

University of Canterbury

---

# VOICE ACTIVATED

EXPLORING THE EFFECTS OF VOICES ON BEHAVIOURS

---

by

Andrew Euan MacFarlane, MA (Hons), MLitt (with distinction)

A thesis submitted to the University of Canterbury

in fulfilment of the requirements of the degree of

Doctor of Philosophy in Psychology

at

the Department of Psychology

and New Zealand Institute of Language, Brain and Behaviour

*Thesis examined by:*  
*Professor John A. Bargh*  
*&*  
*Dr. Matt Crawford*

---

# Contents

<b>Contents</b>	<b>iii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>Acknowledgements</b>	<b>xii</b>
<b>Abstract</b>	<b>xiv</b>
<b>I Introduction and Literature</b>	<b>1</b>
<b>1 Prelude</b>	<b>2</b>
1.1 Aims of Thesis . . . . .	3
1.2 Terminological Consistency . . . . .	4
1.3 Introduction . . . . .	5
1.4 Order of Thesis . . . . .	7
<b>2 Priming Behaviour</b>	<b>8</b>
2.1 Macro Processes . . . . .	8
2.2 Priming Studies . . . . .	10
2.3 Measuring Priming . . . . .	15
2.4 Perceptual Fluency and Mere Exposure . . . . .	17
2.5 Automaticity . . . . .	19
<b>3 Perceiving Others</b>	<b>23</b>
3.1 Social Representations . . . . .	24
3.2 Schemas and Stereotypes . . . . .	25
3.2.1 Schematic Thinking . . . . .	25
3.2.2 Stereotypes . . . . .	28

3.3	Level of Explanation . . . . .	31
<b>4</b>	<b>Language and Behaviour</b>	<b>33</b>
4.1	Style, Convergence and Divergence . . . . .	33
4.2	Lexical and Syntactic Influences . . . . .	35
4.3	Speech and Non-Linguistic Behaviour . . . . .	38
4.3.1	Speech, Environment and Behaviour. . . . .	38
4.4	Voice and Personality . . . . .	39
4.5	A Gap in Sociolinguistics . . . . .	42
<b>II</b>	<b>The Experiments</b>	<b>45</b>
4.6	Introduction to the Experiments . . . . .	46
4.7	A Note on Baselines . . . . .	47
4.8	Statistical Techniques . . . . .	47
<b>5</b>	<b>Experiment 1: Voice and Social Values</b>	<b>51</b>
5.1	Introduction and hypothesis . . . . .	51
5.2	Self Presentation . . . . .	52
5.3	Experiment 1A . . . . .	53
5.3.1	Participants . . . . .	53
5.3.2	Statements . . . . .	54
5.3.3	Voices . . . . .	54
5.3.4	Stimuli Creation . . . . .	55
5.3.5	Results: Social Supportiveness . . . . .	57
5.4	Experiment 1B . . . . .	61
5.4.1	1B Hypothesis . . . . .	61
5.4.2	Participants and Procedure . . . . .	61
5.4.3	Qualitative Comments . . . . .	63
5.4.4	Results . . . . .	64
5.5	Re-analysis . . . . .	66
5.5.1	Experiment 1A Re-analysed . . . . .	66
5.5.2	1A Re-analysis Hypothesis . . . . .	67
5.5.3	Experiment 1B Re-analysed . . . . .	69
5.5.4	Experiments 1A and 1B Combined . . . . .	71
5.6	Experiment 1C: Silent Condition . . . . .	73
5.6.1	Procedure . . . . .	73

5.6.2	Results . . . . .	74
5.7	Experiment 1 Discussion . . . . .	76
5.7.1	Summary . . . . .	76
5.8	Evaluating Men and Women . . . . .	77
5.9	Gender Schema . . . . .	79
5.10	Persuasive appeal . . . . .	80
5.10.1	Voice Affect and Processing . . . . .	81
5.10.2	Conclusion . . . . .	83
<b>6</b>	<b>Priming, Heuristics and Ambiguous Stimuli</b>	<b>85</b>
6.1	Introduction . . . . .	85
6.2	Anchoring, Accessibility and Availability . . . . .	87
6.2.1	Non-numerical Anchoring . . . . .	89
6.3	Experiment 2: Ambiguous stimuli . . . . .	91
6.4	Method . . . . .	91
6.5	Hypotheses and Questions . . . . .	91
6.5.1	Data Analysis . . . . .	92
6.6	Results . . . . .	94
6.6.1	Gay Themed Questions . . . . .	94
6.6.2	Gender Themed Questions . . . . .	97
6.6.3	Ethnicity Themed Questions . . . . .	100
6.6.4	Age Themed Questions . . . . .	104
6.6.5	Speaker Effect Overall . . . . .	108
6.7	Summary . . . . .	108
6.8	Experiment 2 Discussion . . . . .	110
6.9	When category activation isn't enough . . . . .	114
<b>7</b>	<b>Experiment 3: Voice and Reaction Times</b>	<b>116</b>
7.1	Introduction . . . . .	116
7.2	The Elderly Stereotype . . . . .	117
7.2.1	Elderly Speech . . . . .	120
7.3	Experiment 3A: Younger Participants . . . . .	121
7.3.1	Reaction Time Task . . . . .	122
7.3.2	Audio Stimuli . . . . .	122
7.3.3	Speaker Ratings . . . . .	124
7.3.4	Procedure . . . . .	125

