, surprised expressions share few category-diagnostic features with expressions of sadness (e.g. Smith et al., 2005; Schurgin et al., 2014). As such, participants may learn that, when an emotional face is shown, it does not display an expression of surprise. This may motivate participants to respond 'no' following the repetition of the words 'Surprise' and 'Gasp', independent of the type of expression shown (neutral or emotional). In the present work then, irrelevant labels may be used strategically to influence emotionality decisions (see further arguments

concerning strategic uses of language in section 8.3 of the general discussion e.g. Pylyshyn, 1999).

#### 6.4.5 Conclusions and future work

In conclusion, the present findings suggest that emotion-related language does not *directly* influence the perception of emotional expression. Participants were no better to judge the emotionality of a face when presentation followed exposure to an emotion word that was relevant (vs. irrelevant or unrelated) to the expression displayed. Cue-relevance effects were absent both when participants encountered relevant emotion labels (e.g. ‘Sad’) and verbs (e.g. ‘Sob’). Rather than *directly* influencing the perception of emotional expressions, language may simply ‘constrain’ the visual information we receive and/or aid an inferential decision process (e.g. Barrett et al., 2007). This may suggest that language differently influences the perception of objects and emotional expressions (e.g. Lupyan, 2007; Lupyan & Thompson-Schill, 2012).

However, several methodological limitations should be addressed before this conclusion is accepted. The present work should be replicated with two modifications. First, to ensure that participants have sufficient time to process cue words, a typical presentation method should be used, such as aural cueing (e.g. Chen & Spence, 2011; Lupyan & Thompson-Schill, 2012). Second, researchers should select a set of emotional expressions that have a clear set of category-diagnostic features. Using stimuli that clearly possess these distinctive features is an important requirement when testing the assumptions of the label-feedback model (Lupyan, 2007, 2012). This may be achieved by choosing a different expression of interest (e.g. ‘anger’), though researchers should be mindful of the potential for stereotypical gender/expression conjunctions to influence their results (e.g. Hess et al., 2000, 2004, 2009). Alternatively researchers could morph targeted action units of the face to exaggerate the few category-diagnostic features that are associated with expressions of sadness (e.g. Matsumoto & Hwang, 2014).



## Chapter 7

### **Do emotion-related, action words proportionally evoke discrete affective knowledge? Using a categorical method to generate and validate a stimulus set**

#### **7.1 Introduction**

Emotion words are not just ‘words’. Various researchers suggest that emotional words are tightly linked to wider ‘conceptual knowledge’ about emotion, or a repository of information that we accrue through our own affective experiences (e.g. Lindquist, 2009; Vigliocco, Meteyard, Andrews & Kousta, 2009; Kousta, Vigliocco, Vinson, Andrews & Del Campo, 2011; Barrett, 2013; Lindquist & Gendron, 2013). Arguably, a functional relationship is formed; words provide a context for meaningfully interpreting and talking about new emotional experiences, by activating previous knowledge (e.g. Lindquist et al., 2006; Barrett et al., 2007; Lindquist, 2009; Lindquist & Gendron, 2013). The ability to understand, or ‘quantify’ these word-to-knowledge links will be important for researchers interested in using word cues for knowledge activation, or affect induction, in an experimental setting. Previous quantification attempts commonly explore the magnitude to which words evoke knowledge and/or feelings (e.g. Lang, 1980; Newcombe, Campbell, Siakaluk & Pexman, 2012; Siakaluk, Knol & Pexman, 2014; Moffat, Siakaluk, Sidhu & Pexman, 2015). At a higher level of specificity, other researchers attempt to quantify the valence (e.g. Lang, 1980; Estes & Adelman, 2008; Larsen, Mercer, Balota & Strube, 2008; Kousta et al., 2009; Vinson et al., 2014; Kuperman, 2014; Kuperman, Estes, Brysbaert & Warriner, 2014) and/or categorical association of the knowledge evoked by words (e.g. knowledge discretely related to the ‘basic’ emotional states, like ‘fear’; e.g. Stevenson, Mikels & James, 2007a; Briesemeister, Kuckinke & Jacobs, 2011a; Westbury, Keith, Briesemeister, Hofmann & Jacobs, 2014). The present work specifically assesses how best to quantify the knowledge evoked by emotional verbs, or action-related words, which are largely neglected from such work (Pavlenko, 2008).

### 7.1.1 What are emotional verbs and why are we interested in them?

Affective databases, which include extensive lists of rated emotional words, enable researchers to select well-controlled stimuli. The Affective Norms for English Words (ANEW, Bradley & Lang, 1999) and Berlin Affective Word List (BAWL/BAWL-R; Vö, Jacobs & Conrad, 2006; Vö, Conrad, Kuchinke, Urton, Hofmann & Jacobs, 2009) are two such, widely used, examples. Each database includes three classes of word; nouns, adjectives and, to a lesser extent, verbs (Bradley & Lang, 1999; Vö et al., 2009).

Researchers argue that each class of word is represented in the emotion lexicon, and plays a different role (e.g. Pavlenko, 2008; Altarriba & Basnight-Brown, 2010). Adjectives are used as 'emotion words', which commonly refer directly to basic emotional states (e.g. 'Anger', 'Sad'). Verbs denote 'emotion-related' terms, or words that refer to concrete actions and facial expressions associated with an emotional state, without actually naming that emotion, (e.g. 'Cry'). Nouns reflect 'emotion-laden' terms, or words which indirectly express or elicit emotions via their connotations (e.g. 'Cancer'). Theoretically, different classes of word may underpin separate components of an individuals' emotional script.

Prototype theorists suggest that scripts are important for internally storing a repository of emotionally-salient information (e.g. Fehr & Russell, 1984; Shaver, Kirson, O'Connor & Schwartz, 1987; Russell, 1991a). Crucially, these scripts may facilitate emotional attribution via a comparison process; when people observe or experience emotional events they compare current, incoming information with that internally held e.g. 'Sally is going to a funeral; when I went to a funeral I experienced sadness' (e.g. Scherer, 1984; Smith & Ellsworth, 1985; Barrett et al., 2007). With regards to the role of the lexicon in these scripts, researchers argue that the three classes of word reflect a hierarchy of terms, each differing in their strength of connection to the emotion concepts to which they refer (Pavlenko, 2008; Altarriba & Basnight-Brown, 2010; Knickerbocker & Altarriba, 2013).

If a hierarchical view is accurate then emotion-related words, or verbs, may claim an intermediate position. As overt behaviours/expressions often accompany emotional states, they may form an important diagnostic cue for emotional interpretation, and therefore, emotion-concept activation (e.g. Frijda, 1986; Shaver et al., 1987; see also the 'basic' emotion approach; Ekman, 1992). Therefore, the present work seeks to generate a set of

emotion-related verbs, increasing the currently sparse representation in affective databases. Ratings are also generated for those verbs, which capture their position in emotional scripts, or their potential to activate different, discrete emotional concepts.

### **7.1.2 How can we quantify affective grounding for verbs?**

Various measures have been developed to quantify emotional word meaning. These measures are derived from the idea that abstract words are grounded in the emotional experiences that they refer to; words have the power to re-activate or evoke these internal feelings or states (e.g. Wiemer-Hastings & Xu, 2005; Vigliocco et al., 2009; Wilson-Mendenhall, Barrett, Simmons & Barsalou, 2011; Vinson, Ponari & Vigliocco, 2014). For example, the word 'Justice' is understood because it easily evokes certain emotional connotations, such as feelings commonly associated with receiving a jury verdict (e.g. 'joy', 'frustration', 'dismay'; example taken from Newcombe et al., 2012). Dimensional ratings are most commonly used; 'valence' measures the extent to which a word is positive or negative and 'arousal' the intensity to which that word is emotional (e.g. Lang, 1980). In lexical decision tasks these dimensional ratings better predict abstract vs. concrete word processing (Kousta, Vinson & Vigliocco, 2009; Kousta et al., 2011) and emotional vs. neutral word processing (e.g. Estes & Adelman, 2008; Larsen et al., 2008; Kousta et al., 2009; Vinson et al., 2014; Kuperman, 2014; Kuperman et al., 2014). That dimensional ratings predict the way in which emotion words are processed suggests that they appropriately quantify word meaning, or grounding.

Semantic richness measures may also be used (e.g. Newcombe et al., 2012). In particular, the 'emotional experience' variable captures the ease with which a word evokes affective knowledge; making it similar to body-object interaction (Siakaluk, Pexman, Aguilera, Owen & Sears, 2008) and imageability ratings (Schock, Cortese & Khanna, 2012), which represent the strength to which words elicit the motor and sensory properties of their referents. Predictably, Newcombe et al. (2012) found that abstract nouns elicited higher emotional experience ratings than concrete nouns. Subsequent validation attempts also show that emotional experience ratings are significant predictors of abstract (vs. concrete) word processing in Stroop (e.g. Siakaluk et al., 2014) and semantic categorisation tasks (Newcombe et al., 2012; Moffat et al., 2015).

Categorical (or proportional) ratings present a related way to assess affective grounding. Arguably, emotional experience ratings provide an ‘undifferentiated’ quantification, suggestive of the ease with which a word evokes knowledge relevant to a range of emotional states e.g. the word ‘Funeral’ may simultaneously evoke feelings related to ‘sadness’, ‘anguish’ and ‘fear’ (Newcombe et al., 2012; Siakaluk et al., 2014). However, we might be explicitly interested in the likelihood that the word ‘funeral’ evokes feelings of ‘sadness’, in proportion to feelings related to these other possible emotional states. By initially posing the emotional label ‘Sad’, a constrained ‘affective context’ is created, under which participants’ judge the specific relationship between the emotional concept of ‘sadness’, and the word ‘Funeral’. This conceptualisation brings a greater degree of specificity to the notion that words evoke feelings (see also Schwanenflugel & Shoben, 1983; Pecher, Boot & van Dantzig, 2011).

Several researchers already provide categorical ratings for emotion words (e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a; Westbury et al., 2014). For example, Stevenson et al. (2007a) and Briesemeister et al. (2011a) provide a set of participant-generated ratings which represent the strength of association between a set of six ‘basic’ emotional labels/states (e.g. ‘Happiness’, Ekman, 1992) and the words included in the ANEW (Bradley & Lang, 1999) and BAWL-R databases, respectively (Võ et al., 2006, 2009). In addition, Westbury et al. (2014) sampled a large text corpus (HiDeX; e.g. Shaoul & Westbury, 2010) to provide a set of semantic distance values, representative of the linguistic co-occurrence between a set of emotion words and basic emotion labels. In both examples discrete emotional states, denoted by a label, create a constrained ‘affective context’. Participant ratings/distance values indicate the likelihood to which each emotion word proportionally evokes knowledge relevant to those emotional states (e.g. Vigliocco et al., 2009).

Importantly, subsequent work shows that categorical ratings for both English and German words predicted lexical decision latencies (e.g. Briesemeister et al., 2011a, 2011b, Briesemeister, Kuchinke & Jacobs, 2014; Westbury et al., 2014). In particular, words very strongly related to the discrete state of ‘Happiness’<sup>17</sup> were processed faster than neutral

---

<sup>17</sup> Stevenson et al. (2007a, 2007b) assumed that a very strong, or ‘discrete’ association was present when the rating given for the word/label pair was one standard deviation higher than ratings given to that word when paired with all other emotion labels. Using this criterion, 44.54% of the 1,034 words tested were ‘discretely’ related to one or two emotion labels, only. Using



words and words strongly associated with negative states, like ‘Disgust’, ‘Fear’ (Briesemeister et al., 2011a) and ‘Anger’ (Briesemeister et al., 2011b). Briesemeister, Kuchinke & Jacobs, (2014) and Briesemeister, Kuchinke, Jacobs & Braun, (2014) both show that behavioural facilitation was not simply driven by the positive valence of these words. Temporally dissociable ERP components (Briesemeister, Kuckinke & Jacobs, 2014) and topographically distinct brain activity (Briesemeister, Kuckine, Jacobs & Braun, 2014) were found when participants processed words that differed in ‘Happiness’ association (high vs. low), but were matched on valence and arousal. These findings suggest that categorical ratings, or measures which quantify the proportional association between emotion labels and words, are useful for characterising the way abstract words are processed.

### **7.1.3 Why might categorical ratings provide a particularly useful way to quantify the emotional meaning of verbs?**

Currently it remains unclear whether categorical ratings are equally useful for quantifying affective grounding for all classes of emotion word. Work using the DENN-BAWL focused exclusively on emotional nouns (Briesemeister et al., 2011a; Briesemeister, Kuchinke & Jacobs, 2014) and it unclear whether nouns, adjectives and verbs were equally sampled in Briesemeister et al’s. (2011b) and Westbury et al’s. (2014) work. Importantly, two arguments are posed to suggest why a categorical approach may be particularly advantageous for characterising emotion verbs, or emotion-related words (as defined by Pavlenko, 2008).

First, use of categorical ‘method’, in which emotion labels are posed to create an explicit affective context may be useful, as emotion-related action words arguably have ‘dual’ experiential grounding. To demonstrate, although action words like ‘Jump’ may have several types of emotional connotation (e.g. ‘to jump for joy’, to jump is response to frightening or surprising stimulus), the word is still predominantly classed as ‘concrete’, as it refers to an overt action with obvious sensorimotor correlates. As such, verbs have the potential to be ‘understood’ both via re-activation of affective and sensorimotor knowledge; a situation not shared by other classes of emotion word (e.g. nouns and

---

similar methods, Briesemeister et al. (2011a) found that 25.18% of all nouns included in BAWL/BAWL-R (Vö et al., 2006, 2009) were ‘discretely’ associated with a particular emotion label.

adjectives, Vigliocco et al., 2009; Vinson et al., 2014). Indeed, processing of words directly related to emotional behaviours e.g. 'Smile', commonly evoke motor potentials in face and body-specific regions for performing that action (e.g. . Hauk, Johnsrude & Pulvermüller, 2004; Foroni & Semin, 2009; Niedenthal et al., 2009; Moseley, Carota, Hauk, Mohr & Pülvermüller, 2012) and improve understanding of these expressions, when shown by actors (e.g Halberstadt et al., 2009).

Physiological evidence supports the notion that it is comparatively difficult to extract emotional meaning from verbs. Comparing across paradigms, the event-related potentials commonly associated with early affective processing of single emotion words (e.g. the Early Posterior Negative, or EPN; Junghöfer, Bradley, Elbert & Lang, 2001) are commonly evidenced at a later onset for emotional verbs (Schacht & Sommer, 2009; Palazova, Mantwill, Sommer & Schacht, 2011) than for emotional nouns (e.g. Kanske & Kotz, 2007; Kissler, Herbert, Peyk & Junghöfer, 2007) or adjectives (Herbert, Kissler, Junghöfer, Peyk & Rostroh, 2006; Herbert, Junghöfer & Kissler, 2008). However, these differences are attenuated when verbs are processed within a context that makes affective processing salient. For example, when participants made a lexical decision about a verb preceded by an emotionally-relevant noun (e.g. 'lover-kiss') Schacht and Sommer (2009) reported comparable onset potentials to those for emotional nouns and adjectives. Applying a similar manipulation, Palazova, Sommer and Schacht, (2013) found comparable EPN onsets when emotional verbs referred to more concrete, context-invariant behaviours, which had clear affective connotations (e.g. 'to dance' vs. 'to hope'). Similarly, it is anticipated that providing a specific affective context, in the form of basic emotion labels, will bias participants to consider the affective grounding of words that refer to emotional actions.

Second, it is argued that categorical ratings may be more fluently generated for emotional action-words, than nouns and adjectives. Researchers suggest that emotional scripts are 'diagnostic', facilitating interpretation through a comparison process (currently experienced vs. stored knowledge; e.g. Fehr & Russell, 1984; Shaver et al., 1987; Russell, 1991a). In order for script content to be 'diagnostic' in this way, stored information is likely structured around 'hubs' that correspond to discrete, basic emotional states (e.g. 'sadness'; Shaver et al., 1987). When scripts are used in this way, behaviours (represented by verbs) arguably represent a crucial component; overtly presented facial expressions and behaviours are a quick and convenient indicator that someone feels a particular way (e.g. Ekman, 1992). In a task where specific affective contexts are posed, participants may refer

directly to their internal script to judge the affective grounding of actions words, rather than simply relying on the connotative value of those words, as they might for nouns and adjectives.

#### **7.1.4 Summary of the present work**

In sum, the present work aims to explore whether a categorical approach can be used to examine the affective, experiential knowledge that partially grounds action-word meaning. Importantly, in the first study six basic emotion labels are posed to create a constrained 'affective context' (e.g. 'Sad'). Participants will self-generate emotional action words that they commonly associate with each, basic emotional state. Generation frequencies, per action word, will be indicative of the likelihood that the word evokes affective knowledge relevant to the paired emotion label. In the second study a rating task will be conducted to validate use of generation frequencies as a measure of associative strength. Action words are paired with the emotion labels to which they have been most frequently generated, and rated according to the strength of that association.