7.3.5	Participants . . . . .	125
7.3.6	Analysis . . . . .	126
7.3.7	Results . . . . .	127
7.4	Experiment 3B: Older Participants . . . . .	132
7.4.1	Method and Participants . . . . .	132
7.4.2	Data Procedure . . . . .	132
7.4.3	Results . . . . .	133
7.5	Experiment 3C: Silent Comparison . . . . .	138
7.5.1	Method and Participants . . . . .	138
7.5.2	Data Procedure . . . . .	139
7.5.3	Results for Younger Participants . . . . .	139
7.5.4	Results for Older Participants . . . . .	142
7.6	Experiment 3D: Voice Familiarity . . . . .	144
7.6.1	Method and Participants . . . . .	144
7.6.2	Results . . . . .	146
7.6.3	Presence of Voice Overall . . . . .	147
7.7	Discussion . . . . .	149
7.7.1	Stereotype Activation . . . . .	150
7.7.2	Social Facilitation . . . . .	153
7.7.3	Processing Fluency and Familiarity . . . . .	154
7.8	Experiment 3 Conclusion . . . . .	154
<b>III Discussion and Conclusion</b>		<b>156</b>
<b>8 General Discussion</b>		<b>157</b>
8.1	Speech Variation as a Tool . . . . .	159
8.2	Social Behaviour . . . . .	160
8.3	Limitations . . . . .	162
8.4	Implications and Future Directions . . . . .	167
	References . . . . .	170
<b>Appendices</b>		<b>192</b>
<b>A Statement List for Experiments 1A, 1B and 1C</b>		<b>193</b>
<b>B Question list for Experiment 2</b>		<b>195</b>

<b>C Information and Debrief Sheets</b>	<b>197</b>
C.1 Information and Debrief for Experiments 1 and 2 . . . . .	197
C.2 Information and Debrief for Experiment 3 . . . . .	200

# List of Tables

5.1	Exp 1A: Speaker information . . . . .	55
5.2	Exp 1A: Mean support values . . . . .	57
5.3	Exp 1A: Social support analysis . . . . .	59
5.4	Exp 1B: Speaker information . . . . .	63
5.5	Exp 1B: Mean support values . . . . .	64
5.6	Exp 1B: Social support analysis . . . . .	64
5.7	Exp 1A: Re-analysis . . . . .	67
5.8	Exp 1B: Re-analysis . . . . .	69
5.9	Exp 1A & 1B: Combined analysis . . . . .	71
5.10	Exp 1C: Combined analysis with 1A and 1B . . . . .	74
6.1	Exp 2: Analysis of heterosexual photograph question . . . . .	94
6.2	Exp 2: Analysis of heterosexual audio question . . . . .	95
6.3	Exp 2: Analysis of Lady Gaga question . . . . .	96
6.4	Exp 2: Analysis of gender themed questions . . . . .	98
6.5	Exp 2: Analysis of male crowd question . . . . .	99
6.6	Exp 2: Analysis of music origin question . . . . .	101
6.7	Exp 2: Analysis of pleasant speech question . . . . .	102
6.8	Exp 2: Analysis of human rights question . . . . .	103
6.9	Exp 2: Analysis of best age question . . . . .	104
6.10	Exp 2: Analysis of photograph age question . . . . .	106
6.11	Exp 2: Analysis of doctors visits question . . . . .	107
7.1	Exp 3A: Mixed model output . . . . .	128
7.2	Exp 3B: Mixed model output . . . . .	134
7.3	Exp 3B: Mixed model output . . . . .	136
7.4	Exp 3A & 3B combined: Mixed model output . . . . .	137
7.5	Exp 3C: Mixed model output . . . . .	140



7.6	Exp 3C: Mixed model output . . . . .	141
7.7	Exp 3C: Mixed model output . . . . .	143
7.8	Exp 3C: Mixed model output . . . . .	144
7.9	Exp 3D: Mixed model output . . . . .	146
7.10	Exp 3D: Mixed model output . . . . .	148

# List of Figures

5.1	Example layout of the questionnaire . . . . .	57
5.2	Exp 1A: Predicted support for each voice condition . . . . .	60
5.3	Exp 1A: Voice condition and social theme interaction . . . . .	60
5.4	Exp 1B: Participant sex and social theme interaction . . . . .	65
5.5	Exp 1A reanalysis: Predicted agreement by voice condition . . . . .	67
5.6	Exp 1B reanalysis: Predicted agreement by voice sex . . . . .	69
5.7	Exp 1B reanalysis: Predicted agreement by voice condition . . . . .	70
5.8	Exp 1A & 1B combined: Predicted agreement by voice sex . . . . .	71
5.9	Exp 1A & 1B combined: Predicted agreement by voice condition . . . . .	72
5.10	Exp 1C: Predicted agreement by mode of presentation . . . . .	74
5.11	Exp 1C: Predicted agreement by mode of presentation . . . . .	75
5.12	Exp 1C: Predicted support by participant sex . . . . .	75
6.1	Example presentation of ambiguous stimuli question . . . . .	92
6.2	Exp 2: Predicted response for male photograph question . . . . .	95
6.3	Exp 2: Predicted response for heterosexual audio question . . . . .	96
6.4	Exp 2: Predicted response for Lady Gaga question . . . . .	97
6.5	Exp 2: Predicted response for gay themed questions . . . . .	99
6.6	Exp 2: Predicted response for male crowd question . . . . .	100
6.7	Exp 2: Predicted response for music question . . . . .	101
6.8	Exp 2: Predicted response for pleasant speech question . . . . .	102
6.9	Exp 2: Predicted response for music question . . . . .	104
6.10	Exp 2: Predicted response for best age question . . . . .	105
6.11	Exp 2: Predicted response for age of male question . . . . .	106
6.12	Exp 2: Predicted response for doctors visits question . . . . .	107
6.13	Spreading activation diagram . . . . .	113
7.1	Exp 3: Experiment set-up . . . . .	122

7.2	Spread of reaction times for younger participants . . . . .	127
7.3	Exp 3A: Predicted reaction times . . . . .	129
7.4	Exp 3A: Predicted reaction times . . . . .	130
7.5	Spread of reaction times for older participants . . . . .	133
7.6	Exp 3B: Predicted reaction time . . . . .	134
7.7	Exp 3B: Predicted reaction times . . . . .	135
7.8	Exp 3A & 3B combined: Predicted reaction times . . . . .	137
7.9	Spread of reaction times for younger participants in silent condition . . . . .	139
7.10	Exp 3C: Predicted reaction times . . . . .	141
7.11	Spread of reaction times for older participants in silent condition . . . . .	142
7.12	Exp 3C: Predicted reaction times . . . . .	143
7.13	Spread of reaction times for younger participants in Experiment 3D . . . . .	145
7.14	Exp 3D: Predicted reaction times . . . . .	147
7.15	Exp 3 overall: Predicted reaction times . . . . .	148

# Acknowledgements

I was enormously lucky to pursue my doctoral work in New Zealand, with its impossibly arresting landscapes, intriguingly stress-free citizens, and unbeatably stimulating coffee. I extend sincere gratitude to the *New Zealand Institute of Language, Brain and Behaviour* for providing me with that opportunity, and for funding and supporting me during my happy time there. Leading the institute is my friend and mentor Professor Jen Hay. Jen's combination of practical research skills and left-field thinking proved to be a neat fit for me - my own research skills benefited greatly, and my often unusual ideas were met with enthusiastic support.

Emma Parnell's friendship, administrative support, colourful turns of phrase and gossip kept even the rainiest of days bright. The tedious paperwork and policies for international students is never-ending, so having these dealt with in a timely and expert manner by Emma helped enormously. Our morning coffee sessions with rowdy laughter and occasionally (frequently) inappropriate humour are sorely missed.

Professor Lucy Johnston provided much help, academic and personal, throughout my studies both in a supervisory capacity, and in her role as Dean of Postgraduate Research. Lucy's thorough readings of my work often challenged me to reconsider my own thinking and line of argumentation, and her comments prompting me to *write, rinse, repeat*.

Thanks also to Dr. Ewald Neumann and Dr. Kumar Yogeeswaran. Ewald was attentive in making sure I stuck to APA rules and formatting, and sent me interesting and relevant articles to consider for my work. Kumar was an ideal boss to have when I taught social psychology labs. His enthusiasm for social psychology and his advice to me as I formulated my ideas was much

appreciated. I would also like to extend thanks to Associate Professor Paul Warren and the Victoria University of Wellington, where I was welcomed for a time after the 2011 earthquake in Christchurch. I enjoyed my time there immensely.

My appreciation and thanks must also go to my examiners, Professor John Bargh, and Dr. Matt Crawford. John's work was perhaps the single biggest inspiration to my work, and so to receive such positive, thoughtful comments as I did on this thesis was a personal highlight of my PhD adventure. Matt's comments were careful and incisive, and as oral examiner, engaged me in a thoughtful and reflective discussion of this work. This thesis was strengthened by their reviews. Thank you both.

I must extend thanks to all my colleagues and friends at the institute, with special mention for Jacqui, Romain, Vica, Mari and Pat. Afternoons on the grass, evenings in pubs, and nights gorging on dumplings whilst discussing everything from schema theory to Seneca kept me amused and informed. To my colleagues, who provided comments on this research at seminars and were there when I needed their particular expertise, my gratitude. Thanks also to Professor Jane Stuart-Smith and the Glasgow University Laboratory of Phonetics who gave me an 'institute away from institutes' after I returned to Scotland - it was like I had never been away.

Ten years of higher education, from MA through MLitt and now PhD, does not come cheap. A great deal of that expense has been shouldered by my parents, who have never questioned why I spent so long at University, or why what I do matters. Their faith that I somehow knew what I was doing was unwavering. For the countless bank transfers, flights paid, free dinners (with nice wine), sly £20 notes in jacket pockets, not to mention holidays to nice places...thank you very, very much indeed. Couldn't have done it without you!

# Abstract

Decades of priming research have revealed that environmental stimuli feed into our behaviours, often without any awareness of our using this information to guide our behaviour. This has been shown using plentiful stimuli across multiple contexts. One of the most socially rich stimuli in our environment is voice, and yet this has featured surprisingly little in behavioural research, particularly within social psychology. This thesis was written as a step towards addressing this gap, and it explores how voices might affect particular behaviours in different contexts.

Three broad experiments, each with their own sub-experiments, investigated how voices, acting as proxies for social categories, could influence one's behaviour. In the first experiment, the responses to socially themed statements were influenced by the sex of the voice presenting those statements. Female voices primed more agreement to these statements than did male voices. In the second experiment, judgements of ambiguous stimuli and questions were also affected by voices, albeit in less clear ways. In the third experiment, the reaction times of participants were again affected by voices. Younger participants' reaction times were slower when listening to an older voice, and older participants' reaction times were faster when listening to an older voice. Across these three experiments, I found too that the presence of a voice led to task differences compared to when voice was absent.

The combination of these experiments is, to my knowledge, the first to look at voice-based behavioural priming. How these results fit with selected existing theories, the potential to specify theories based on these results, and the possible practical applications of voice based priming are

discussed.