One potential use for this dataset was presented in Chapter 6. Here verbs proportionally associated with discrete emotional states were used to activate emotion-related knowledge. Participants encountered an emotion-related verb before deciding whether a briefly presented face did or did not display an emotional expression. We explored whether emotionality decisions were facilitated when the presented verb was strongly, and discretely, associated with the emotional expression displayed. Findings suggest that activated conceptual knowledge did not have a top-down influence on emotion perception. However, other paradigms are later identified that may specifically benefit from using verbs to activate affective knowledge.

## 7.2 Study 1- Identifying action words that proportionally evoke affective knowledge

In study 1 emotion labels will be used to provide a constrained, ‘affective context’.

Following Stevenson et al. (2007a) and Briesemeister et al. (2011a), the accepted, ‘basic’ emotion labels will be presented (‘Happy’, ‘Sad’, ‘Fear’, ‘Anger’, ‘Disgust’ and ‘Surprise’ e.g. Ekman, 1992). Basic, or commonly experienced emotions, should be fluently associated with behavioural referents (e.g. Shaver et al., 1987).

In contrast to previous rating work, a highly constrained semantic feature-generation task will be conducted. Participants are instructed to self-generate multiple single-word actions that they commonly associate with experiencing each discrete emotional state (see McRae, Cree, Seidenberg & McNorgan, 2005; Vinson & Vigliocco, 2008 and Buchanan, Holmes, Teasley & Hutchison, 2013 for broader examples of semantic feature generation). Explicit instructions are important as action words are rarely generated in response to emotion labels during feature generation (e.g.; Hutchison et al., 2010; Buchanan et al., 2013).

Encouraging participants to engage separately with each emotion label will likely lead to a wider stimulus set, as rating methods often produce a ‘happiness asymmetry’ (many words are strongly associated with ‘happiness’, but far fewer words are associated with discrete, negative states e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a).

Overall, the likelihood that an action word evokes discrete affective knowledge will be measured according to the frequency of participants who endorse the pair (e.g. McRae et al., 2005).

### 7.2.1 Method

#### 7.2.1.1 Participants

Twenty-five participants generated action words (17 female, 8 male;  $M_{\text{age}} = 27.24$ ,  $SD_{\text{age}} = 7.63$ ). All participants reported themselves to be native English speakers; although 7 participants spoke a second language, they did not consider themselves fluent. An opportunity recruitment method was used. Participants responded to links posted on research recruitment websites and completed the study online (e.g. <http://www.psych.hanover.edu/research/exponnet.html>;

<http://www.onlinepsychresearch.co.uk>; <http://www.in-mind.org/content/online-research>; <http://www.reddit.com/r/SampleSize>).

### **7.2.1.2 Procedure**

All materials, including informed consent items, were presented using the Survey Monkey platform (<http://www.surveymonkey.com>, Survey Monkey Inc. Palo Alto, California, USA). Participants ticked a series of boxes to confirm that they understood task instructions and gave their informed consent to take part. Participants were then asked to carefully read Pavlenko's (2008) definition of an emotion-related action word. Definitions were edited to include relevant examples.

*“ ‘Emotion-related’ words are used to describe behaviours related to a particular emotional state, without naming the actual emotion. For example, the word ‘Cry’ might describe the behaviour of someone feeling Sad while the word ‘Smile’ may describe the behaviour of somebody who is Happy.”*

Participants were directed to six basic emotion labels, listed below the definition ('Sad', 'Happy', 'Anger', 'Disgust', 'Surprise' and 'Fear', Ekman, 1992). They were asked to generate as many action words as they could, which they personally associated with each emotion label. Separate boxes were provided for participants to type their examples. Participants were instructed to provide single-word answers and to avoid label synonyms or adverbs (e.g. 'Sadness', 'Sadly'). They were also discouraged from using the internet to generate responses. Participants were asked to work on the basic labels sequentially and labels were presented in a randomised order across participants. There was no time limit imposed on word generation.

## **7.2.2 Results**

### **7.2.2.1 Data modifications and modal exemplars**

In total, participants generated 362 unique words, across the six emotion labels. On average, participants each generated 27.32 words during the task (SEM = 3.05). The data were parsed in various ways to determine an acceptable set of action words, which were 'modally' associated with one or more emotion labels (see McEvoy & Nelson, 1982; Doost, Moradi, Taghavi, Yule & Dalglish, 1999; McRae et al., 2005; Vinson & Vigliocco, 2008 for similar methods). The Cambridge Online English Dictionary

(<http://www.dictionary.cambridge.org/>) and an online Thesaurus (<http://www.Thesaurus.com>) were consulted to support modifications.

First, words were deemed unacceptable if (a) they did not describe a concrete action (e.g. 'Tearful'), or (b) were synonyms for the emotion label itself (e.g. 'Afraid', generated in response to 'Fear'). Second, multiple-word responses or phrases were only retained if they could be simplified to a single word with the same, or similar, meaning, for example, 'Sharp intake of breath' was replaced with 'Gasp'. Third, merging techniques were used either when participants provided grammatical derivatives or plurals of the same word (e.g. 'Run', 'Run', 'Runs', 'Running', 'Ran away') or generated synonyms for action words that had already been provided by themselves or other participants (e.g. 'Scream' and 'Shriek'). In the former case, plurals were changed to their singular form and grammatical derivatives were merged and represented by the simplest version, provided their meaning did not change (e.g. 'Run').

The second type of merging (non-derivative words) was wholly motivated by the need to develop stimuli for our subsequent validation attempt (study 2). This study required only six action words, each of which held the strongest association with one of the six emotion labels. Therefore it was important to ensure that words with the same or very similar meanings were grouped together, and their frequencies summed, to aid assessment of how strongly those related words evoked similar types of affective knowledge<sup>18</sup>. Strict criteria were imposed for this form of merging. Action words were only classed as synonymous if there was evidence of forward and backward association e.g. when 'Laugh' was entered into the thesaurus 'Giggle' was given as a synonym, and when 'Giggle' was entered into the thesaurus, 'Laugh' was given as a synonym. It is important to acknowledge that some action words could have multiple meanings when presented in isolation (e.g. Schacht & Sommer, 2009). For example, the action word 'Jump' could mean 'to leap, spring or skip', 'to recoil' or 'to avoid' (definitions taken from <http://www.thesaurus.com>). In

---

<sup>18</sup> Although this type of merging helped to identify the top-six modal action words, for use in study 2, it necessarily inflated the apparent frequency-based strength of association between those core action words and corresponding emotion labels. Interested readers are encouraged to consult Appendix B, in which all modal exemplars are listed alongside unmerged generation frequencies, which provide a clearer estimation of the strength with which individual action word evokes affective knowledge relevant to different emotion states.

these cases the participants' intended meaning was discerned by considering the emotion label to which the word had most frequently been generated. As the word 'Jump' was frequently endorsed for the emotion labels 'Surprise' and 'Fear' it went unmerged with 'Skip', which although a synonym, was only given in response to the label 'Happy'. Here it was considered likely that the two words likely had different intended meanings, each congruent with the core emotion concept to which they had been modally generated (see Buchanan et al., 2013 for similar methods).

Where merging occurred, generation frequencies for both/all action words were added together. For non-derivative synonyms the dominant response was retained, based on existing frequencies (i.e. the action word generated by the highest number of participants). This exemplar became the 'core' action word and non-dominant responses were subsumed and became 'subsidiary' action words. For example, in response to the label 'Sad', 'Cry' became a core action word and the synonyms 'Weep' and 'Sob' became subsidiaries<sup>19</sup>. The number of participants who generated the action words 'Cry', 'Weep' and 'Sob' were added together to provide a frequency total for the core action word ('Cry'). Note that frequencies could exceed 25 if participants had provided both core and subsidiary action words in response to the same emotion label.

---

<sup>19</sup> It was particularly difficult to make merging decisions about the exemplar 'Cry'. As this exemplar was given in response to the 'Sad', 'Anger', 'Fear', 'Happy' and 'Surprise' categories, consideration of cue word could result in two (or more) definitions being accepted. To illustrate, when generated in response to 'Sad(ness)' the definition 'to weep or make sad sounds' would be most relevant, but when generated in response to 'Anger' the definition 'to call out/yell' was most appropriate (definitions taken from <http://www.Thesaurus.com>). Arguably participants may have had either meaning in mind when they generated the exemplar in response to the remaining emotion labels, which complicated the issue. The decision was made to merge 'Cry' contingent on the first sadness-related definition, only, as the exemplar was most frequently given in response to the 'Sad' category. 'Cry' become the core action word, and 'Weep' and 'Sob' the subsidiary action words. As 'Cry' was already the unmerged, top modal exemplar for 'Sad(ness)', this merging decision did not change the modal response that was chosen for the 'Sad' label in study 2. If the decision had been made to alternatively (or additionally) chosen to merge according to the second definition, 'Cry' could have been grouped with 'Scream', 'Shout' and 'Shriek'. This was problematic as our criteria suggested that 'Scream' and 'Shriek' could be merged with 'Yell', but 'Yell' could not be merged with 'Cry'. Therefore, the strategy adopted was both simpler, and more conservative.

Following these steps the stimulus set still contained a large number of 'idiosyncratic' responses, generated by only one participant in response to a particular label (124 words, 56.88% of remaining responses). These exemplars are unlikely to represent words which commonly evoke discrete affective knowledge; therefore, the decision was made to exclude these responses (see Buchanan et al., 2013). Following removal there remained 15 unique core action words, 19 subsidiary action words and a further 17 'modal' action words (51 unique action words in total). Here 'modal' refers to an action word that was generated by two or more participants, but was not synonymous with other responses and went unmerged. This final selection represents 14% of the total number of unique words originally generated.

The top three most frequently generated action words, per emotion label, are shown in Table 7.1. Response frequencies are shown in parenthesis, in the second column. For 'core' exemplars, frequencies also include the number of participants who generated subsidiary action words (corresponding subsidiary words are shown in the column three, where appropriate). The full set of acceptable action words (core, subsidiary and modal), are provided in Appendix E (see Portch et al., 2015 for further supplementary materials).



**Table 7.1 Top three, most frequently generated action words, per emotion label.**

*Top responses are presented alongside subsidiary responses (where appropriate). Response frequencies for each action word are presented within parenthesis in the second column. These frequencies represent merged totals where a corresponding subsidiary action word is shown in the third column.*

<b>Emotion Label</b>	<b>Most frequent action words (<i>response frequency</i>)</b>	<b>Corresponding, subsidiary action words (<i>core action word</i>)</b>
<b>Anger</b>	Scream (34); Hit (13); Cry (7)	Shout/Yell/Shriek ( <i>Scream</i> ); Punch ( <i>Hit</i> ); Sob/Weep ( <i>Cry</i> )
<b>Happy</b>	Smile (27); Laugh (20); Dance (10)	Grin ( <i>Smile</i> ); Giggle ( <i>Laugh</i> ); Skip ( <i>Dance</i> )
<b>Sad</b>	Cry (23); Frown (9), Withdraw (7)	Sob/Weep ( <i>Cry</i> ); Grimace ( <i>Frown</i> )
<b>Disgust</b>	Recoil (7); Frown (6); Gag/Vomit (5 each)	Cringe ( <i>Recoil</i> ); Grimace ( <i>Frown</i> ); Retch ( <i>Gag</i> )
<b>Fear</b>	Hide/Run (13 each); Shiver (11); Cry (9)	Avoid ( <i>Hide</i> ); Shake ( <i>Shiver</i> ); Sob/Weep ( <i>Cry</i> )
<b>Surprise</b>	Jump (15); Gasp (13); Scream (12)	Inhale ( <i>Gasp</i> ); Shout/Yell/Shriek ( <i>Scream</i> )

Analysing by exemplar, 78.43% of all modal action words were generated in response to one emotion label only, leaving 21.57% that were generated for multiple labels. This distinction was present even for the most frequently generated action words, shown in Table 7.1. When only these exemplars were considered, 15.79% represented the most frequent responses for more than one emotion label, and 68.75% were at least modally generated in response to one or more other emotion labels.

### 7.2.3 Discussion

In the present study several constrained, 'affective contexts' were introduced to identify action words with the potential to evoke affective knowledge, relevant to different emotional states (e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a). The greater number of participants who generated a particular action word in response to an emotion label, the greater likelihood that that action word would be situated in, and evoke affective knowledge relevant to that emotion. Both acceptable action words (modal, core and subsidiary) and generation frequencies are available in Appendix E.

Our findings suggest that participants generated two types of action word. First, there were a number of action words that were generated in response to one emotion label only, suggesting a very strong association with a discrete emotional state. This supports the view that emotional words, like concrete words, may be understood by evoking very specific representations of situations in which their referents appear (e.g. Schwanenflugel & Shoben, 1983; Pecher et al., 2011). Second, there were also a subset of words that were proportionally related to a number of different emotional states (overlapping exemplars e.g. Newcombe et al., 2012). This supports the notion that some words are 'generally' evocative and have the potential to re-activate affective knowledge, proportionally related to a range of emotional states (e.g. Newcombe et al., 2012). The word 'Cry' is a particularly good example; this exemplar was frequently generated in response to the 'Sad', 'Anger' and 'Fear' labels, and also by a smaller number of participants for the labels 'Happy' and 'Surprise'.

Importantly, both 'discrete' and overlapping exemplars have been found in previous rating work, using the ANEW (Bradley & Lang, 1999; Stevenson et al., 2007a) and BAWL-R databases (Vö et al., 2009; Briesemeister et al., 2011a). Specifically, Stevenson et al. (2007b) found that 44.54% of ANEW words obtained ratings to suggest that they were proportionally associated with one (or two) discrete emotions, and 22.70% of words were associated with three or more emotion labels (overlapping exemplars). In contrast to earlier expectations, this may suggest that a categorical approach is equally useful for action words, as it is for other classes of emotional words (Briesemeister et al., 2011a exclusively used nouns and Stevenson et al., 2007a, 2007b, heavily sampled nouns and adjectives).

In study 2 a rating task was used in order to validate the assumption that generation frequencies are an appropriate measure of associative strength.

### **7.3 Study 2- Validating associations between action words and emotion labels**

In study 2 the typicality of self-generated action words, and the stability of action word-to-label associations, would be assessed. A rating task, similar to that used by Stevenson et al. (2007a), was chosen. Participants were asked to rate the relationship between the six most frequently generated action words, and each discrete, emotion label. Emotion labels and action words would be presented within a sentence e.g. “if you see someone ‘Recoil’ how likely are you to think that they are feeling the following emotion?... ‘Disgust’ ”. Primarily, it was expected that ratings would indicate a comparatively stronger association between action words and the emotion labels to which they were most frequently generated, during study 1. This would confirm that the word is understood because it disproportionately activates affective knowledge relevant to that emotional state.

Further validation was particularly important to assess whether the top exemplars ‘Cry’ and ‘Smile’ were as strongly linked to the respective emotional states of ‘Sad(ness)’ and ‘Happ(iness)’ as their generation frequencies implied. This was a concern as both action word/label pairs had been included as examples in the task instructions for study 1, so frequent endorsement may not reflect spontaneous generation. This may also explain why the word ‘Cry’ was given so frequently across all affective contexts. In addition, although participants were discouraged from using the internet to generate their responses during study 1, it was difficult to definitively rule out the possibility that they had done so. Use of external sources may have inflated frequencies, artificially creating modal exemplars. Although this seems unlikely, given the larger number of idiosyncratic vs. modal exemplars obtained, it is important to address this possible methodological issue.

### 7.3.1 Method

#### 7.3.1.1 Participants

Eighty-six participants completed the rating task (57 female;  $M_{\text{age}} = 27.78$ ,  $SD_{\text{age}} = 10.04$ ). Each participant indicated whether they spoke any languages in addition to English and estimated how many years they had been able to do so. Those judged to be fluent bilinguals or multi-linguals were omitted from the sample. This resulted in seven exclusions (four females), leaving a remaining sample of seventy-nine participants (53 female;  $M_{\text{age}} = 26.99$ ,  $SD_{\text{age}} = 9.24$ ).

An opportunity recruitment method was used; participants responded online, to links posted on social media sites (see Study 1). The study was presented using the Survey Monkey platform (<http://www.surveymonkey.com>, Survey Monkey Inc. Palo Alto, California, USA). As before, informed consent items were embedded in the online survey and participants agreed to take part by ticking a series of boxes. There was no time limit imposed.

#### 7.3.1.2 Design and Procedure

A 6 (Emotion Category: 'Happy', 'Surprise', 'Sad', 'Fear', 'Anger', 'Disgust')  $\times$  2 (Typicality: Typical, Atypical) within-participants design was employed. Participants each made 36 ratings, based on all combinations of six discrete emotion labels and the action words most frequently endorsed in response to each of these labels (study 1). Feature generation data determined whether emotion label / action word pairings were typical (e.g. six pairs, 'Happy' and 'Smile'), or atypical (30 pairs, e.g. 'Sad' and 'Smile').