*Man himself may be controlled by his environment, but it is an environment which is almost wholly of his own making.* Skinner (1971, p 201)

---



## Part I

# Introduction and Literature

# Chapter 1

## Prelude

More than four decades ago, B.F. Skinner offered a characteristically pragmatic approach to investigating the behaviour of people in society when he wrote his classic *Beyond Freedom and Dignity* (Skinner, 1971). Central to his philosophy was a staunch belief in the illusory nature of free will - the ability of us humans, possessed of agency, to will ourselves however we so choose. Not only did Skinner refute free will, he found it to be a dangerous belief that obstructed progress towards a greater good. Whilst this thesis explores priming, it is more broadly about behaviours, directly inspired by Skinner's thinking. His work has remained controversial, and his ideas have fallen out of favour, a reaction against a perceived mechanism on his part (Dinsmoor, 1999). Emergent afterwards was what came to be called the cognitive revolution (see Goldstein, 2008). Whilst the behaviourist tradition of assigning one's responses to learned behaviour is outdated (although not without contemporary examples), the topic of free will has been subject to fierce debate in psychology.

In *Bypassing the Will*, Bargh (2005) argued that not only is conscious control over certain behaviours a myth, but also largely unnecessary as we go about our day-to-day lives. Bargh continued this theme in *Free Will is Unnatural* (Bargh, 2008), reviewing literature from across disciplines to build a case for, at the least, investing less faith in our belief in human agency. A crucial distinction between a classical behaviourist and researchers currently pursuing work on automatic behaviours is, as Bargh and Ferguson (2000) acknowledged, the "behaviourists refusal to consider mediating internal constructs and processes...in explanations of human behaviour" (Bargh & Ferguson, 2000, p 1). Central to my own work

is a classically behaviourist assumption: our actions can, and in some cases should, be explained with an appeal to observable causes. Deliberate learning and reinforcement need not be required to explain all of our behaviours. Simple exposure may, in some cases, suffice. In the experiments detailed in this thesis, this takes the form of judgements, decisions, and reaction speeds being traceable to a voice.

Influenced by current research into social cognition (see for example Wheeler & Berger, 2013; Shanks et al., 2013; Bargh, Schwader, Hailey, Dyer, & Boothby, 2012; Cameron, Brown-Iannuzzi, & Payne, 2012; Doyen, Klein, Pichon, & Cleeremans, 2012; Schwarz, 2011), it must be acknowledged that mental processes mediate this *cause and effect* behaviour, and where appropriate, possible theories relevant to the results in the thesis are outlined. I then elaborate on whether such mechanisms may be premature for the type of effects that are investigated, and ask whether it is a fruitful line of enquiry to be concerned with the processes rather than the context and environment, at least presently. Stroebe and Strack (2014) have written on the importance of having multiple results across multiple paradigms to illuminate a theory. It is because of this importance that theoretical conclusions are not drawn from individual results from the separate experiments. The objectives for this thesis were twofold: to demonstrate that voices can affect particular behaviours in comparison to when no voices are present, and to demonstrate that certain voices act differently in this respect to others. Satisfying these two objectives then has implications for current theories on language variation and change, and also provides new evidence in relation to existing priming results in social psychology. By the end of the thesis it should be apparent that the voices we hear around us have potential effects over and above mere speech content. That is to say that the voices themselves may affect our behaviour in subtle, yet likely unpredictable ways.

## 1.1 Aims of Thesis

The overarching aim for this thesis is to offer evidence showing that non-linguistic behaviour can be affected by voices we hear around us. This has not, at the time of writing, been shown in past research. Hearing different voices as we do different

tasks is a large part of many of our lives, making it surprising that more work has not been done to address the non-linguistic consequences of voices. To achieve this aim, evidence is sought to show that voices can affect behaviour in different types of tasks: tasks requiring social opinions, tasks requiring the judgement of ambiguous stimuli, and tasks requiring physical reactions. These types of tasks serve as a proxy for behaviours that people typically exhibit in real-life, and so demonstrating these effects under laboratory conditions is a crucial first step. A secondary aim of this thesis is to show that the presence of a voice in a task could lead to behavioural differences overall, compared to when voice was absent. In addition, there is an aim to show that particular voices could lead to particular effects, and may be open to prediction under well specified conditions.

## 1.2 Terminological Consistency

At the outset there are some terminological clarifications to set out. Some of these distinctions come from other researchers, others are set by the American Psychological Association style guide. This research has nothing to add to the debates at large surrounding terminology, and what constitutes as what, but rather this is for the reader's clarity and consistency. Throughout this thesis *sex* and *gender* are referred to. The terminological distinction was set out by Deaux (1985), and this is upheld here. Sex refers to biological categories, and gender to psychological differences between these categories. As such, where male and female participants are selected, they will be referred to by their sex. In discussing why male and female behaviours may differ, and the social constructs which operate on their biological category, gender will be used. *Gay* is used rather than *homosexual* in reference to relevant populations and individuals. Homosexuality has a history of being associated with pathology, and as I do not like being called it, prefer not to use it in writing. In referring to the social issues faced by gay people, such as same sex marriage, these are referred to simply as gay issues. *Heterosexual* is used for clarity as not all English speakers will be familiar with the term straight, and as the antonym of straight is bent, heterosexual seems preferable.

The terms Asia and Asian are used in this thesis. Different countries have different meanings for these terms. In the USA, it typically means someone from

East Asia, such as a Japanese person. In the UK, it primarily refers to people of South Asian origin, such as Pakistan. In this thesis, Asian refers to the ASEAN+3 criteria as used by the Association of East Asian Nations. This is a trade and political organization linking the respective countries, which are: the Philippines, Indonesia, Thailand, Malaysia, Singapore, Brunei, Vietnam, Myanmar, Cambodia and Laos - and the three East Asian nations of Japan, People's Republic of China and the Republic of Korea. Grouping these countries together is still not ideal, but it is widely used and, once defined, a reasonably accessible construct. Where I can be more specific, individual nations or peoples will be named.

The terms *unconscious* and *automatic* are used throughout. For these terms, the distinctions set out by Banaji, Lemm, and Carpenter (2001) are followed. Unconscious refers to any process that occurs outside of conscious awareness, with no ability of control, or with no intention to occur. Thinking about martial arts after hearing an Asian person speaking would constitute an unconscious process here. The term automatic is used to refer to the behavioural side of these processes, such as pressing a button more slowly upon hearing an older person would constitute an automatic response. More typical examples include aspects of driving, or changing your walking pace depending on the music you are listening to.

### 1.3 Introduction

Hearing other people speak is for many of us a large part of our daily lives. When we speak, we hear ourselves, and when we listen, we hear others. It is an integral part of our educational and working lives, and many make a career out of listening and speaking. We make friends, meet partners, network in our chosen fields, pick political candidates, enjoy films, and buy products in conjunction with hearing other people speak. We know that what people say is important. Being told you are fat has direct consequences on how you feel and perhaps even on your behaviour. When someone at a cocktail party says 'we should do lunch', we have a feeling of whether or not to get our diaries out. Moreover, we know that how people speak is important. Their tone of voice, intonation, tempo, and pitch all go towards informing you of the meaning (Egan, 1980; Kreuz & Roberts, 1995; Attardo, Eisterhold, Hay, & Poggi, 2003). In varying combinations, our

experience with these voice features tells us whether ‘What a great day’ was said with sincerity or sarcasm (Woodland & Voyer, 2011), and whether ‘I’m...busy right now’ is work related or because someone is romantically indisposed. When we are listening face to face, another important factor, gesturing and body language, feeds into the conversational mixture (McNeill, 1992; Beattie, 2003). Folded arms, a posed smile, or wild hand waving add to both the linguistic content and form to produce meaning.

This thesis adds voice into the mix. This does not entail speech content, nor the manner in which one speaks, nor even any visible features accompanying one’s speech, but solely the voice and how listeners might construe social characteristics from that voice. In a physical conversation we have a wealth of information with which to evaluate the speaker. A person’s sex, ethnicity, race, stature, clothing and many other features contribute to how we evaluate them through their speech (Braun, 1996; Schötz, 2006; Lim, 1997; Baugh, 2003; MacFarlane & Stuart-Smith, 2012). These other features are deliberately excluded in this thesis. In many situations however, such as on radio, telephone, call centres, advertisements and GPS systems to name a few, we have no accompanying information about the speaker. In these situations, the hypothesis is that the socio-demographic information about a speaker, carried in their speech, could have consequences outside of mere interpretation, and outside of that temporary speech act.

This thesis does not look at how listeners perceive speech, or their affective reactions to it as has been done broadly in the past (see Section 4.4). This limits the ability to draw conclusions about whether or not the voices in this thesis primed socially congruent behaviour, or whether the voice was simply pleasing or displeasing to people, resulting in task differences also. This is addressed in Section 8.3. Voice, as used in this thesis, serves as proxy for social category membership, so that a specific voice is a proxy for all voices of the relevant category. Whilst affective reactions were not explicitly tested for, alternative theories are offered, particularly where the results do not support a category based approach based on assimilation or contrast. I look at whether this voice proxy could have behavioural and attitudinal consequences, such that a person with a Chinese voice might encourage people to judge an ambiguous stimulus as more Chinese sounding, or whether a person with an elderly voice might affect the reaction times of people in a monotonous

reaction time task. In short, this thesis looks at the ability of voice to prime behaviour. Specifically it addresses whether or not a voice, in absence of a physical speaker, can affect different domains of behaviour: attitudes, evaluations, decisions, and reaction times. It is, to my knowledge, the first time this question has been investigated.

## 1.4 Order of Thesis

The thesis is presented in three parts, each with its own chapter(s). Part 1 is an introduction to the topic and the literature that has informed, and contextualizes, this work. Part two presents the three different experimental paradigms used to investigate the effects of voice. This is split into three chapters. Chapter 5 looks at attitudes and opinions, Chapter 6 at judgements and evaluations, and Chapter 7 at reaction times. Each of these chapters has its own shorter introduction with relevant literature, and its own discussion pertinent to that literature. Part 3 of the thesis broadens out to review these experiments and their relation to existing literature, and the discussion chapter in particular serves to offer future suggestions for this type of research, and poses questions that have arisen as a result of this work.

## Chapter 2

# Priming Behaviour

The primary focus in this thesis is the priming of behaviour. The working definition of priming borrows from Bargh, Chen, and Burrows (1996), who described the process as “the incidental activation of knowledge structures, such as trait concepts and stereotypes, by the current situational context” (Bargh et al., 1996, p 1). Priming can be considered a subset of the much larger work on heuristics in social psychology, particularly accessibility and availability which is detailed more in Section 6.2, in Chapter 6. This translates in the current thesis to investigating whether voices in the current situational context (the experiments) can impact upon the behaviour of participants who are listening to these voices and completing another task. Whilst there is a wealth of knowledge showing that the environment and its social factors can affect our linguistic behaviour (see Section 4.1), there is scant work describing the converse of this. This chapter begins by describing different types of priming research along with relevant studies, moving on to the different measurements that can be used. Passing reference is made to the current controversy surrounding priming research (although this is discussed more in the final chapter, see Section 8.3) before moving on to a discussion of automaticity, a theory that runs through this thesis and was the inspiration for this work.