Participants were presented with an open-ended sentence for each rating, which included either an emotion label or action-word e.g. "if you are feeling 'Sad', how likely are you to act in the following way?". Participants were invited to substitute each of the six action words (or labels) into the end of this sentence (e.g. 'Cry'), and to provide a likelihood rating for each label/action word pairing. Ratings were always made on a five-point Likert-style scale, anchored 'Very Unlikely' (1) to 'Very Likely' (5). After all six ratings were submitted,

participants were presented with the next open-ended sentence, which included a new emotion label (or action word)<sup>20</sup>.

Overall, participants made ratings in six, separate blocks. In each block a different label (or action word) would be rated against each action word (or label), respectively. Block order was counterbalanced across participants. Within a particular block, participants encountered each of the six ratings in a fixed order. Although fixed per participant, this order was randomised per block, to ensure that the typical pairing was not always presented in the same rating position [e.g. in the 'Sad' block participants rated associations with action words in the following order: 'Smile', 'Cry'(typical) 'Jump'..., but in the 'Happy' block they rated action words in a different order: 'Hide', 'Scream', 'Smile'(typical)..]. Therefore, while block order differed, rating order within blocks was the same for all participants within a particular condition.

### 7.3.1.3 Materials

The six basic emotion labels were re-used from study 1 ('Fear', 'Happy', 'Sad', 'Disgust', 'Anger' and 'Surprise', e.g. Ekman, 1992). The most frequently generated action words for each emotion label were selected from the merged, feature generation data. They were as follows: 'Scream' (matched with 'Anger'); 'Smile' ('Happy'), 'Cry' ('Sad'), 'Recoil' ('Disgust'), 'Hide' ('Fear') and 'Jump' ('Surprise').

---

<sup>20</sup> Please note that 'rating direction' was varied, between-participants (i.e. whether participants were primarily presented with an emotion-related action-word to rate against all six emotion labels, or an emotion label to rate against all six emotion-related actions words, see Portch et al., 2015). 40 participants were assigned to the label-to-action word rating condition (31 female,  $M_{\text{age}} = 25.65$ ,  $SD_{\text{age}} = 9.56$ ) and 39 participants were assigned to the action-to-label rating condition (29 female;  $M_{\text{age}} = 28.35$ ;  $SD_{\text{age}} = 8.70$ ). There were no significant interactions between (a) rating direction and emotion category ( $p = .90$ ), nor (b) rating direction and typicality ( $p = .27$ ). Therefore the rating direction factor was omitted from further analyses.

## 7.3.2 Results

### 7.3.2.1 Data preparation

For each emotion label, two mean ratings were calculated, per participant. The ‘typical’ mean was the rating given to the most typical label and action word pairing, according to the feature generation data (e.g. ‘Cry’ and ‘Sad’). The five remaining ratings were summed and then averaged to produce a grouped ‘atypical’ score (mean scores for the full set of 36 label/action word ratings are shown in Appendix F). ‘Typicality’ reflects the strength of association between action words and emotion labels (operationalised here as high or low).

### 7.3.2.2 Analysis

A 6 (Emotion Category: ‘Happy’, ‘Surprise’, ‘Sad’, ‘Fear’, ‘Anger’, ‘Disgust’) × 2 (Typicality: Typical, Atypical) within-participants ANOVA was performed. Bonferroni corrections were applied to account for multiple comparisons<sup>21</sup>.

Importantly there was a significant main effect of Typicality,  $F(1, 78) = 701.49, p < .001$ ,  $MSE = 1.04, \eta_p^2 = 0.90$ . Participants gave higher mean likelihood ratings to typical pairings ( $M = 4.31, SEM = 0.063$ ), than grouped atypical pairings, ( $M = 2.56, SEM = 0.055$ ), on a scale from 1-5. This effect supports the use of feature-generation frequencies (study 1) to assess the strength of label-action word associations.

There was also a significant main effect of Emotion Category,  $F(4.41, 343.94) = 19.17, p < .001, MSE = 0.51, \eta_p^2 = 0.20$ , and a significant interaction between Typicality and Emotion Category,  $F(4.23, 330.62) = 34.56, p < .001, MSE = 1.04, \eta_p^2 = 0.31$ . This interaction prompted further investigation of the Typicality effect for each discrete, emotion category (see Table 7.2 and Figure 7.1).

---

<sup>21</sup> Greenhouse-Geisser corrections are reported for the main effect of ‘Emotion Category’ and the interaction between ‘Emotion Category’ and ‘Typicality’.

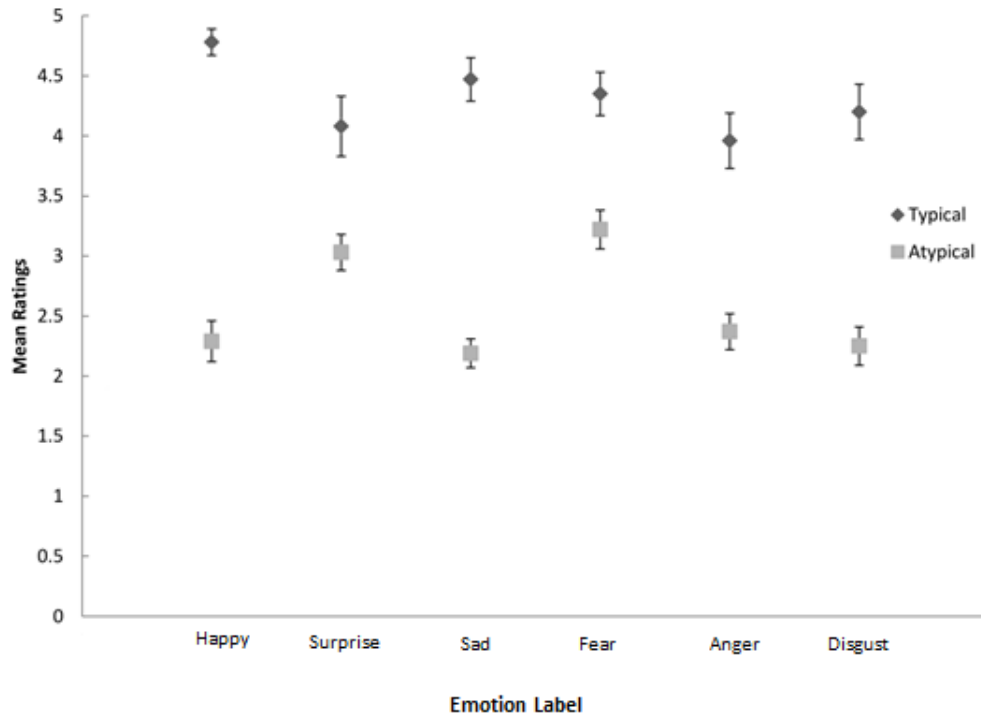
**Table 7.2 Mean typical and atypical ratings, t, p and d statistics, per emotion label.**

<i>Emotion Category</i>	<i>Typical Mean (SEM)</i>	<i>Atypical Mean (SEM)</i>	<i>t</i>	<i>p</i>	<i>d</i>
<b>Happy</b>	4.78 (0.053)	2.29 (0.084)	24.24	< .001	4.00
<b>Surprise</b>	4.08 (0.12)	3.03 (0.074)	9.06	< .001	1.58
<b>Sad</b>	4.47 (0.091)	2.19 (0.062)	23.86	< .001	3.31
<b>Fear</b>	4.35 (0.090)	3.22 (0.079)	13.04	< .001	1.51
<b>Anger</b>	3.96 (0.11)	2.37 (0.075)	13.22	< .001	1.85
<b>Disgust</b>	4.20 (0.12)	2.25 (0.081)	13.12	< .001	2.21
<b>Totals</b>	4.31 (0.098)	2.56 (0.077)	-	-	-

*\*Degrees of Freedom were always (1, 78)*

All six paired samples t-tests were significant and in the anticipated direction (typical category/action word pairings received higher likelihood ratings than the grouped atypical pairings). Therefore, the interaction likely reflects general differences in the strength with which typical action words evoke affective knowledge related to corresponding emotions, all effects being conventionally large ( $d > 0.8$ ).

In support, Figure 7.1 (below) shows that the 95% confidence intervals for average typical ratings and the summed average of atypical ratings did not overlap for any emotion category. That typicality predictions were supported weakens the suggestion that participants used the internet to generate their responses during study 1. In addition, typicality effects were present for the specific pairings of 'Happy' / 'Smile' and 'Sad' / 'Cry', reducing the likelihood that these associations were simply generated as a result of their inclusion in task instructions.



**Figure 7.1 Mean typical and atypical ratings, per emotion label (error bars represent 95% CI).**

### 7.3.3 Discussion

Ratings confirm that participants were more likely to associate action words with the emotional label to which they had been typically generated in study 1, validating the use of generation frequencies to assess proportional activations. This lessens the likelihood that endorsement was inflated by the examples included in task instructions, or use of the internet.

However, it is important to acknowledge the following limitation: task design meant that participants rated one label (or action word) in association with all six action words (or labels) before they were presented with the next label (or action word). Participants might then have considered it appropriate to adopt a 'relative' rating strategy, in which they simultaneously compared the likely association between all six items and the dominant label, or action word. Typical pairings may then receive the highest likelihood ratings because they represent the 'best option', rather than giving a true indication of the way in which action words proportionally evoke affective knowledge, relevant to the presented



label. Per block participants also responded to the six pairings in the same order. Therefore, any biases that this presentation strategy encouraged would be applicable to all participants, despite care to vary the position in which typical pairings were presented. However, our data suggest it is unlikely that participants automatically employed a comparative rating strategy. If they had all atypical pairings would be expected to receive very low likelihood ratings. Although some of the averaged, atypical ratings were below the scale midpoint (2.5; 'Happy', 'Anger', 'Sad' and 'Disgust', see Table 7.2), others were higher (3.22 for 'Fear' and 3.03 for 'Surprise'). These particular discrepancies are predicted as some of the typical action words included in the task were also amongst the top, three modal exemplars generated for other emotion labels, during study 1. This was the case for the three labels that attracted the highest average atypical ratings ('Fear', 'Surprise' and 'Anger'). For example, although the action word 'Cry' represented the typical exemplar for the label 'Sad', it was also frequently generated in response to the emotion labels 'Fear' and 'Anger' (see Table 7.1). Including these overlapping exemplars meant that, for some emotion labels, not all 'atypical' exemplars were equally 'atypical', inflating the averaged atypical ratings. Importantly, these findings indicate that participants judged each action word/label pair based on the 'absolute' association between the two words, rather than making a comparative judgment that was biased by the presence of an obviously 'typical' pairing. They also support the idea that 'typicality' is expressed as a matter of degree, as action words may simultaneously evoke affective knowledge relevant to several emotional states (e.g. Newcombe et al., 2012).

## 7.4 General Discussion

The present work provides a set of emotion-related action words alongside data to show how strongly each word evokes emotional knowledge relevant to several, discrete emotional states. This work is consistent with the proposal that emotion words are grounded in affective knowledge (e.g. Vigliocco et al., 2009) and complements previous research, by exploring whether word-to-knowledge links are constructed, at least partially, in a categorical or proportional fashion (e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a, 2011b; Briesemeister, Kuchinke & Jacobs, 2014; Westbury et al., 2014). Despite suggesting that a categorical approach might be most useful for quantifying emotion verb meaning, a similar proportion of 'categorisable' words were found in present and previous work (e.g. Stevenson et al., 2007a, 2007b; Briesemeister et al., 2011a). Perhaps then the

approach is similarly compatible for all three classes of word within the emotion lexicon (nouns, adjectives and verbs).

Action words were elicited during a constrained feature-generation task (e.g. McRae et al., 2005; Vinson & Vigliocco, 2008; Buchanan et al., 2013). Emotion labels were used to create (and constrain) six, different 'affective contexts' (e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a). This method allowed assessment of the strength with which each action word elicited specific affective knowledge; the larger the number of participants who endorsed the pair the greater the likelihood that the word proportionally evokes knowledge relevant to that emotional state. Data from the rating task (study 2) confirm that the action words most frequently elicited in study 1 were more likely to be associated with the emotion label to which they had been generated (typical pairs), than to other emotion labels (atypical pairs).

The current approach, and data produced, may provide an alternative way to select emotional stimuli (e.g. Briesemeister et al., 2011a, 2011b; Briesemeister, Kuchinke & Jacobs, 2014). Work that uses this resource was presented in Chapter 6. Here participants read action words proportionally associated with different discrete emotional states before making affective decisions about emotional faces. Decisions were compared after participants read action words that should strongly evoke affective knowledge, relevant or irrelevant to the emotional expression displayed. In contrast to predictions, participants performed similarly, independent of the congruence between the repeated action-word and emotional expression. Findings may suggest that activated emotional knowledge does not provide a useful top-down constraint for the perception of emotional expression (e.g. Pylyshyn, 1999).

Arguably, this set of stimuli may be used in a similar way for facial mimicry work. Previous research shows that participants mimic facial expressions when they encounter expression-congruent emotion words (e.g. Foroni & Semin, 2009), and that mimicry leads to enhanced processing of subsequently presented emotional stimuli e.g. valence-congruent sentences (e.g. Havas, Glenberg & Rinck, 2007), emotion-related cartoons (Foroni & Semin, 2009) and facial expressions (e.g. Halberstadt et al., 2009). We might expect emotion-related action words to more strongly elicit congruent facial mimicry, given their dual grounding in affective (Vinson et al., 2014) and sensorimotor knowledge (e.g. Hauk et al., 2004;

Niedenthal et al., 2009; Moseley et al., 2012). However, few studies incorporate action words and those that do find inconsistent evidence for a verb (vs. adjective) advantage (Feroni & Semin, 2009; Niedenthal et al., 2009, see also Halberstadt et al., 2009 for use of other word stimuli). If these findings reflect inconsistent use of linguistic stimuli then the present data may help by providing a larger set from which to select stimuli. Further, by choosing words that are both proportionally related to a particular emotional state and related to facial actions, researchers may extend investigations into whether language-mediated facial mimicry is 'category' or 'valence' driven. Specifically, whether reading an action word associated with 'fear' specifically induces mimicry in category-diagnostic features of a fearful face, (Niedenthal et al., 2009; Ponari, Conson, D'Amico, Grossi, & Trojano, 2012) or whether reading any negatively valenced word induces a similar pattern of negative mimicry.

In conclusion, the present work provides a set of English action words, characterised by their proportional likelihood to evoke affective knowledge relevant to different emotional states. Basic emotion labels were presented to create a set of constrained 'affective contexts', both for initial generation of action words (study 1) and validation of the most typical exemplars (study 2). This set of stimuli both complement and extend existing linguistic databases that contain categorical norms (e.g. Stevenson et al., 2007a; Briesemeister et al., 2011a). One possible use for this stimulus set is presented in Chapter 6. Here emotion-related action words are used to activate conceptual knowledge, associated with different, discrete emotional states. This work assesses whether activated knowledge has an impact on subsequent processing of a different type of emotional stimuli (expressive faces). However, the following data are provided to support future research: a list of all acceptable words (generated by at least two participants in response to an label; Appendix E), accompanied by generation frequencies and merging suggestions, and a full set of ratings for the 36 action word/label pairs (Appendix F).



## Chapter 8

### General Discussion

#### 8.1 Key findings

This thesis explored the ways in which language, and linked conceptual knowledge, may help us to perform a critical social task: interpreting emotional expressions. The key findings from the semantic satiation and threshold paradigms are first detailed in the paragraphs below. This thesis also included a third type of work, which examined a related question; specifically, how particular types of emotion word might underpin and activate a network of semantic, affective knowledge (Chapter 7). Due to its contrasting focus and methodology, the key findings, theoretical implications and potential future directions for this work are separately discussed in section 8.7.

The semantic satiation findings, described in Chapters 2 – 5, suggest that reducing access to emotion-relevant semantic knowledge did not specifically impair participant’s ability to discriminate between emotional expressions. Instead, participants demonstrated performance decrements after massed repetition of *any* linguistic label, or non-linguistic activity (repetitive tapping or passive watching). A general massed repetition decrement may implicate the role of non-semantic and/or linguistic factors, inherent when using an immediate repetition paradigm (see full discussion of possible alternate mechanisms in Chapters 4 and 5). In addition, participants showed relative facilitation in performance when they responded to a pair of matching emotional expressions, but only after repeating a judgment-relevant emotion label (e.g. repeat the word ‘Sad’, before seeing a pair of ‘Sad’ expressions). Semantic mechanisms are unlikely to explain this effect as facilitation was shown after both 3 and 30 repetitions of that emotion label. Instead, repetition of a relevant label may either *directly* influence the perception of those two expressions (e.g. Lupyan, 2007, 2012), or may impact on the strategic decision process that governs visual discrimination of categorical stimuli (e.g. Roberson & Davidoff, 2000; Kikutani et al., 2008).

The threshold work (experiment 5), described in Chapter 6, showed that participants were no better to make an emotionality decision about a briefly presented face after being exposed to a judgment-relevant emotion word vs. an irrelevant emotion word, or neutral word. Similar results were obtained both when participants encountered emotion labels (e.g. 'Sad') and emotion-related action words/verbs (e.g. 'Sob'), and when there was a short or long inter-stimulus interval inserted between exposure to words and faces. Instead there were some indications that participants made strategic use of the words that they repeated, perhaps based on the task-wide relationships between those labels and the emotional faces that they encountered (e.g. Pylyshyn, 1999).