### 2.1 Macro Processes

A number of different mechanisms have been proposed for particular priming effects, mirroring somewhat the theories of behaviour in social psychology more generally. Smith and Queller (2001) provide an overview of the approaches to mental representations in social psychology, contrasting the associative models



with the schematic ones (discussed further in Section 3.2.1). The authors make clear that these need not be competing theories, but can be complementary, which is the position taken in this thesis. Based on the strength of the existing research, and the relatedness of the mechanisms to the results, a select few of these theories are discussed in this chapter. Whatever differences they have, they typically share in common an underlying reliance on an associative network (Collins & Loftus, 1975), where the activation of a particular concept has knock on effects on other traits and concepts that share links in this associative network. Indeed, much of the research reviewed in the coming chapters deals with schematic representations using associative-network concepts (e.g., Valentino, 1999; Fazio, 2001; Mussweiler, 2006). Once the concept has been activated it heightens the accessibility for the concepts that share links with it. This is stronger for the more closely related concepts, and weakens as the distance from the initial concept to other related concepts increases. For example, the concept *cheese* is closely related to other concepts such as food, but as you travel further in the associative network, concepts such as *chalk* may be activated. This knock on effect is often referred to as spreading activation (Collins & Loftus, 1975; Neely, 1977). This has historically been shown using semantic decision tasks, such as when word pairs like DOCTOR-NURSE were judged as real words more quickly than were unrelated word pairs such as DOCTOR-BREAD (Meyer & Schvaneveldt, 1971).

Examples of research that indirectly use some form of spreading activation are Fazio (2001), Wittenbrink, Judd, and Park (2001) and Fitzsimons, Chartrand, and Fitzsimons (2008), who have demonstrated that not only are related lexical items made more accessible, but also attitudes and behaviours. This can be extended too, so that even an abstract concept such as *5 degrees Celsius* can make more accessible other concepts (or ‘links’ in Collins and Loftus terminology) such as *kalt* (cold) or *schlitten* (sleigh) (Mussweiler & Strack, 2000). Research varies on what happens after this initial associative activation. There are motivational accounts that suggest inhibitory mechanisms (Bodenhausen & Macrae, 1998), accounts that offer ease of retrieval as a mediator of following effects (Schwarz, Bless, & Bohner, 1991; Gawronski & Bodenhausen, 2005), and accounts that suggest a motivated preparation to interact (Cesario, Plaks, & Higgins, 2006). Research in the past few decades has also looked at spreading inhibition. To understand why inhibition may be required, D. Carlston (2010) gives the example of a friend called Bill. If

*Bill* is made active in the associative network, then it would not take long before ducks and retired U.S. presidents became more accessible too. This would not be conducive to a person who is, at the time, attempting to engage only with the concept of their friend Bill. To quote, he says that “Too many environmental stimuli could leave people drowning in a soup of excitation” (D. Carlston, 2010, p 46). Neumann and Deschepper (1992), who proposed an inhibitory process, have suggested that by temporarily making irrelevant information less accessible, processing advantages can be gained for whatever task is at hand.

## 2.2 Priming Studies

The research discussed in this section has directly influenced the theoretical and methodological approach to the research questions set out in this thesis. The focus is mostly on supraliminal priming, which is the type of priming utilized in the experiments, where participants are consciously aware of the primes but not on their intended effect on concurrent or subsequent tasks (see for example Higgins, Rholes, & Jones, 1977; Decoster & Claypool, 2004; Fitzsimons et al., 2008). Studies with subliminal priming are mentioned where relevant. Further, the focus is mostly on studies that have used language, such as words in isolation, written sentences or asking participants to write the stimuli themselves, as the priming stimuli. Although the definition of priming given by Bargh et al. (1996) is used, provided at the outset of this chapter, other definitions help to highlight the wide applicability of the term priming. Collins and Loftus (1975) wrote that “Priming (or preparation) involves the same tracing process that was described for memory search. When a concept is primed, activation tags are spread by tracing an expanding set of links in the network out to some unspecified depth” (Collins & Loftus, 1975, p 409). They noted in their article that an implication of this is that not only will nodes be primed (where a node can be a concept, or a lexical item, in an associative cognitive network), but the links associated with those nodes. Spreading activation, as discussed above, underlies many of the studies in priming that require trait or concept associations for their predictions. Arndt, Greenberg, and Cook (2002) investigated how the spreading activation of the mortality concept facilitated access to individual world-view constructs. Their results showed that having people think about mortality triggered a heightened

accessibility of world-view relevant content. Bargh, Raymond, Pryor, and Strack (1995) found that priming power constructs, with words such as *executive* and *macho*, in male participants led to more sexually themed evaluations of female targets. As these results indicate, exposing people to concepts such as power, sex and death heightens their accessibility to related concepts. In the present research, voices are used to activate concepts such as Asian, elderly, and gay, so that related concepts should be heightened for the participants, affecting their task behaviours.

Another proposed method of priming, sharing some similarities to spreading activation, is the compound cue account (see McKoon & Ratcliff, 1992). In this model, the strength of priming is related to whether the target and prime are already compounded in memory. The authors wrote that “People, words, and objects do not occur in isolation; rather, they occur in some larger context, and memory must provide the means of integrating the individual parts into the unified context” (McKoon & Ratcliff, 1992, p 1). The authors argued in the paper that their model accounted for results not predicted by a spreading activation account, such as the word *deer* activating the word *grain*. Classic spreading activation accounts would not prohibit this from happening, however, and it would vary from individual to individual depending on their experiences both historical and recent. Contemporary models that invoke an associative network, such as that of Mussweiler and Damisch (2008), have associations that are invoked, retrieved and compared against other concepts to form judgements. Similarly Gawronski and Bodenhausen (2005), who argue that spreading activation accounts may be overused, still rely on concept relatedness and make no mention of compounded items. In the experiments for this thesis, hearing a particular voice that one recognises as belonging to a particular social group is likely to be similarly compounded. Some participants hearing the elderly voice might have their concept of old tied to their concept of their grandmother, whilst for others, it may be more strongly tied to their job. Whether we invoke a simple spreading activation account, or a compound-cue account, is less of a concern because both should activate concepts that the participants relate to the voices they hear.

Other studies have looked at ease of retrieval as a mediator of priming, rather than a semantic association account. Schwarz et al. (1991) asked participants to write down either 6 or 12 examples of times they had been assertive. In a following

questionnaire, they were asked to rate themselves along lines of anxiety, insecurity and assertiveness. Contrary to a spreading activation account, Schwarz et al. (1991) found that participants rated themselves as less assertive after recalling 12 examples rather than 6. They argued that the difficulty in generating 12 examples from memory led participants to feel they must not be particularly assertive, as the ease of recall affected the availability of the construct. Conversely, those participants who only described 6 examples were argued to have found that somewhat easier, and the ease of this accessibility led to them feeling more assertive.

In a similar study that looked not at self ratings but at implicit associations, Gawronski and Bodenhausen (2005) conducted three experiments looking at stimulus compatibility accounts versus response compatibility accounts - that is to say, an account based on semantic association, and one based on ease of retrieval and valence respectively. The authors showed that in two response compatibility tasks (an IAT, and a lexical decision task), priming effects were mediated by the ease of retrieval from memory. In one of their experiments, participants listed either 3 or 10 women they considered strong. In a following task, they were presented with words related to either strength or weakness, and then asked to perform an Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998) on gender stereotypes. The IAT is a computer based response task that typically asks participants to assign a single stimulus, such as a word or image, to pairs of words, some of which fit into the same category, and some which do not. As an example, a Chinese face might serve as the stimulus, with participants asked to assign the face to the words BAD and CRIMINAL or to GOOD and LAWYER and many other pairs. It has proven to be a popular tool in social psychology, but is not without its criticisms (see Gawronski (2002) for a discussion). Participants who had listed 10 strong women were more susceptible to gender stereotyping, while those who had listed only 3 strong women were more resistant to implicit stereotyping. This is in contrast to a spreading activation account which would predict that listing a higher number of counter-stereotypical women would make related concepts more accessible, thus inhibiting gender stereotyping. That this did not happen, the authors suggested, indicated that the subjective ease or difficulty of listing 3 or 10 women was responsible for the priming effects.

Another variant of priming is affective priming, whereby the immediate,

affective reaction to the stimulus causes a behavioural change. Murphy and Zajonc (1993) wrote that “positive and negative affective reactions can be evoked with minimal stimulus input and virtually no cognitive processing” (Murphy & Zajonc, 1993, p 1). It is, then, quite different to the processes described above which require cognitive mediation of one type or another. Robert Zajonc (Zajonc, 1980), taking inspiration from Wundt (1907) and Bartlett (1937), argued that the affective system is separate to the cognitive system; it is the first system to respond to stimuli, and for many organisms, it is the only system (Zajonc, 1980; Winkielman, Zajonc, & Schwarz, 1997). Zajonc was clear to note, however, that cognition and affect are not always entirely distinct. Both in the laboratory, and even more-so so in everyday experience, affect likely influences cognition, from choosing to marry or divorce, to committing suicide (Zajonc, 1980, p 153). Murphy and Zajonc (1993) tested the automatic affect theory by displaying Chinese characters that had been preceded either by smiling or frowning faces subliminally. The ideographs elicited more negative reactions when preceded by the frowning faces, suggesting this near-instant, affective reaction carried over into a proceeding stimuli. This was replicated by Winkielman et al. (1997), using a nearly identical method with positively and negatively valenced faces preceding neutral ideographs. They built upon Murphy and Zajonc’s 1993 study by showing that even when subjects were told that the subliminal primes might affect their judgements, the priming effects remained. Similar results have also been shown by Payne, Cheng, Govorun, and Stewart (2005), who suggested that the misattribution effects reported in such studies was difficult to control because participants did not believe they were experiencing it (Payne et al., 2005). Results like these are in contrast to supraliminal studies that have shown when participants are made aware of the potential to be primed by a stimulus, the effects weaken or disappear (Lombardi, Higgins, & Bargh, 1987; Strack & Schwarz, 1993).

Outside of psychology, linguists have demonstrated that priming certain expectations can affect how participants evaluate language stimuli. Niedzielski (1999) played her Detroit participants 50 sentences recorded by a Detroit speaker. Following the sentences, the participants were played six computer resynthesized vowels, and asked to pick which one was closest to the vowels of the speaker they had heard speaking the sentences. The priming manipulation had half of the participants told the speaker had been a fellow Detroiter, and half told that

the speaker was from Ontario. In the vowel selection task, the two groups chose significantly different vowel sounds. Specifically, those who were told the speaker was Canadian chose more raised vowels, as is commonly associated with Canadian speech. Hay, Nolan, and Drager (2006) conducted a similar experiment. In their paradigm, participants heard a male speaker saying various sentences. In the following task, they were asked to pick a synthesized vowel from six choices which they felt was most like the vowel they had previously heard. Roughly half of the participants had *New Zealand* written on their answer sheet, and the other half *Australia*. In contrast to Niedzielski, participants were not told that these labels represented the nationality of the speaker, nor was their attention explicitly drawn to it. Their findings add to those by Niedzielski by showing that not only is explicit information about a speaker capable of affecting people's perceptions of speech, but so too the mere presence of information that could be construed as task relevant.