## **8.2 Theoretical implications: the construction and label-feedback accounts.**

Key findings are first discussed in relation to the two theoretical accounts described in the general introduction; the construction (e.g. Barrett, 2006b, 2009; Barrett et al., 2007) and label-feedback positions (Lupyan, 2007, 2012). The basic assumptions of these accounts are briefly re-introduced before findings are discussed within the context of each.

The construction account suggests that our interpretation of newly-encountered emotional stimuli is driven by a comparison process, between this new information and the information we have accrued through our previous emotional experiences (e.g. Barrett, 2006b, 2009; Barrett et al., 2007). Experience-based semantic knowledge is stored in an internal repository, organised and re-activated by basic language labels (e.g. 'Sad'). Although there are several possible mechanisms by which activated semantic knowledge may influence perception of facial expression (e.g. Thielscher & Pessoa, 2007; Fox et al., 2009; Anderson et al., 2011; Herbert et al., 2013), it is argued that language labels themselves play an *indirect* role in the process. That is, language labels are responsible only for reactivating semantic knowledge, when required. In contrast, the label-feedback account (Lupyan, 2007, 2012) suggests that language labels may have a *direct* influence on the detection and perception of categorical stimuli, albeit a temporary one. This process relies on the intrinsic links that form between category labels and their perceptual referents; exposure to a category label temporarily creates a perceptual expectancy that, when met, simultaneously accentuates and draws our attention towards the category-diagnostic features of that object. To illustrate, when we see a 'chair' following exposure to

the label 'chair' we are better to strategically process the 'chair-like' properties that differentiate that object from other items of furniture e.g. has four legs, has a seat (e.g. Lupyan, 2008a). The label-feedback hypothesis has been extensively tested and validated in the object perception literature but there are indications that the theory may apply to the processing of other complex categorical stimuli, like emotional expressions (e.g. Roberson et al., 2007; Kikutani et al., 2008; Puri & Wojciulik, 2008; Esterman & Yantis, 2010; Calvo et al., 2012; Lupyan & Thompson-Schill, 2012).

### **8.2.1 Strand 1: The semantic satiation work (Chapters 2–5)**

Following Lindquist et al. (2006), the modified semantic satiation experiments (1-4) were used to directly test the construction position. However, there was no strong evidence in the current work to suggest that discrimination of emotional expressions suffered after participants completed an activity that should temporarily reduce access to emotion-relevant, semantic knowledge (repeating an emotion label 30 vs. 3 times; see also Gendron et al., 2012 and Lindquist et al., 2014). This was even the case when the satiated emotion label matched both of the expressions shown at test (experiments 2a and 2b); a circumstance under which discrete semantic knowledge, highly relevant for the interpretation of the expression pair, should be unavailable (e.g. the satiation comparison, Smith & Klein, 1990; Lindquist et al., 2006). As such, conclusions arising from the current work are not in line with Lindquist et al. (2006), nor the construction position. Importantly, the present experiments (2a – 4) introduced several baseline conditions to robustly delineate between semantic and non-semantic explanations for the massed repetition decrement (see Chapters 3 and 4). These baseline conditions were lacking from Lindquist et al.'s. (2006) original experiments. As such, the present findings confirm that it is difficult to conclude whether the massed repetition decrements reported in Lindquist et al.'s. (2006) work were indicative of a semantic mechanism (emotion-specific) or non-semantic mechanisms (e.g. Tian & Huber, 2010).

Although the present findings have implications for the way in which Lindquist et al.'s. (2006) results are interpreted, it is also important to consider how both sets of findings fit with those reported by Gendron et al. (2012). These researchers continued to find evidence of an emotion-specific massed repetition decrement when using a modified semantic satiation paradigm, which included a differently operationalised set of baseline conditions (see further discussion of Gendron et al., 2012 in Chapter 5). Crucially, Gendron

et al.'s. (2012) experiments differed from a standard 'immediate repetition' design in several other ways. In particular, they used a critical decision that was based on an implicit index of repetition priming, rather than the explicit categorisation of emotional faces. Below, in section 8.4, we describe how differences in task demands (i.e. the critical decision) may have driven a divergent pattern of findings across the experiments reported here, and those in Gendron et al. (2012). This perhaps occurs as task demands bias the types of language-mediated conceptual knowledge that participants bring to the task.

The label-feedback account may provide an explanation for a second consistent finding that arose in the semantic satiation work: the relative facilitation found in the judgment-relevant, expression-match trials (i.e. trials in which the repeated emotion label and pair of expressions all match each another; experiments 1, 2a and 2b). If label exposure creates a relevant perceptual expectancy then participants may experience facilitated processing of both expressions within the pair. A perceptual warping process is implicated; relevant labelling accentuates and draws attention toward the category-diagnostic features that the faces share, and abstracts across those that they do not, temporarily making the pair of faces seem more perceptually similar and facilitating a 'match' decision (e.g. Lupyan, 2007, 2012; Lupyan & Spivey, 2010b).

However, two findings weaken the applicability of this explanation (see Chapter 5 for further discussion). The first consideration is that the label-feedback account might also predict similar facilitation in performance when participants repeated an emotion label that matched only one of the emotional expressions shown within the pair. This would depend on a perceptual warping mechanism; emotional expressions that match the repeated label would be temporarily perceived as closer to a category prototype and, those that mismatched the label, more disparate (e.g. Lupyan, 2008b; Lupyan & Spivey, 2010b though see Lupyan, 2009; Lupyan & Thompson-Schill, 2012). As such, perceptual warping should help the participant to respectively accept one expression as providing a match for the repeated label, and reject the other, facilitating an overall *mismatch* decision. However, common facilitation in judgment-relevant, expression match *and* mismatch trials were not found in any experiment that included these two conditions. The second consideration is that participants in experiment 1 showed similar, relative facilitation in performance when they encountered trials that presented a decision-level analogue (i.e. when all three stimuli mismatched each other; repeat the emotion label 'Sad', followed by the presentation of an



'Angry' and 'Fearful' expression pair). Parallel facilitation may instead support a *strategic* explanation (e.g. Roberson & Davidoff, 2000). Here the outcome of two respective comparisons, between the repeated label and the two expressions within each pair, would heuristically bias responses to the crucial face-to-face comparison that followed. By this account language labels do not act in a top-down manner to directly influence the way expressions are perceived. Instead we may use categorical labels to supplement (and/or replace) the visual codes we generate for stimuli; language-based codes may be highly useful, dependent on the particular demands of the task (e.g. Pylyshyn, 1999; Roberson & Davidoff, 2000; Kikutani et al., 2008). Other key findings within the threshold work, described below, may also support an account based on the *strategic* use of language labels. The possible application of this theoretical account is fully discussed in section 8.3.

### **8.2.2 Strand 2: The threshold work (Chapter 6)**

When considering the threshold work (experiment 5), the label-feedback and construction accounts make similar predictions concerning the presence of cue-relevance effects; that participants should be faster/more accurate to decide that a 'sad' face displays an emotional expression following repetition of a relevant emotion word (vs. an irrelevant emotion word, or neutral word). First, according to the construction position, words activate emotion-relevant, semantic knowledge, therefore repeating a relevant emotion word should facilitate emotionality decisions, perhaps by constraining perception of that expression to be in line with the repeated word (e.g. Herbert et al., 2013; Nook et al., 2015). Arguably exposure to both emotion labels and verbs should have the potential to activate relevant semantic aspects of conceptual knowledge, though perhaps to a different extent. Indeed, verbs constitute direct referents for the behaviours we observe as emotional in everyday life and should therefore occupy ground in our repository of knowledge (e.g. Fehr & Russell, 1984; Shaver et al., 1987; Russell, 1991a). In contrast, the label-feedback account would likely predict cue-relevance effects to only, or more strongly, emerge following exposure to relevant emotion labels vs. verbs, particularly when a short ISI is imposed between cue and target exposure (e.g. Lupyan & Thompson-Schill, 2012; Edmiston & Lupyan, 2015). This would reflect the assumed privileged links that form between category-labels and perceptual forms of conceptual knowledge about a referent object (e.g. Lupyan & Thompson-Schill, 2012; Edmiston & Lupyan, 2015; Lupyan & Casasanto, 2015).

The present work found no consistent cue-relevance effects. Participants were no faster or more accurate to infer that a 'sad' expression was emotional following exposure to relevant emotion words (labels and verbs), in comparison to irrelevant emotion words or neutral words. Cue-relevance effects were absent at both a short and long ISI. However, it is important to note that this study was specifically designed to assess whether language influences the early processing of emotional expression. As such, null findings may suggest that language does not robustly exert its influence here, but at a later processing stage. For example, language may contribute towards an inferential decision process, where evidence is drawn from a number of different sources, external to the face. This interpretation would be broadly consistent with ideas expressed by proponents of the construction approach (see further discussion in sections 1.3.1 and 6.4 e.g. Nook et al., 2015; though see Gendron et al., 2012). However, replication is likely necessary to explore this possibility.

An alternative explanation is also available for the lack of cue-relevance effects. The atypical presentation of cue words in experiment 5 may have reduced the potential for cue consolidation and subsequent conceptual knowledge activation (e.g. Chen & Spence, 2011; Lupyan & Thompson-Schill, 2012; but see Lupyan & Swingley, 2011). In a possible replication attempt, simple modifications are available to rectify this issue. For example, cues words could be presented aurally, as is typical in previous work, which may lead to faster conceptual activation by reducing the need for lexical translation of the cue word (e.g. Chen & Spence, 2010, 2011; but see Lupyan & Swingley, 2012). Alternatively, consolidated processing of the cue word could be achieved by presenting the word more than once, before a face is shown (e.g. Iordanescu et al., 2008, 2011). Indeed, this second idea would address an important methodological difference between the semantic satiation (experiments 1-4) and threshold work (experiment 5). If facilitation in the judgment-relevant, expression match trials of the semantic satiation work constitutes a 'cue-relevance' effect, then this may support the notion that multiple repetitions of the relevant word cue are required for consolidation and conceptual activation. Though, of course, while the difference in cue-relevance effects across strands of the present work may relate to the number of times a relevant cue is presented, they may also reflect the ways in which cue words are used to inform different critical decisions (i.e. categorical discrimination vs. perception).

### 8.3 Using language labels strategically to inform categorical decisions

The label-feedback account suggests that language labels aid (categorical) perception of referent objects by *directly* modifying our ongoing visual experience (e.g. through perceptual warping processes, Lupyan, 2007, 2012). The construction account may be interpreted to extend a similar claim. Whether label-activated semantic knowledge constrains our perception of an expression (e.g. Herbert et al., 2013; Nook et al., 2015), or helps us to form a new percept of the face, independent of the ambiguous perceptual information contained therein (e.g. Fox et al., 2009), semantic information somehow shapes what we see. As such, both accounts fundamentally accept the notion that higher-level knowledge can have a top-down influence on a low-level visual processing (e.g. Collins & Olson, 2014). However, not all researchers accept that vision is ‘cognitively penetrable’, particularly in early processing stages (e.g. Pylyshyn, 1999). By this view, conceptual knowledge allows us to label or identify an already-perceived stimulus, but the application of a label neither influences the visual processing that arises prior, or subsequent, to the application of that label. Arguably this account would be consistent with failures to find cue-relevance effects in the threshold work; a preceding label should not have an active influence on the way we initially perceive an emotional expression.

However, arguing that labelling a stimulus does not affect its perception does not necessarily preclude the *strategic* benefits that labelling can have, dependent on the demands posed by a particular task. These strategic benefits have been documented in categorical perception work (e.g. Roberson & Davidoff, 2000; Pilling et al., 2003; Kikutani et al., 2008). In brief, when attempting to discriminate between visually-presented categorical stimuli participants likely engage in an automatic labelling process. When applied, categorical/verbal codes may supplement visual codes, providing a second, more durable, form of evidence for the discrimination process. Indeed, automatic application and use of verbal labels have been implicated in the oft-reported ‘cross-category’ advantage (e.g. participants are likely faster to correctly decide that two stimuli present a categorical mismatch vs. match as stimuli both look different, and can be labelled differently, from one another; Roberson & Davidoff, 2000; Roberson et al., 2007; Gilbert et al., 2008; Kikutani et al., 2008).

Labels may have been used strategically in the semantic satiation work. In experiment 1, participants showed facilitated performance in two types of trial; those in which the two possible label-to-face comparisons and single face-to-face comparison all provided a congruent outcome (all ‘matches’ or ‘mismatches’, respectively). Due to the restricted time window provided for face-to-face comparisons, participants may have relied on the initial label-to-face comparisons to make their final decisions, foregoing final perceptual analysis and comparison of the face pair (e.g. ‘if the emotion label that I have repeated mismatches both expressions then it is likely that the two expressions themselves mismatch’). In these two trial types, feedback from initial label-to-face comparisons are sufficient to drive accurate final discriminations.

Similarly, there may be indications of strategic label use in the threshold work (experiment 5). The response bias data showed that participants were more likely to adopt a conservative response bias, judging all expressions as ‘non-emotional’ following exposure to an irrelevant emotion word (both labels and verbs e.g. ‘Surprise’ and ‘Gasp’) than a relevant word (‘Sad’ and ‘Sob’). This effect may suggest that participants were able to form a task-wide strategy, based on the relationships between the emotion words and emotional expressions that they repeatedly encountered. Expressions of ‘surprise’ were never shown during the threshold task (only ‘sad’ expressions were used); when participants encountered the label ‘Surprise’ then (or, to a lesser extent, the emotion-related verb, ‘Gasp’) they may have been primed to respond ‘non-emotional’, independent of the expression shown on the test face.

#### **8.4 Task demands and the selective activation of conceptual knowledge**

The conceptual knowledge we hold about any referent is likely multi-faceted, containing both semantic and perceptual elements (the meaningful information we hold about an object’s potential purpose/use, as well as the features that signal membership to a particular category e.g. Lupyan, 2009). Researchers suggest that conceptual knowledge activation, via language labels, is a flexible and selective process, and will be largely determined by inherent task demands (e.g. Barsalou et al., 2003; James & Gauthier, 2004; Casasanto & Lupyan, 2015). The work presented in this thesis suggests that researchers should be mindful of this possibility when designing future work. Below, tentative

arguments are put forward to suggest how the task demands posed both in the semantic satiation and threshold paradigms may have resulted in preferential activation of perceptual forms of conceptual knowledge. These arguments must be tentative for two, interlinked reasons. First, when participants bring pre-existing knowledge to the task there are difficulties inherent in delineating between the contribution of different facets of that knowledge to performance (James & Gauthier, 2003, i.e. people accrue perceptual and semantic information about emotional states through their personal, affective experiences). Second, delineation is further problematic when the same behavioural outcomes are predicted, independent of the type of knowledge activated (i.e. cue-relevance effects in experiment 5, described in Chapter 6). For this second reason, it is recommended that future research pose divergent predictions, dependent on possible *conceptual* knowledge activation.

Following Lindquist et al. (2006), the semantic satiation work used a task in which participants discriminated between a pair of facial expressions. Previous immediate repetition paradigms have employed paired discrimination tasks with word stimuli (e.g. 'King' and 'Royalty'; Smith & Klein, 1990). Researchers suggest that judging the relationship between a pair of words should place heightened demands on the semantic system, in comparison to simply judging whether a single target word is meaningfully associated with the repeated cue word (e.g. an earlier version of the immediate repetition paradigm e.g. Smith, 1984; Smith & Klein, 1990; Balota & Black, 1997). In addition, when a single cue word is presented in the semantic satiation paradigm, performance decrements may reflect inability to re-access the repeated word itself, rather than the semantic knowledge base that potentially connects the cue and target words (Esposito & Pelton, 1971; Smith & Klein, 1990; Balota & Black, 1997; see Lindquist et al., 2006, experiment 1, for a similar argument). However, it is unclear to what extent these arguments apply when a pair of pictures, rather than words, are presented for discrimination.

Indeed, it seems likely that participants engage in an additional translation process when they attempt to discriminate between a pair of pictures vs. words. When we view a pair of pictures we likely engage in automatic, covert labelling to translate visual information into a categorical code; this process is unnecessary when word stimuli are used (e.g. Roberson & Davidoff, 2000; Roberson et al., 2007; Gilbert et al., 2008; Kikutani et al., 2008).

Acknowledging this difference, various researchers have suggested that the massed

repetition decrements originally reported in Lindquist et al.'s. (2006) work might reflect covert label inaccessibility vs. semantic inaccessibility (e.g. Gendron et al., 2012; Lindquist et al., 2014). However, results from the present work are the first to demonstrate the limitations of this alternate explanation (see Chapter 4). Participants continued to show evidence of a massed repetition decrement when they engaged in 30 (vs. 3) repetitions of two non-linguistic activities (foot tapping and passive watching), which should neither cause semantic, nor label inaccessibility.

Although present findings may not be consistent with the use of a *covert* labelling strategy, there was strong evidence that participants attended to *overtly* repeated labels, when congruent with task goals (e.g. Faulkner et al., 2002). This was best exemplified by the consistent finding of facilitated performance in trials where the repeated emotion label matched both expressions at test (experiments 1, 2a and 2b). As described above, two possible explanations are available for this pattern of findings; first, that labels were used *strategically* to meet the demands of the discrimination task, irrespective of the perceptual information contained within the expressions shown (e.g. Pylyshyn, 1999; Roberson & Davidoff, 2000), or second, that repeated labels *directly* helped to accentuate and draw attention towards the category-diagnostic features, contained within the presented expressions (e.g. Lupyan, 2007, 2012). In either case, failures to find an emotion-specific massed repetition decrement in the present work may simply suggest that semantic knowledge does not support emotion perception as it was operationalised here (i.e. in a simultaneous categorical discrimination task). As a consequence the present findings do not definitively rule out the possibility that semantic knowledge is important for the interpretation of emotional expression in other tasks, and settings (e.g. Barrett et al., 2007; Gendron et al., 2012).

The threshold work (experiment 5) in this thesis was specifically designed to assess the contribution of language (and linked conceptual knowledge) to the *early* perception of emotional expressions, rather than later, discrete categorisation. To ensure that this aim was met, several methodological aspects were adopted from Gendron et al. (2012) and related object recognition paradigms (e.g. Lupyan & Spivey, 2010a; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013). For example, participants saw emotional faces for very brief durations and made a simple 'emotionality' decision, which should not require the

explicit retrieval of a language label to categorise the expression displayed on the face (e.g. Russell & Barrett, 1999).

However, it may be the case that some of the very same methodological aspects that allow monitoring of the early effects of language on emotion perception also encourage preferential activation of perceptual forms of conceptual knowledge, when participants encounter an emotion word. First, researchers suggest that brief presentation of categorical stimuli may generally encourage a perceptual processing strategy (Marsolek et al., 2006). Second, although accurate emotionality decisions could be reached based on a consideration of basic affective properties of the face (i.e. valence and arousal, Russell & Barrett, 1999), rather than via application of a discrete label to the expression, researchers argue that emotionality decisions may still be broadly ‘categorical’ in the way that they encourage participants to consider stimuli (an expression can either be categorised as ‘emotional’ or ‘non-emotional/neutral’ e.g. Roesch et al., 2010; Schurgin et al., 2014). Similar criticisms have been applied when emotionality decisions are used to explore the category-diagnostic features that participants associate with different discrete expressions (e.g. Smith & Gosselin, 2002; Schurgin et al., 2014). Similar to the threshold work, participants here must decide whether a presented face is emotional or neutral in expression. It is argued that this task may then encourage strategic and unnatural reliance on a perceptual processing strategy that highlights the particular features important for differentiating neutral expressions from select emotional expressions (e.g. Smith & Gosselin, 2002; Smith et al., 2005; Schurgin et al., 2014).

Arguing that the demands of the threshold task lead to preferential label-mediated activation of perceptual knowledge does not yet explain the lack of cue-relevance effects reported in this work. Indeed, the label-feedback account would predict cue-relevance effects; after hearing a relevant emotion label the category-diagnostic features of the subsequently presented expression would be accentuated, easing the ability to identify that face as emotional (e.g. Lupyan & Spivey, 2010a; Lupyan & Thompson-Schill, 2012). However, as discussed in Chapter 6, it is possible that we sampled sub-optimal stimuli in experiment 5. Based on piloting work the decision was made to use only male expressions of sadness; it is argued that expression of sadness may include fewer, and less distinctive, category-diagnostic features than those associated with other emotional states (e.g. Calvo & Lundqvist, 2008; Goeleven et al., 2008; Sánchez & Vázquez, 2013). Given the importance

that the label-feedback approach places on the typicality assumption (e.g. Lupyan, 2008a; Lupyan & Thompson-Schill, 2012), it is perhaps unsurprising that cue-relevance effects were absent; activated perceptual knowledge cannot accentuate category-diagnostic features if few exist, particularly if these features overlap with those present in their neutral counterparts (e.g. Sánchez & Vázquez, 2013). As discussed in Chapter 6, future researchers may vary the typicality of stimuli to better assess the application of the label-feedback account to the case of emotional expressions.

## **8.5 Modifying the threshold paradigm to assess the time-course of language-mediated effects on emotion perception**

As previously reported, findings from the threshold work may have indirect implications for the construction account (see Chapter 6). Nook et al. (2015) report facilitative label-relevance effects in a task that used similar task components, but presented them in a reversed order. Here participants first saw an emotional face, followed by an emotion label, and decided whether the emotional expression and label ‘matched’ or ‘mismatched’ one another. Facilitative effects were weaker when the label was replaced with a second face, and participants judged whether the two faces matched in emotional expression. Nook et al. (2015) interpreted these findings to suggest that participants initially generate emotion-related category predictions when they view the first emotional face, but that these predictions may be weak due to the perceptual ambiguities inherent within emotional expressions (e.g. Barrett et al., 2007; Mauss & Robinson, 2009). When a second face is shown it is unlikely to enhance the participant’s confidence in their categorisation, even when it presents a matching expression, as it offers the same perceptually ambiguous information. In contrast, label-activated semantic knowledge may constrain the range of possible categorisation options generated in response to the initial face, allowing us to abstract across any perceptual ambiguities to make our decision (e.g. Carroll & Russell, 1996; Halberstadt & Niedenthal, 2001; Halberstadt, 2005).

This conceptualisation supports some construction ideas about the possible way in which semantic knowledge influences emotion perception (e.g. Herbert et al., 2013). In contrast, the label-feedback approach (Lupyan, 2007, 2012) may only predict cue-relevance effects either when the relevant word precedes, or is presented concurrent, with the emotional face; hence the decision to adopt this task format in previous object perception work (e.g.



Lupyan & Spivey, 2010a; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013). Arguably it is unclear how language labels would temporarily accentuate/draw attention toward category-diagnostic features if they were presented after the visual display.

Crucially, the comparison between Nook et al.'s. (2015) findings and those presented in the current threshold work offer an interesting avenue for future research. Varying the temporal placement, or proximity, of the cue word in relation to face stimuli in the threshold paradigm would both tell us more about the time-course across which emotional language has an impact on emotion 'perception' and would likely have theoretical implications for the accounts described in this thesis. The temporal proximity of verbal cues, in relation to visual targets, have previously been manipulated in similar object recognition work (e.g. Spence, Shore & Klein, 2001; Roelofs, 2005; Chen & Spence, 2010; see further discussion in Chapter 6).

## **8.6 Future strategies to assess the influence of semantic knowledge on emotion perception**

When assessing the possible contribution of semantic knowledge to emotion perception, advocates of the construction approach have mainly focused on behavioural outcomes. For example, the differences in accuracy and reaction time when participants respond to emotional stimuli after access to semantic knowledge has been manipulated (e.g. Lindquist et al., 2006, 2014; Gendron et al., 2012). However, alternative methods of assessment may prove more effective (e.g. Abdel Rahman & Sommer, 2012). In training paradigms, where participants are taught to associate biographical, semantic and/or affective knowledge with novel faces, differences in response to trained and untrained stimuli often emerge in physiological outcomes (e.g. event-related potentials, Galli et al., 2006; Heisz & Shedden, 2009; Herzmann & Sommer, 2010; Abdel-Rahman, 2011; Abdel-Rahman & Sommer, 2012, see also semantic knowledge benefits in object training paradigms e.g. James & Gauthier, 2004; Collins & Curby, 2013). Findings frequently suggest that trained vs. untrained unfamiliar faces are associated with enhanced early amplitudes, indicative of efficient structural processing, for example, the P100 (Abdel Rahman & Sommer, 2012), the N170 (Galli et al., 2006; Heisz & Shedden, 2009; Herzmann & Sommer, 2010, but see Abdel Rahman & Sommer, 2012) and the Early Posterior Negative/Late Positive Potential (Abdel

Rahman, 2011). Crucially, these differences are also often accompanied by activations indicative of knowledge access (N400), despite no requirement for explicit knowledge retrieval during response to the stimuli (e.g. Abdel Rahman & Sommer, 2012). In future work then, physiological measures may be used to assess whether emotional faces are processed differently when access to emotion-related semantic knowledge has been manipulated. That previous work documents effects in early potentials may suggest that this type of work could be useful to assess whether initial ‘perception’ of a face is modulated by semantic knowledge activation.

## **8.7 Self-generation and rating of emotion-related verbs: theoretical implications and future directions**

A third strand of work, described in Chapter 7, was driven by the notion that emotion words are grounded in, and have the potential to re-activate, affective knowledge (e.g. Vigliocco et al., 2009; Kousta et al., 2011; Wilson-Mendenhall et al., 2011). A simplified feature-generation task (e.g. Vinson & Vigliocco, 2008; Buchanan et al., 2013) was used to specifically elicit the emotion-related verbs commonly included within the participant’s semantic script. Arguably, generated verbs are the lexical representations for the emotional behaviours we have previously experienced, and have begun to associate with different emotional states (e.g. Russell & Fehr, 1984; Shaver et al., 1987; Russell, 1991a).

Findings from the present task provide some support for the construction account. It was observed that participants generated two types of emotion-related verb; a small number that were disproportionately related to only one emotion label/state (e.g. the verb ‘Vomit’ was only generated by participants in response to the label ‘Disgust’) and a larger number that were generated in response to multiple emotion labels/states, differing only in endorsement strength (e.g. the verb ‘Cry’ was generated by at least two participants in response to five of the six emotion labels). That participants generate the second type of *overlapping* exemplar is of particular interest. Support was found in the subsequent validation study as associative label-to-verb ratings mirrored generation frequencies. When an *overlapping* exemplar (verb) was encountered by raters, their ratings reflected relative endorsement strength. To illustrate, in study 1 the verb ‘Scream’ was both generated by a large number of participants in response to the label ‘Fear’ and a smaller number of

participants in response to the label 'Anger', therefore the associative ratings given by participants in study 2 for the 'Scream/Fear' pair were higher than those that they gave for the 'Scream/Anger' pair. The existence of overlapping exemplars stands in direct contrast to 'basic' or 'discrete' views (e.g. Ekman, 1992, 2003; Ekman & Cordaro, 2011). This approach suggests that the human ability to discretely categorise emotional experience depends on the non-overlapping nature of behaviours and physiological responses, which comprise the hard-wired program for each basic emotional state (e.g. Tomkins, 1984; Ekman & Cordaro, 2011). In contrast, construction (and componential e.g. Scherer, 1984) accounts suggest that we consider information from several different sources when we interpret emotion, meaning that behaviours themselves do not need to be solely diagnostic for categorisation (e.g. Barrett et al., 2007; Scherer & Ellgring, 2007). In addition, both approaches emphasise the importance of variability and context-dependence when considering how we interpret emotion (e.g. Barrett et al., 2011). Specifically, that people may only experience a behaviour as emotional if they have personally experienced or observed that behaviour in an appropriate context (e.g. Somerville & Whalen, 2006), and that the same behaviour might have a very different meaning dependent on the context in which it occurs (e.g. finding out that you have passed or failed a difficult exam might cause you to feel different emotions, whilst prompting the same behaviour e.g. 'crying', Barrett, 2006a, 2009, 2011).

Further work may also be conducted to extend the database of emotion-related verbs reported here. The extent to which these words proportionally evoke discrete emotional knowledge will likely depend on the nationality of the speaker; strength of knowledge activation may only match the reported generation frequencies and ratings for English speakers (e.g. Pavlenko, 2008). That emotion words differ in affective value across languages has prompted cross-cultural translations and extensions of Bradley and Lang's (1999) ANEW database (e.g. Vö et al., 2006, 2009; Redondo, Fraga, Padrón, & Comesaña, 2007; Eilola & Havelka, 2010; Soares, Comesaña, Pinheiro, Simões, & Frade, 2012). It may be particularly interesting to compare the strength of discrete knowledge activation for emotion-related verbs across speakers of British English and the Slavic languages (e.g. Russian, Serbian, Polish). It may be argued that the emotion lexicon is constructed differently for these two sets of speakers (e.g. Pavlenko, 2008; Wierzbicka, 1992, 1994).

When we consider the emotion lexicon as a hierarchy, basic labels should sit atop for English speakers (e.g. 'Anger'). These words are used to directly refer to the emotional states that we *internally* experience as part of our mental lives and should therefore have the strongest potential to activate discrete, emotion-related knowledge (Pavlenko, 2008). In comparison, emotion-related words (verbs) and emotion-laden words (commonly nouns) are connotative in their relation to discrete emotional states and therefore have weaker potential to evoke knowledge (e.g. Altarriba, 2006; Pavlenko, 2008; Altarriba & Basnight-Brown, 2010; Knickerbocker & Altarriba, 2013). In contrast, speakers of Slavic languages view emotional states as dynamic and experienced as part of an *external*, and interpersonal environment. As a consequence, these individuals frequently use emotion-related words, or verbs, to communicate that they are experiencing a particular emotion (Wierzbicka, 1992, 1994; Pavlenko, 2008). As such, emotion-related words, or verbs, may sit atop the lexical hierarchy for these speakers and have a comparatively larger potential to activate discrete, affective knowledge. In a self-generation task then, we speculate that Slavic speakers will show more frequent endorsement of a smaller set of emotion-related verbs. This will reflect the strong mappings between emotion-related verbs and emotional states for these speakers.

## 8.8 Potential applications

The ability to 'correctly' interpret the emotional state that another is feeling is a critical social task, determining the way in which we respond towards that individual in the immediate context (Mauss & Robinson, 2009) and possibly future contexts, too (e.g. Barrett, 2006a, 2009; Barrett et al., 2007). The work included in this thesis joins a body of research that may prompt exploration into why specific subsets of the population experience difficulties when attempting to interpret emotional expression. In common, these populations may show deficits when attempting to access and apply discrete category labels to visual stimuli (e.g. Roberson et al., 1999; Lupyan, 2009; Lupyan & Mirman, 2013; but see Sauter et al., 2011) and/or when attempting to access emotion-related conceptual representations (e.g. Lindquist et al., 2006; Gendron et al., 2012). The latter problem may arise if that knowledge base is still under construction, as may be the case in children (i.e. Russell & Widen, 2002; Widen & Russell, 2008; Widen, 2013). Alternatively, conceptual representations may have become detached from their language-based anchors, preventing future access to that knowledge, as is the case in patients with semantic dementia (e.g. Calabria et al., 2009; Lindquist et al., 2014). Recent

research extends these arguments to those experiencing alexithymia, defined as an inability to spontaneously label and describe self and other-relevant emotional states (e.g. Grynberg et al., 2012; Cook, Brewer, Shah & Bird, 2013). Importantly, work conducted by Nook et al. (2015) offers the first indication that explicitly presenting emotion-relevant labels and concepts within a task can alleviate the emotion recognition deficits associated with spontaneous retrieval of those elements. As such, future work may explore whether the emotion recognition deficits, present in populations identified above, can be attenuated with explicit training in the development and application of emotion labels and concepts (see Golan & Baron-Cohen, 2006; Nook et al., 2015).

## References

- Abdel Rahman, R. (2011). Facing good and evil: early brain signatures of affective biographical knowledge in face recognition. *Emotion, 11*(6), 1397–1405.
- Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review, 15*(6), 1055–1063.
- Abdel Rahman, R. & Sommer, W. (2012). Knowledge scale effects in face recognition: An electrophysiological investigation. *Cognitive, Affective, & Behavioral Neuroscience, 12*(1), 161–174.
- Altarriba, J. (2006). Cognitive approaches to the study of emotion-laden and emotion words in monolingual and bilingual memory. In Pavlenko, A. (ed.), *Bilingual Minds: Emotional experience, expression and representation* (pp. 232–256). Clevedon, Buffalo, Toronto: Multilingual Matters Ltd.
- Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology, 117*(3), 389–410.
- Altarriba, J., & Basnight–Brown, D. M. (2010). The representation of emotion vs. emotion-laden words in English and Spanish in the Affective Simon Task. *International Journal of Bilingualism, 15*, 310–328.
- Anderson, E., Siegel, E. H., Bliss–Moreau, E., & Barrett, L. F. (2011) The visual impact of gossip. *Science, 332* (6036), 1446–1448.
- Augustyn, C., & Peters, M. (1986). On the relation between footedness and handedness. *Perceptual and Motor Skills, 63*(3), 1115–1118.
- Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., ... Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19*(7), 724–732.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: evidence from task switching. *Journal of Experimental Psychology: General, 130*(4), 641–657.
- Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology, 33*(4), 439–454.
- Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36*(2), 233–252.

- Balota, D. A., & Black, S. (1997). Semantic satiation in healthy young and older adults. *Memory & Cognition*, *25*(2), 190–202.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.
- Barrett, L. F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science*, *1* (1), 28–58.
- Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10* (1), 20–46.
- Barrett, L. F. (2009). Variety is the spice of life: A psychological constructionist approach to understanding variability in emotion. *Cognition and Emotion*, *23* (7), 1284–1306.
- Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, *20*(6), 400–406.
- Barrett, L. F. (2012). Emotions are real. *Emotion*, *12* (3), 413–429.
- Barrett, L. F. (2013). Psychological construction: A Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389.
- Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, *21*(4), 595–599.
- Barrett, L.F., Gendron, M., Huang, Y–M. (2009). Do discrete emotions exist? *Philosophical Psychology*, *22*(4), 427–437.
- Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332.
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York: Cambridge University Press.
- Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *The Quarterly Journal of Experimental Psychology*, *66*(11), 2135–2160.

- Black, S. R. (2001). Semantic satiation and lexical ambiguity resolution. *The American Journal of Psychology*, *114*(4), 493–510.
- Black, S. R., Leonard, R. C., Willis, S., Burton, P., McCown, S., Lookadoo, R., & Wayde, E. (2013). Effects of homograph meaning frequency on semantic satiation. *Canadian Journal of Experimental Psychology*, *67*(3), 175–187.
- Blaxton, T. A., & Neely, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition*, *11*(5), 500–510.
- Bornemann, B., Winkielman, P., & van der Meer, E. (2012). Can you feel what you do not see? Using internal feedback to detect briefly presented emotional stimuli. *International Journal of Psychophysiology*, *85*(1), 116–124.
- Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology*, *99*(2), 167–189.
- Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion*, *22*(8), 1569–1587.
- Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Technical Report C–1) The Center for Research in Psychophysiology, Gainesville: University of Florida.
- Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011a). Discrete emotion norms for nouns: Berlin affective word list (DENN–BAWL). *Behavior Research Methods*, *43*(2), 441–448.
- Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011b). Discrete emotion effects on lexical decision response times. *PloS ONE*, *6*(8), e23743. doi: 10.1371/journal.pone.0023743
- Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later? *Brain Research*, *1564*, 62–71.
- Briesemeister, B. B., Kuchinke, L., Jacobs, A. M., & Braun, M. (2014). Emotions in reading: Dissociation of happiness and positivity. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 287–298.
- Brown, A. S., Zoccoli, S. L., & Leahy, M. M. (2005). Cumulating retrieval inhibition in semantic and lexical domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 496–507.
- Buchanan, E. M., Holmes, J. L., Teasley, M. L., & Hutchison, K. A. (2013). English semantic word-pair norms and a searchable Web portal for experimental stimulus creation. *Behavior Research Methods*, *45*(3), 746–757.



- Calabria, M., Cotelli, M., Adenzato, M., Zanetti, O., & Miniussi, C. (2009). Empathy and emotion recognition in semantic dementia: A case report. *Brain and Cognition*, *70*(3), 247–252.
- Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 527–551.
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*(2), 81–118.
- Calvo, M. G., Fernández–Martín, A., & Nummenmaa, L. (2012). Perceptual, categorical, and affective processing of ambiguous smiling facial expressions. *Cognition*, *125*(3), 373–393.
- Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display–duration conditions. *Behavior Research Methods*, *40*(1), 109–115.
- Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition and Emotion*, *23*(4), 782–806.
- Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, *137*(3), 471–494.
- Cambridge Dictionaries Online. (2015). Retrieved from:  
<http://www.dictionary.cambridge.org/>
- Camos, V., & Tillman, B. (2008). Discontinuity in the enumeration of sequentially presented auditory and visual stimuli, *Cognition*, *107*(3), 1135–1143.
- Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*(2), 205–218.
- Casasanto, D. & Lupyan, G. (2015). All Concepts are Ad Hoc Concepts. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New Directions in the study of concepts* (pp. 543–566). Cambridge: MIT Press.
- Chapman, J. P., Chapman, L. J., & Allen, J. J. (1987). The measurement of foot preference. *Neuropsychologia*, *25*(3), 579–584.
- Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404.

- Chen, Y. C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1554–1568.
- Cohene, L. S., Smith, M. C., & Klein, D. (1978). Semantic satiation revisited with a lexical decision task. *Memory & Cognition*, *6*(2), 131–140.
- Collins, J. A., & Curby, K. M. (2013). Conceptual knowledge attenuates viewpoint dependency in visual object recognition. *Visual Cognition*, *21*(8), 945–960.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.
- Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*(4), 843–860.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.
- Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*, *24*(5), 723–732.
- Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*, 92–107.
- Curby, K. M., Hayward, G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective & Behavioral Neuroscience*, *4*(1), 100–111.
- Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1372–1384.
- Damjanovic, L., Roberson, D., Athanasopoulos, P., Kasai, C., & Dyson, M. (2010). Searching for happiness across cultures. *Journal of Cognition and Culture*, *10*(1), 85–107.
- Dilkina, K., McClelland, J. L., & Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Research*, *1365*, 66–81.
- Doost, H. T. N., Moradi, A. R., Taghavi, M. R., Yule, W., & Dalgleish, T. (1999). The development of a corpus of emotional words produced by children and adolescents. *Personality and Individual Differences*, *27*(3), 433–451.
- Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*(3), 286–303.

- Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100.
- Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cerebral Cortex*, *17*(9), 2123–2133.
- Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, *42*(1), 134–140.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3/4), 169–200.
- Ekman, P. (2003). *Emotions revealed*. New York, NY: Holt Paperbacks.
- Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.
- Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. New Jersey, NJ: Prentice Hall.
- Ekman, P. and Friesen, W. (1976). *Pictures of Facial Affect*. Palo Alto, California: Consulting Psychologists Press.
- Ekman, P., and Friesen, W.V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, California: Consulting Psychologists Press.
- Elfenbein, H.A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235.
- Ellison, J. W., & Massaro, D. W. (1997). Featural evaluation, integration, and judgment of facial affect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 213–226.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582.
- Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*(1), 148–168.
- English, M. C., & Visser, T. A. (2014). Exploring the repetition paradox: The effects of learning context and massed repetition on memory. *Psychonomic Bulletin & Review*, *21*(4), 1026–1032.

- Esposito, N. J., & Pelton, L. H. (1971). Review of the measurement of semantic satiation. *Psychological Bulletin*, *75*(5), 330–346.
- Esterman, M., & Yantis, S. (2009). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253.
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, *8*(4), 453–457.
- Esteves, F., & Öhman, A. (1993). Masking the face: recognition of emotional facial expressions as a function of the parameters of backward masking. *Scandinavian Journal of Psychology*, *34*(1), 1–18.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240.
- Fabes, R. A., & Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin*, *17*(5), 532–540.
- Faulkner, T. F., Rhodes, G., Palermo, R., Pellicano, E., & Ferguson, D. (2002). Recognizing the un-real McCoy: Priming and the modularity of face recognition. *Psychonomic Bulletin & Review*, *9*(2), 327–334.
- Fearing, M. K., Browning, C. A., Corey, D. M., & Foundas, A. L. (2001). Dual-task performance in right- and left-handed adults: A finger-tapping and foot-tapping study. *Perceptual and Motor Skills*, *92*(2), 323–334.
- Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464–486.
- Fernández-Dols, J. M., Carrera, P., Barchard, K. A., & Gacitua, M. (2008). False recognition of facial expressions of emotion: causes and implications. *Emotion*, *8*(4), 530–539.
- Fernández-Dols, J. M., Wallbott, H., & Sanchez, F. (1991). Emotion category accessibility and the decoding of emotion from facial expression and context. *Journal of Nonverbal Behavior*, *15*(2), 107–123.
- Fetterman, A. K., Robinson, M. D., & Meier, B. P. (2012). Anger as “seeing red”: evidence for a perceptual association. *Cognition and Emotion*, *26*(8), 1445–1458.
- Fiorentini, C., & Viviani, P. (2009). Perceiving facial expressions. *Visual Cognition*, *17*(3), 373–411.
- Froni, F., & Semin, G. R. (2009). Language that puts you in touch with your bodily feelings: The multimodal responsiveness of affective expressions. *Psychological Science*, *20*(8), 974–980.

- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124.
- Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. (2009). The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. *Neuroimage*, *44*(2), 569–580.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton–Mifflin.
- Frank, M. G., & Stennett, J. (2001). The forced–choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, *80*(1), 75–85
- Frenck–Mestre, C., Besson, M., & Pynte, J. (1997). Finding the locus of semantic satiation: An electrophysiological attempt. *Brain and Language*, *57*(3), 406–422.
- Friedman, D., Cycowicz, Y. M., & Dziobek, I. (2003). Cross–form conceptual relations between sounds and words: Effects on the novelty P3. *Cognitive Brain Research*, *18*(1), 58–64.
- Frijda, N. H. (1986). *The Emotions*. Cambridge, UK: Cambridge University Press.
- Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2010). Reading Chimpanzee faces: A test of the structural and conceptual hypotheses. *Emotion*, *10* (4), 544–554.
- Galli, G., Feurra, M., & Viggiano, M. P. (2006). “Did you see him in the newspaper?” Electrophysiological correlates of context and valence in face processing. *Brain Research*, *1119*, 190–202.
- Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, *20*(3–6), 507–523.
- Gendron, M., Lindquist, K., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*(2), 314–325.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*(2), 251–262.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98.
- GNU Image Manipulation Software (GIMP, Version 2.8) [Computer Software]. Retrieved from: <http://www.gimp.org>

- Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*(1), 13–42.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32–41.
- Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, *22*(4), 1628–1635.
- Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: a validation study. *Cognition and Emotion*, *22*(6), 1094–1118.
- Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology*, *18*(2), 591–617.
- Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.
- Gosselin, F., & Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.
- Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): systematic review, unanswered questions and further perspectives. *PLOS one*, *7*(8), e42429.
- Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion*, *13*(3), 225–266.
- Halberstadt, J. (2005). Featural shift in explanation-biased memory for emotional faces. *Journal of Personality and Social Psychology*, *88*(1), 38–49.
- Halberstadt, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology*, *81*(4), 587–598.
- Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional conception how embodied emotion concepts guide perception and facial action. *Psychological Science*, *20*(10), 1254–1261.
- Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review*, *18*(2), 355–363.

- Harnad, S. (1987). Introduction: psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.). *Categorical perception: The groundwork of cognition*, New York: Cambridge University Press.
- Hauk, O., I. Johnsrude & F. Pulvermüller. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301–307
- Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, 14(3), 436–441.
- Heisz, J. J., & Shedden, J. M. (2009). Semantic learning modifies perceptual face processing. *Journal of Cognitive Neuroscience*, 21(6), 1127–1134.
- Herbert, C., Junghöfer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, 45(3), 487–498.
- Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology*, 43(2), 197–206.
- Herbert, C., Sfarlea, A., & Blumenthal, T. (2013). Your emotion or mine: labelling feelings alters emotional face perception - an ERP study on automatic and intentional affect labelling. *Frontiers in Human Neuroscience*, 7, 378.
- Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39(1), 3–36.
- Herzmann, G., & Sommer, W. (2010). Effects of previous experience and associated knowledge on retrieval processes of faces: An ERP investigation of newly learned faces. *Brain Research*, 1356, 54–72.
- Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision*, 9(19), 1–8.
- Hess, U., Adams, R. B., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion*, 4(4), 378–388.
- Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, 21(4), 241–257.
- Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, 24(4), 265–283.

- Higgins, J. A., & Johnson, M. K. (2013). Lost thoughts: Implicit semantic interference impairs reflective access to currently active information. *Journal of Experimental Psychology: General*, *142*(1), 6–11.
- Holle, C., Neely, J. H., & Heimberg, R. G. (1997). The effects of blocked versus random presentation and semantic relatedness of stimulus words on response to a modified Stroop task among social phobics. *Cognitive Therapy and Research*, *21*(6), 681–697.
- Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: Temporal segregation through synaptic depression. *Cognitive Science*, *27*(3), 403–430.
- Hutchison, K. A., Balota, D. A., Cortese, M. J., Neely, J. H., Niemeyer, D., Bengson, J. J., ... Buchanan, E. (2010). The Semantic Priming Project. *Behavior Research Methods*, *45*(4), 1099–1114.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241.
- Inquisitive mind [online recruitment forum]: Retrieved from: <http://www.in-mind.org/content/online-research>
- Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*, *72*(7), 1736–1741.
- Iordanescu, L., Grabowecky, M., & Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta Psychologica*, *137*(2), 252–259.
- Iordanescu, L., Guzman–Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*(3), 548–554.
- Izard, C. E. (1992). Basic emotions, relations amongst emotions and emotion-cognition relations. *Psychological Review*, *99*(3), 561–565.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2), 288–299.
- Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*(18), 1543–1548.
- Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19–25.



- Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241–7244.
- James, T. W., & Gauthier, I. (2003). Auditory and Action Semantic Features Activate Sensory-Specific Perceptual Brain Regions. *Current Biology*, *13*(20), 1792–1796.
- James, T. W., & Gauthier, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Research*, *44*(5), 429–439.
- Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, *3*(2), 81–123.
- Joyce, C. A., Schyns, P. G., Gosselin, F., Cottrell, G. W., & Rossion, B. (2006). Early selection of diagnostic facial information in the human visual cortex. *Vision Research*, *46*(6), 800–813.
- Junghöfer, M., Bradley, M. M., Elbert, T. R., & Lang, P. J. (2001). Fleeting images: a new look at early emotion discrimination. *Psychophysiology*, *38*(2), 175–178.
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, *1148*, 138–148.
- Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying emotions on the basis of neural activation. *PLoS ONE*, *8*(6), e66032. doi:10.1371/journal.pone.0066032
- Kayyal, M. H., & Russell, J. A. (2012). Language and emotion: certain English–Arabic translations are not equivalent. *Journal of Language and Social Psychology*, *32*(2), 261–271.
- Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, *15*(4), 787–794.
- Kissler, J., Herbert, C., Peyk, P., & Junghöfer, M. (2007). Buzzwords—Early cortical responses to emotional words during reading. *Psychological Science*, *18*(6), 475–480.
- Knickerbocker, H., & Altarriba, J. (2013). Differential repetition blindness with emotion and emotion-laden word types. *Visual Cognition*, *21*(5), 599–627.
- Kounios, J., Kotz, S. A., & Holcomb, P. J. (2000). On the locus of the semantic satiation effect: Evidence from event-related brain potentials. *Memory & Cognition*, *28*(8), 1366–1377.

- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34
- Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.
- Kuhl, B. A., & Anderson, M. C. (2011). More is not always better: paradoxical effects of repetition on semantic accessibility. *Psychonomic Bulletin & Review*, *18*(5), 964–972.
- Kuperman, V. (2014). Virtual experiments in megastudies: A case study of language and emotion. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1693–1710.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168.
- Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex.
- Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(8), 1249–1268.
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, *8*(4), 445–452.
- Lewis, M. B., & Ellis, H. D. (1999). Repeated repetition priming in face recognition. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *52*(4), 927–955.
- Lewis, M. B., & Ellis, H. D. (2000a). Satiation in name and face recognition. *Memory & Cognition*, *28*(5), 783–788.
- Lewis, M. B., & Ellis, H. D. (2000b). The effects of massive repetition on speeded recognition of faces. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(4), 1117–1142.

- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265–294.
- Lindquist, K. A. (2009). Language is powerful. *Emotion Review*, *1*(1), 16–18.
- Lindquist, K., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science*, *19*(9), 898–903.
- Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66–71.
- Lindquist, K., Barrett, L. F., Bliss–Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, *6*(1), 125–138.
- Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B.C. (2014). Emotion perception but not affect perception is impaired with semantic memory loss. *Emotion*, *14*(2), 375–387.
- Lindquist, K. A., Siegel, E. H., Quigley, K., & Barrett, L. F. (2013). The hundred years emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, & Bench (2011). *Psychological Bulletin*, *139*(1), 255–263.
- Logie, R. H., & Baddeley, A. D. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 310–326.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, ISBN: 91-630-7164-9.
- Lupyan, G. (2007). *The Label Feedback Hypothesis: Linguistic Influences on Visual Processing* (Doctoral dissertation). Carnegie Mellon University, Pittsburgh, PA.
- Lupyan, G. (2008a). From Chair To “Chair:” A Representational Shift Account Of Object Labeling Effects On Memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369.
- Lupyan, G. (2008b). The Conceptual Grouping Effect: Categories Matter (and named categories matter more). *Cognition*, *108*(2), 566–577
- Lupyan, G. (2009). Extracommunicative Functions of Language: Verbal Interference Causes Selective Categorization Impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718.
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label–feedback hypothesis. *Frontiers in Psychology*, *3*(54), 1–13.
- Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193.

- Lupyan, G., & Mirman, D. (2013). Linking language and categorization: evidence from aphasia. *Cortex*, *49*(5), 1187–1194.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083.
- Lupyan, G. & Spivey, M.J. (2010a). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS ONE*, *5*(7), e11452.  
doi:10.1371/journal.pone.0011452
- Lupyan, G. & Spivey, M.J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253.
- Lupyan, G., Swingley, D., (2011). Self-directed speech affects visual processing. *Quarterly Journal of Experimental Psychology*, *65*(6), 1068–1085.
- Lupyan, G., & Thompson–Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186.
- Lupyan, G., Thompson–Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, *21*(5), 682–691.
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201.
- MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*(3), 474–488.
- Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 21–57). Hillsdale, NJ: Erlbaum.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.
- Mahr, A., & Wentura, D. (2014). Time-compressed spoken word primes cross-modally enhance processing of semantically congruent visual targets. *Attention, Perception, & Psychophysics*, *76*(2), 575–590.

- Marsolek, C. J., Schnyer, D. M., Deason, R. G., Ritchey, M., & Verfaellie, M. (2006). Visual antipriming: evidence for ongoing adjustments of superimposed visual object representations. *Cognitive, Affective & Behavioral Neuroscience, 6*(3), 163–174.
- Martin, R. A., Berry, G. E., Dobrański, T., Horne, M., & Dodgson, P. G. (1996). Emotion perception threshold: individual differences in emotional sensitivity. *Journal of Research in Personality, 30*(2), 290–305.
- Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., Ray, R., & Uchida, H. (2002). American–Japanese cultural differences in judgements of emotional expressions of different intensities. *Cognition and Emotion, 16*(6), 721–747.
- Matsumoto, D., & Hwang, H. C. (2014). Judgments of subtle facial expressions of emotion. *Emotion, 14*(2), 349–357.
- Matsumoto, D., Keltner, D., Shiota, M. N., O’Sullivan, M & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J.M. Haviland-Jones, & L.F. Barrett (Eds.), *The Handbook of Emotion* (3rd ed., pp. 211–234). New York: Guilford Press.
- Matsumoto, D., Olide, A., Schug, J., Willingham, B., & Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. *Journal of Nonverbal Behavior, 33*(4), 213–238.
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209–237.
- Maxwell, J., & Davidson, R. (2004). Unequally masked: Indexing differences in the perceptual salience of "unseen" facial expressions. *Cognition and Emotion, 18*(8), 1009–1026.
- Mayr, E. (1974) Behavior programs and evolutionary strategies. *American Scientist, 62*, 650–659.
- McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *The American Journal of Psychology, 95*(4), 581–634.
- McKenna, F. P. (1986). Effects of unattended emotional stimuli on color–naming performance. *Current Psychology, 5*(1), 3–9.
- McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: the role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 382–392.

- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and non-living things. *Behavior Research Methods, 37*(4), 547–559.
- Milders, M., Sahraie, A., & Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cognition and Emotion, 22*(1), 63–82.
- Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*(2), 123–142.
- Moffat, M., Siakaluk, P. D., Sidhu, D. M., & Pexman, P. M. (2014). Situated conceptualization and semantic processing: effects of emotional experience and context availability in semantic categorization and naming tasks. *Psychonomic Bulletin & Review, 22*(2), 408–419.
- Moseley, R., Carota, F., Hauk, O., Mohr, B., & Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex, 22*(7), 1634–1647.
- Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion, 7*(4), 736–744.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*(3), 226–254.
- Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review, 5*(1), 8–15.
- Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience, 6*(113), 1-15.
- Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology, 96*(6), 1120–1136.
- Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A New Look at Emotion Perception: Concepts Speed and Shape Facial Emotion Recognition. *Emotion, 15*(5), 569–578.
- Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin, 131*(5), 763–784.
- Online Psychology Research [Online recruitment website]: Retrieved from: <http://www.onlinepsychresearch.co.uk>
- Palazova, M., Mantwill, K., Sommer, W., & Schacht, A. (2011). Are effects of emotion in single words non-lexical? Evidence from event-related brain potentials. *Neuropsychologia, 49*(9), 2766–2775.

- Palazova, M., Sommer, W., & Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: An event-related potential study with verbs. *Brain and Language, 125*(3), 264–271.
- Paller, K. A., Gonsalves, B., Grabowecky, M., Bozic, V. S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. *Neuroimage, 11*(2), 98–110.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52*(1), 629–651.
- Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition, 11*(2), 147–164.
- Pecher, D., Boot, I., & van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. Ross (Ed.), *The Psychology of Learning and Motivation (Vol. 54)* (pp. 217–248). Burlington: Academic Press.
- Pell, P. J., & Richards, A. (2011). Cross-emotion facial expression aftereffects. *Vision Research, 51*(17), 1889–1896.
- Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience, 7*(122), 1–4.
- Peters, M., & Durning, B. M. (1979). Footedness of left- and right-handers. *The American Journal of Psychology, 92*(1), 133–142.
- Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. (2003). Is color “categorical perception” really perceptual?. *Memory & Cognition, 31*(4), 538–551.
- Pilotti, M., Antrobus, J. S., & Duff, M. (1997). The effect of presemantic acoustic adaptation on semantic “satiation”. *Memory & Cognition, 25*(3), 305–312.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Attention, Perception, & Psychophysics, 15*(2), 285–290.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik, & H. Kellerman (Eds.). *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3–33). New York: Academic Press.
- Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., & Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion, 12*(6), 1398–1403.
- Portch, E., Havelka, J., Brown, C., & Giner-Sorolla, R. (2015). Using affective knowledge to generate and validate a set of emotion-related, action words. *PeerJ, 3*:e1100 <https://doi.org/10.7717/peerj.1100>.

- Psychology Software Tools, Inc. [E-Prime Version 2.0: Professional]. (2012). Retrieved from <http://www.pstnet.com>.
- Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research, 48*(4), 589–597.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? : The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341–365.
- Pynte, J. (1991). The locus of semantic satiation in category membership decision and acceptability judgment. *Journal of Psycholinguistic Research, 20*(4), 315–335.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 55*(4), 1339–1362.
- Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods, 39*(3), 600–605.
- Reisenzein, R. (2009). Emotional experience in the computational belief–desire theory of emotion. *Emotion Review, 1*(3), 214–222
- Renoult, L., Brodeur, M. B., & Debrulle, J. B. (2010). Semantic processing of highly repeated concepts presented in single–word trials: Electrophysiological and behavioral correlates. *Biological Psychology, 84*(2), 206–220.
- Ridgeway, D., Waters, E., & Kuczaj, S. A. (1985). Acquisition of emotion–descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology, 21*(5), 901–908.
- Roberson, D., Damjanovic, L., & Kikutani, M. (2010). Show and tell: The role of language in categorizing facial expression of emotion. *Emotion Review, 2*(3), 255–260.
- Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a “category adjustment” model. *Memory & Cognition, 35*(7), 1814–1829.
- Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*(6), 977–986.
- Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*(1), 1–42.
- Roberts, T., & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception, 17*(4), 475–481.



- Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin*, *30*(11), 1472–1484.
- Roelofs, A. (2005). The visual–auditory color–word Stroop asymmetry and its time course. *Memory & Cognition*, *33*(8), 1325–1336.
- Roesch, E. B., Sander, D., Mumenthaler, C., Kerzel, D., & Scherer, K. R. (2010). Psychophysics of emotion: The QUEST for emotional attention. *Journal of Vision*, *10*(3), 1–9.
- Rosenberg, E., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion*, *19*(2), 111–138.
- Russell, J. A. (1991a). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology*, *60*(1), 37 – 47.
- Russell, J.A. (1991b). Culture and the categorization of emotions. *Psychological Bulletin*, *110*(3), 426–450.
- Russell, J. A. (1993). Forced–choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41–51.
- Russell, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of cross–cultural studies. *Psychological Bulletin*, *115*(1), 102– 141
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819.
- Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development*, *11*(1), 30–52.
- Saeki, E., & Saito, S. (2004). Effect of articulatory suppression on task-switching performance: Implications for models of working memory. *Memory*, *12*(3), 257–271.
- Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory & Cognition*, *22*(2), 181–187.
- Saito, S. (1997). When articulatory suppression does not suppress the activity of the phonological loop. *British Journal of Psychology*, *88*(4), 565–578.
- Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*(4-6), 313–322.

- Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research, 40*(1), 10–18.
- Salverda, A. P., & Altmann, G. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance, 37*(4), 1122–1133.
- Sánchez, A., & Vázquez, C. (2013). Prototypicality and intensity of emotional faces using an anchor–point method. *The Spanish Journal of Psychology, 16*, E7, 1–11.
- Sauter, D. A., LeGuen, O., & Haun, D. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion, 11*(6), 1479–1483.
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 28–43.
- Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In K. R. Scherer & P. Ekman (Eds.). *Approaches to Emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.
- Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*(1), 113–130.
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology, 62*(10), 1904–1914.
- Schmidt, J., & Zelinsky, G. J. (2011). Visual search guidance is best after a short delay. *Vision Research, 51*(6), 535–545.
- Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two–way crossmodal priming paradigm. *Experimental Psychology, 55*(2), 121–132.
- Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(2), 374–379.
- Schurigin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision, 14*(13), 1–16.
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 82–102.

- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402–409.
- Sekiguchi, T. (1999). Locus of semantic satiation examined with a serial matching task. *Perceptual and Motor Skills, 89*(3), 753–759.
- Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413.
- Shariff, A. F., & Tracy, J. L. (2011). Emotion Expressions On Signals, Symbols, and Spandrels- A Response to Barrett (2011). *Current Directions in Psychological Science, 20*(6), 407–408.
- Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*(6), 1061–1086.
- Shivde, G., & Anderson, M. C. (2011). On the existence of semantic working memory: evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1342–1370.
- Siakaluk, P. D., Knol, N., & Pexman, P. M. (2014). Effects of emotional experience for abstract words in the Stroop task. *Cognitive Science, 38*(8), 1698–1717.
- Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition, 106*(1), 433–443.
- Smith, L. C. (1984). Semantic satiation affects category membership decision time but not lexical priming. *Memory & Cognition, 12*(5), 483–488.
- Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology, 17*(19), 1680–1685.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813–838.
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184–189.
- Smith, L., & Klein, R. (1990). Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 852–861.

- Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. *Behavior Research Methods*, *44*(1), 256-269.
- Somerville, L. H., & Whalen, P. J. (2006). Prior experience as a stimulus category confound: an example using facial expressions of emotion. *Social Cognitive and Affective Neuroscience*, *1*(3), 271–274.
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, *130*(4), 799–832.
- Stevenson, R. A., Mikels, J. A., & James, T. W. (2007a). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods*, *39*(4), 1020–1024.
- Stevenson R. A., Mikels J. A., James T. W. (2007b). Manual for the categorization of the Affective Norms for English Words (ANEW) by discrete emotions. Retrieved from: [http://www.indiana.edu/~panlab/people/ras/Stevenson\\_ANEWManual\\_2007](http://www.indiana.edu/~panlab/people/ras/Stevenson_ANEWManual_2007)
- Storbeck, J., Robinson, M. D., & McCourt, M. E. (2006). Semantic processing precedes affect retrieval: The neurological case for cognitive primacy in visual processing. *Review of General Psychology*, *10*(1), 41–55.
- Subreddits: Sample Size [Online recruitment forum]: Retrieved from: <http://www.reddit.com/r/SampleSize>
- Suegami, T., & Michimata, C. (2010). Effects of Stroop Interference on Categorical Perception in Simultaneous Color Discrimination. *Perceptual and Motor Skills*, *110*(3), 857–878.
- SurveyMonkey Inc. [Questionnaire Presentation Software]: Palo Alto, California, USA. Retrieved from: <http://www.surveymonkey.com>.
- Sweeny, T. D., Grabowecky, M., Suzuki, S., & Paller, K. A. (2009). Long-lasting effects of subliminal affective priming from facial expressions. *Consciousness and Cognition*, *18*(4), 929–938.
- Szczepanowski, R., & Pessoa, L. (2007). Fear perception: can objective and subjective awareness measures be dissociated?. *Journal of Vision*, *7*(4), 1–17.
- Szczepanowski, R., Traczyk, J., Wierzchoń, M., & Cleeremans, A. (2013). The perception of visual emotion: Comparing different measures of awareness. *Consciousness and Cognition*, *22*(1), 212–220.
- Thesaurus.com (2015). Retrieved from: <http://www.Thesaurus.com>

- Thielscher, A., & Pessoa, L. (2007). Neural correlates of perceptual choice and decision making during fear–disgust discrimination. *The Journal of Neuroscience*, *27*(11), 2908–2917.
- Tian, X., & Huber, D. E. (2010). Testing an associative account of semantic satiation. *Cognitive Psychology*, *60*(4), 267–290.
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*(1), 163–173.
- van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX–UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247.
- Vinson, D., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition & Emotion*, *28*(4), 737–746.
- Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190.
- Võ, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross–validating the Berlin affective word list. *Behavior Research Methods*, *38*(4), 606–609.
- Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL–R). *Behavior Research Methods*, *41*(2), 534–538.
- Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250.
- Yik, M., Widen, S. C., & Russell, J. A. (2013). The within–subjects design in the study of facial expressions. *Cognition & Emotion*, *27*(6), 1062–1072.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.
- Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2014). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1–24.
- Wickham, L. H., & Swift, H. (2006). Articulatory suppression attenuates the verbal overshadowing effect: A role for verbal encoding in face identification. *Applied Cognitive Psychology*, *20*(2), 157–169.

- Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, 5(1), 72–77.
- Widen, S. C., Christy, A. M., Hewett, K., & Russell, J. A. (2011). Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition and Emotion*, 25(5), 898–906.
- Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development*, 23(2), 291–312.
- Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39(1), 114–128.
- Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29, 719–736.
- Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion*, 6(3-4), 285–319.
- Wierzbicka, A. (1994) Emotion, language and cultural scripts. In S. Kitayama & H. R. Markus (Eds.). *Emotion and Culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.
- Wilschut, A., Theeuwes, J., & Olivers, C. N. (2014). Priming and the guidance by visual and categorical templates in visual search. *Frontiers in Psychology*, 5(148), 1–12.
- Wilson, M. (1988). MRC Psycholinguistic Database: Machine–usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1), 6–10.
- Wilson–Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, 49(5), 1105–1127.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780–7785.
- Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top–down guidance in visual search. *Vision Research*, 44(12), 1411–1426.
- Young, S. G., Elliot, A. J., Feltman, R., & Ambady, N. (2013). Red enhances the processing of facial expressions of anger. *Emotion*, 13(3), 380–384.
- Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, 63(3), 271–313.

- Yu, H., Garrod, O. G., & Schyns, P. G. (2012). Perception–driven facial expression synthesis. *Computers & Graphics, 36*(3), 152–162.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151–175.





## Appendices

### Appendix A: Volunteer information sheet provided during experiments 1, 2a, 2b and 3.

*Explicit instructions to participants are emboldened.*

#### Information Sheet

**Researcher(s):** Emma Portch (e.s. portch@leeds.ac.uk)

**Supervisors:** Charity Brown (pscbr@leeds.ac.uk) and Jelena Havelka (J.Havelka@leeds.ac.uk)

**We are approaching you as a potential research participant to take part in our study. The experiment will consist of a computer-based emotion judgement task, with two blocks of trials. Per trial you will be required to repeat specific words at the rate at which they appear on the screen. It may not be appropriate for you to partake in this experiment if you suffer from epilepsy or migraines, as these words are presented very quickly and you will be required to maintain a steady rate of repetition. Please inform the researcher if this applies to you. Following the presentation of these words, you will then be shown two photographed faces, positioned side by side. You will be asked to judge whether these faces match each other in emotional expression (e.g. if you think that both faces display a 'sad' expression then you should respond that they 'match'). Importantly, you should make your matching decision (a) irrespective of the words repeated prior to face presentation, and (b) while the faces are still displayed on the screen. You will indicate your matching decision using the '1' and '9' keys on the keyboard; the experimenter will inform you with regards to key mapping and additional instructions will be provided as you progress through the task.**

Once you have completed the study you will be thanked for your participation and then debriefed with regards to the aims of the study. Results gained from this study will be

used for research purposes only. This may include presentation at research conferences or publication in academic journals. This means that we may need to retain the data you provide in this experiment for a period of up to 10 years. However, the data you provide will at all times remain anonymous. Anonymised research data from this project may be archived electronically at the UK Data Archive or in a similar facility in order to make them available to other researchers. This is in line with current data sharing practices and will ensure that the data has longer-term value for the wider research community. We are happy to answer specific questions you may have about this process.

This research is subject to ethical guidelines set out by the British Psychological Society. These guidelines include principles such as obtaining your informed consent before research starts, notifying you of your right to withdraw, and protection of your anonymity. This research has been reviewed and approved by the University of Leeds Ethics Committee (Ethics reference: 12-0179; Date of Approval: 22/10/2012). This sheet will hopefully provide you with enough information about the study to allow you to make an informed decision about participation. However, if you have any questions or would like to discuss anything with me please let me know.

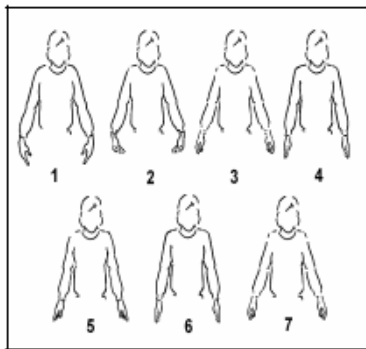
On signing the consent sheet you are giving us permission to log your results and use them for research purposes. However, you will remain anonymous as your name will be replaced with a unique identification code.

If you have any questions, please feel free to ask before the study begins.

## Appendix B: Health and safety exercises used in experiment 2b

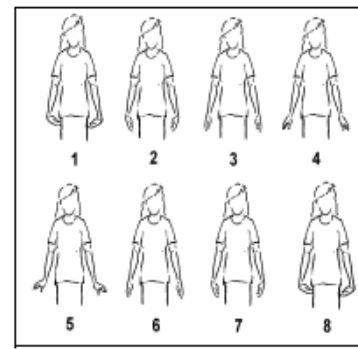
Figure A.1 Reproduced from: *Ergonomics: Health problems and exercises*. Software Educational Resources Ltd. Auckland: New Zealand.

Please note that Software Educational Resources Ltd no longer trades. The relevant copyright permissions to reproduce these instructions and diagrams were obtained from Cheryl Price ([Cherylprice.co.nz](mailto:Cherylprice.co.nz)) via email correspondence (available upon request).



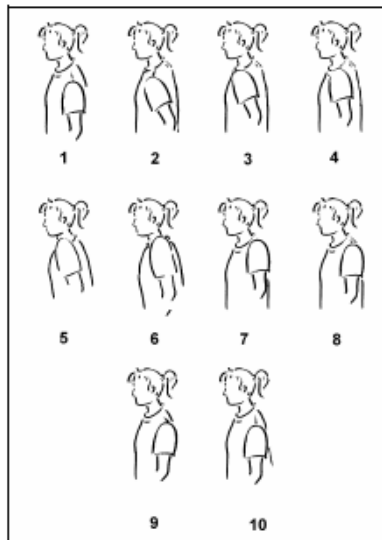
### Exercise One

Let your shoulders go loose and shake your whole arm(s), holding your arms straight up and down.



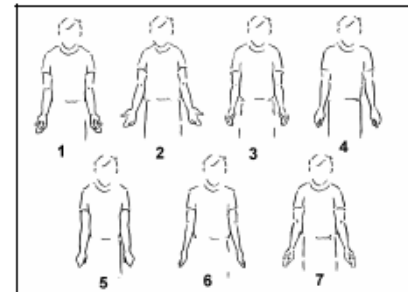
### Exercise Two

Let your wrists go loose, move your arms back and forth to make your hands flap, bending loosely at your wrist.



### Exercise Three

In a relaxed manner, circle your shoulders (hold the rest of your body as still as possible) so they describe the biggest circles possible. Let your arms hang still while your shoulders move. Do big circles in both directions.



### Exercise Four

Sometimes, our work causes us to hold some muscles short and tight. These need gentle lengthening.

Straighten your arms, hold them down by your thighs, both ways. Hold for a few seconds at the extremes.

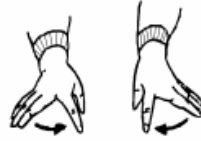
### Exercise Seven

1



With forearms rested, turn palms up then down.

2



With fingers straight, spread apart then together.

3



Touch each finger to thumb in turn. Repeat sliding finger tip to base of thumb.

4



Bend wrist to 90°, fingers straight - make a fist.

## Appendix C: Volunteer information sheet provided for experiment 4.

*Explicit instructions to participants are emboldened.*

### Participant Information Sheet- 'Tapping' language: does repetitive tapping affect your ability to perceive emotion?

**Researchers:** Emma Portch (e.s.portch@leeds.ac.uk)

**Supervisors:** Jelena Havelka (J.Havelka@leeds.ac.uk); Charity Brown (pscibr@leeds.ac.uk), and Russell Hutter (R.R.C.Hutter@leeds.ac.uk)

In this study we want to explore whether verbal repetition and steady-rhythm foot tapping have a similar influence on the ability to perceive emotion from faces.

**If you agree to take part in this study you will complete two, similar blocks of computer-based trials. Per trial you will be presented with a short or long series of fixation symbols on the screen. If you see a DASH symbol (- -) you should tap your right foot, and you should do this every time a dash appears. The researcher will demonstrate how you should tap and a piece of paper will be placed under your foot to help you maintain your tapping rate. If you see a CROSS symbol (+) you should passively watch as each new cross appears on the screen. Please refrain from counting the number of symbols that appear. You will complete several practice trials to familiarise yourself with the two different types of cue. After 'tapping' or 'watching' you will decide whether two briefly presented faces match or mismatch each other in emotional expression (e.g. if you think that both faces display a 'sad' expression then you should respond that they 'match'). You will use the '1' and '9' keys across the top of the keyboard to make your decision; the researcher will instruct you with regards to key mapping. Please try to submit a response while the faces are still displayed on the screen. Further instructions will be provided as you progress through the task.**

**Please note that in order to take part you must be right-handed, have no extensive/formal musical training and have normal or corrected-to-normal vision. In addition, it may not be appropriate for you to participate if (a) you suffer from migraines and/or epilepsy, or (b) have any pre-existing lower leg/foot injuries on your right side, that would prevent you from maintaining a steady tapping rhythm. Please inform the researcher if you are unsure about your suitability. Please also note that you can request a break, or withdraw from the study, if you experience any pain or discomfort associated with the tapping activity. Please inform the researcher if this is the case and please refrain from providing a response if you have avoided the tapping activity on a particular trial.**

Please remember that participation in this study is voluntary, and you can withdraw from it at any time. You also have the right to request your data to be destroyed after you have completed the study, for a period up to and including 1<sup>st</sup> April, 2015. You can do this by contacting the researcher. All information obtained in this study will be strictly confidential and your data will be stored under an individual participant code, which will be given to you at the end of the study. Please note that results will be used for research purposes only. This may include presentation at research conferences or publication in academic journals. This means that we may retain the data you provide for a period of up to 10 years. Anonymised research data from this project may be archived electronically at the UK Data Archive or in a similar facility in order to make them available to other researchers. This is in line with current data sharing practices and will ensure that the data has longer-term value for the wider research community. We are happy to answer specific questions you may have about this process.

This research is subject to ethical guidelines set out by the British Psychological Society and complies with the requirements of and has been reviewed by the Institute of Psychological Sciences Ethics Committee (ethics reference number: 14-0238, date of approval: 10/11/14).

This sheet has hopefully provided you with enough information about the study to allow you to make an informed decision about participation. However, if you feel you would like to ask any further questions please do so.

## Appendix D: Volunteer information sheet provided for experiment 5.

*Explicit instructions to participants are emboldened.*

### Participant Information Sheet- Main experiment

**Researchers:** Emma Portch (e.s.portch@leeds.ac.uk)

**Supervisors:** Jelena Havelka (J.Havelka@leeds.ac.uk), Charity Brown (pscbr@leeds.ac.uk), and Russell Hutter (R.R.C.Hutter@leeds.ac.uk)

We are interested in how efficiently people can extract socially-relevant information from faces and how language might influence this process.

**If you agree to take part in this study you will complete four, similar blocks of computer-based trials. In the first set of trials you will be presented with a series of male faces and will need to decide whether each face displays an emotional or neutral expression. If you think that 'Yes', the face displays an emotional expression, then you should press the Left Shift Key. However, if you think that 'No', the face does not display an emotional expression, then you should press the Right Shift Key. You should respond as quickly and as accurately as possible. In the three remaining blocks, each trial will begin with the presentation of a word on the screen, for example 'Sad', 'Gasp', 'Pen', 'Bus'. You will be required to read this word out loud. Following word repetition, you will see a briefly presented face and will need to make the same emotionality decision as you did before. You will be given additional instructions as you progress through the task. If you suffer from epilepsy or migraines, it may not be appropriate for you to partake in this experiment, as these faces are presented very quickly. Please inform the researcher if this applies to you. In addition, you must be over the age of 18 and a native English speaker to participate.**

Please remember that participation in this study is voluntary, and you can withdraw from it at any time. You also have the right to request your data to be destroyed after you have completed the study, for a period up to and including 1<sup>st</sup> April, 2015. You can

do this by contacting the researcher. All information obtained in this study will be strictly confidential and your data will be stored under an individual participant code, which will be given to you at the end of the study. Please note that results will be used for research purposes only. This may include presentation at research conferences or publication in academic journals. This means that we may retain the data you provide for a period of up to 10 years. Anonymised research data from this project may be archived electronically at the UK Data Archive or in a similar facility in order to make them available to other researchers. This is in line with current data sharing practices and will ensure that the data has longer-term value for the wider research community. We are happy to answer specific questions you may have about this process.

This research is subject to ethical guidelines set out by the British Psychological Society and complies with the requirements of and has been reviewed by the Institute of Psychological Sciences Ethics Committee (reference number: 15-0050, date of approval: 08/02/15).

This sheet has hopefully provided you with enough information about the study to allow you to make an informed decision about participation. However, if you feel you would like to ask any further questions please do so.



## Appendix E: Full list of modal exemplars provided in the feature generation work (Chapter 7, Study 1)

Table B.1 Alphabetised, full set of modal action words, generated by two or more participants during study 1 (N=25). Unmerged response frequencies, per emotion label, are shown. Responses regarded as synonymous, via our criteria, are labelled as 'core' or 'subsidiary' and corresponding core or subsidiary action words are provided in the final column.

	<i>Sad</i>	<i>Happy</i>	<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Surprise</i>	<i>Core/Subsidiary</i>	<i>Synonymous core/subsidiary exemplar(s)</i>
<i>Attack</i>	0	0	2	0	0	0		
<i>Avoid</i>	0	0	0	4	2	0	<i>Subsidiary</i>	<i>Hide</i>
<i>Beat</i>	0	0	2	0	0	0		
<i>Break</i>	0	0	2	0	0	0		
<i>Cheer</i>	0	3	0	0	0	0	<i>Subsidiary</i>	<i>Clap</i>
<i>Clap</i>	0	4	0	0	0	0	<i>Core</i>	<i>Cheer</i>
<i>Clench</i>	0	0	3	0	0	0		
<i>Cower</i>	0	0	0	0	2	0		
<i>Cringe</i>	0	0	0	2	0	0	<i>Subsidiary</i>	<i>Recoil</i>
<i>Cry</i>	18	6	7	0	9	3	<i>Core</i>	<i>Sob; Weep</i>
<i>Dance</i>	0	6	0	0	0	0	<i>Core</i>	<i>Skip</i>
<i>Exclaim</i>	0	0	0	0	0	2		
<i>Fight</i>	0	0	4	0	0	0		

	<i>Sad</i>	<i>Happy</i>	<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Surprise</i>	<i>Core/Subsidiary</i>	<i>Synonymous core/subsidiary exemplar(s)</i>
<i>Frown</i>	9	0	3	4	0	0	<i>Core</i>	<i>Grimace</i>
<i>Gag</i>	0	0	0	3	0	0	<i>Core</i>	<i>Retch</i>
<i>Gape (originally 'open mouth')</i>	0	0	0	0	0	2		
<i>Gasp</i>	0	0	0	0	2	8	<i>Core</i>	<i>Inhale; Sharp Intake</i>
<i>Giggle</i>	0	2	0	0	0	0	<i>Subsidiary</i>	<i>Laugh</i>
<i>Grimace</i>	0	0	0	2	0	0	<i>Subsidiary</i>	<i>Frown</i>
<i>Grin</i>	0	6	0	0	0	2	<i>Subsidiary</i>	<i>Smile</i>
<i>Growl</i>	0	0	4	0	0	0		
<i>Hide</i>	2	0	0	0	11	0	<i>Core</i>	<i>Avoid</i>
<i>Hit</i>	0	0	9	0	0	0	<i>Core</i>	<i>Punch</i>
<i>Hug</i>	0	8	0	0	0	0		
<i>Hum</i>	0	2	0	0	0	0		
<i>Hurt</i>	2	0	0	0	0	0		
<i>Ignore</i>	0	0	0	2	0	0		
<i>Inhale</i>	0	0	0	0	0	2	<i>Subsidiary</i>	<i>Gasp</i>
<i>Isolate</i>	2	0	0	0	0	0		
<i>Joke</i>	0	2	0	0	0	0	<i>Core</i>	<i>Play</i>
<i>Jump</i>	0	2	0	0	5	15		
<i>Kick</i>	0	0	5	0	0	0		

	<i>Sad</i>	<i>Happy</i>	<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Surprise</i>	<i>Core/Subsidiary</i>	<i>Synonymous core/subsidiary exemplar(s)</i>
<i>Laugh</i>	0	20	0	0	0	11	<i>Core</i>	<i>Giggle</i>
<i>Mope</i>	2	0	0	0	0	0		
<i>Panic</i>	0	0	0	0	7	0		
<i>Play</i>	0	2	0	0	0	0	<i>Subsidiary</i>	<i>Joke</i>
<i>Punch</i>	0	0	4	0	0	0	<i>Subsidiary</i>	<i>Hit</i>
<i>Rage</i>	0	0	2	0	0	0	<i>Subsidiary</i>	<i>Rant</i>
<i>Raise Eyebrow</i>	0	0	0	0	0	3		
<i>Rant</i>	0	0	4	0	0	0	<i>Core</i>	<i>Rage</i>
<i>Recoil</i>	0	0	0	5	0	0	<i>Core</i>	<i>Cringe</i>
<i>Retch</i>	0	0	0	2	0	0	<i>Subsidiary</i>	<i>Gag</i>
<i>Run</i>	0	0	0	0	13	0		
<i>Scream</i>	0	0	17	0	7	4	<i>Core</i>	<i>Shout; Shriek, Yell</i>
<i>Shake</i>	0	0	6	0	8	0	<i>Core</i>	<i>Shiver</i>
<i>Sharp intake</i>	0	0	0	0	0	3	<i>Subsidiary</i>	<i>Inhale</i>
<i>Shiver</i>	0	0	0	0	3	0	<i>Subsidiary</i>	<i>Shake</i>
<i>Shout</i>	0	0	13	0	0	4	<i>Subsidiary</i>	<i>Scream</i>
<i>Shriek</i>	0	0	0	0	0	2	<i>Subsidiary</i>	<i>Scream</i>
<i>Shudder</i>	0	0	0	2	0	0		
<i>Sigh</i>	2	0	0	0	0	0		

	<i>Sad</i>	<i>Happy</i>	<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Surprise</i>	<i>Core/Subsidiary</i>	<i>Synonymous core/subsidiary exemplar(s)</i>
<i>Sing</i>	0	5	0	0	0	0		
<i>Skip</i>	0	4	0	0	0	0	<i>Subsidiary</i>	<i>Dance</i>
<i>Sleep</i>	3	0	0	0	0	0		
<i>Slow</i>	2	0	0	0	0	0		
<i>Smile</i>	0	21	0	0	0	4	<i>Core</i>	<i>Grin</i>
<i>Sneer</i>	0	0	0	3	0	0		
<i>Sob</i>	3	0	0	0	0	0		
<i>Socialise</i>	0	2	0	0	0	0		
<i>Squeal</i>	0	0	0	0	0	2		
<i>Stomp</i>	0	0	2	0	0	0		
<i>Stutter</i>	0	0	0	0	2	0		
<i>Swear</i>	0	0	3	0	0	0		
<i>Sweat</i>	0	0	0	0	4	0		
<i>Tense</i>	0	0	2	0	4	0		
<i>Throw</i>	0	0	4	0	0	0		
<i>Vomit</i>	0	0	0	5	0	0		
<i>Wallow</i>	2	0	0	0	0	0		
<i>Weep</i>	2	0	0	0	0	0		
<i>Withdraw</i>	7	0	0	4	2	0		

	<i>Sad</i>	<i>Happy</i>	<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Surprise</i>	<i>Core/Subsidiary</i>	<i>Synonymous core/subsidiary exemplar(s)</i>
<i>Yell</i>	0	0	4	0	0	2	<i>Subsidiary</i>	<i>Scream</i>
<i>Yelp</i>	0	0	0	0	0	2		

## Appendix F: Mean scores for the full set of 36 label/action word ratings (Chapter 7, study 2)

Table C.1 Mean ratings (SEM) for all action word and label pairings. Modal pairings are presented in bold.

	Scream	Smile	Jump	Cry	Recoil	Hide	Totals
<b>Anger</b>	<b>3.96 (0.11)</b>	1.62 (0.081)	2.07 (0.12)	3.18 (0.14)	2.44 (0.12)	2.54 (0.13)	2.64 (0.12)
<b>Happy</b>	2.82 (0.14)	<b>4.78(0.052)</b>	3.03(0.14)	2.59(0.14)	1.43(0.078)	1.60(0.091)	2.71 (0.11)
<b>Surprise</b>	3.72(0.11)	3.27(0.12)	<b>4.06(0.13)</b>	2.62(0.12)	3.37(0.12)	2.13(0.11)	3.20 (0.12)
<b>Sad</b>	2.08(0.11)	1.61(0.083)	1.49(0.075)	<b>4.46(0.078)</b>	2.14(0.12)	3.62(0.13)	2.57 (0.099)
<b>Disgust</b>	2.84(0.14)	1.36(0.074)	1.98(0.12)	2.44(0.13)	<b>4.20(0.11)</b>	2.36(0.12)	2.53 (0.12)
<b>Fear</b>	3.83(0.11)	1.53(0.074)	3.37(0.16)	3.45(0.12)	3.88(0.11)	<b>4.35(0.075)</b>	3.40 (0.11)
<b>Totals</b>	3.21 (0.12)	2.36 (0.067)	2.67 (0.12)	3.12 (0.12)	2.91 (0.11)	2.77 (0.11)	