To end this section on the different types of priming, it needs noting that what has been reviewed is a small selection of the available literature. Cameron et al. (2012) have provided an extensive meta-analysis of priming literature, listing the moderating variables and reasons given for priming effects. The studies discussed above could all be used to help explain why voices might prime behaviour in the experiments in this thesis. I have presented research showing that the activation of concepts results in the activation of related concepts, that context can determine which concepts become active and which are inhibited, and that people typically have immediate, affective reactions to stimulus. Further, people's explicit evaluations can be separated from their implicit evaluations. From this research, it is proposed that voices are clear candidates for behavioural priming. They are a stimulus rich in content and the number of concepts or nodes that people will associate with a voice should be numerous, but given a particular population, should have enough overlap such that a gay voice will be associated with gay concepts, Asian voices with Asian concepts and so on. Once these concepts become active, the task behaviour of people listening to different voices should show differences as a result of the available associations. Secondly, voices are human, and affective reactions to humans are strong. Voices that are more or less liked should then result in different task behaviours. Finally, the expectation of relevance should play its part, with people perhaps inferring a voice and task relationship. For example, if people hear a female voice asking about questions

involving female issues, the expectancy of relevance should, in combination with associative priming and affective priming, lead to behavioural differences between female and non-female voice conditions.

## 2.3 Measuring Priming

I have discussed the different types of priming so far, and some of the mechanisms offered for their action. In their meta-review of the priming literature, Cameron et al. (2012) describe the different measures used to investigate priming, and some of the limitations of those methods. They wrote that reaction time measures have been widely used historically, but that problems with reliability (see Fazio & Olson, 2003) have prompted a shift to accuracy measures. Reaction time measurements allow researchers to target implicitly held attitudes (or context-specific attitudes, *see next paragraph*), those attitudes of which people are largely unaware of, and that go unacknowledged (Greenwald et al., 1998; Ferguson & Bargh, 2007). These beliefs are particularly dangerous because they are under the radar, and people with otherwise good intentions can be unaware of their own prejudices. Often however an explicitly reported attitude is sought (see for example Bodenhausen, Schwarz, Bless, & Wanke, 1995; Cesario et al., 2006; Valentino, Hutchings, & White, 2002). These explicit attitudes are concerned with people's deliberate, consciously aware attitudes, for example when asked whether or not they support immigration. The aim of the prime in such studies is usually not to assess implicit attitudes, but to test what sort of effects primes could have on real-world tasks, such as job candidate selection or presidential election voting (Valentino, 1999).

To overcome the limitations of explicit reporting strategies, many researchers have used implicit measures but these too present difficulties. As the research previously reviewed on priming demonstrated, the preceding and present context of a task can influence that task behaviour. Schwarz and Bohner (2001) have noted that the use of implicit attitudes has not, as might have been hoped, allowed researchers a "context independent window on respondents' true attitudes" and that attempts to document attitudes that are context independent, and stable, "may be overly optimistic" (Schwarz & Bohner, 2001, p 449). This was also the stance taken by Gilbert and Hixon (1991), who argued that automatic activation of information "does not mandate such use, nor does it determine the precise

nature of its use. It is possible for activated information to exert no effect on subsequent judgements or to have a variety of different effects” (Gilbert & Hixon, 1991, p 512). An example of a prime exerting no influence under particular conditions was shown by Macrae and Johnston (1998). In two experiments, they showed that both exogenous features (the environment) and endogenous features (internal factors such as goal states) could inhibit automatic behavioural effects. In their first experiment, they reasoned that being primed with the concept of helpfulness in a scrambled sentence task with words such as ‘supported’ would lead to participants picking up the experimenter’s dropped pen more often than in a control condition. However, if the pen was leaking, the cost of picking it up would override the automatic priming. This was indeed what they found. Their second experiment used the same scrambled sentence paradigm to prime helpfulness. After completion of the test, participants were instructed to go to a second experimenter and were informed either that the experiment was running on time, or was behind schedule. The researchers reasoned that the internal goal state to not be late would override the primed helping behaviour, and again, their results supported this prediction. Across their two experiments, then, they demonstrated that even when the behaviour is task relevant, both the environment and the inner state of the participant can negate the expected behavioural priming. Cameron et al. (2012) have noted the widely varying correlations between implicitly held attitudes, and actual behaviours, and Greenwald, Poehlman, Uhlmann, and Banaji (2009) report a correlation of .27 between the IAT and actual behaviour from their meta-analyses. Use of the IAT does however produce typically large effect sizes, making replication (and finding an effect) more probable (Payne et al., 2005).

Another measure that can be used and which is employed in this thesis is the evaluation or judgement of ambiguous stimuli or questions. Cameron et al. (2012) have suggested that measuring priming effects with ambiguous stimuli can be more robust than timed implicit priming tasks, such as the IAT, and that they are now commonly used in combination with each other. In this approach, the aim is to assess the influence of a prime on the following evaluation of a presented target. The target could be a Chinese pictograph, inkblot style tests, synthesized vowels or ambiguously sized animals. The mechanisms of these studies differ; some seek to prime affect for or against the stimuli, as in Cameron et al. (2012) who used positive and negative primes to influence the affective evaluation of Chinese



pictographs. Others have used primes as a type of anchor for people to base their following judgements from, such as Herr (1986) who used extreme exemplars to prime animal size judgements. Looking ahead to the experiments, where voice will be used as a prime for ambiguous targets, both the affect, and the potential for the voices to be used as anchors (where an anchor is a starting point upon which subsequent judgements are based), are important considerations. The idea of primes as anchors is returned to in Section 6.2.

Many priming studies have used reaction times as their measure of priming effects (see for example Dijksterhuis et al., 1998; Cesario et al., 2006; Levy, 1996; Kouider & Dupoux, 2005). The speed at which a participant presses the button (or whatever the task is) is used as a proxy for the accessibility of a construct, or associations of constructs, for participants. Dijksterhuis et al. (1998), for example, used reaction time measures to assess how quickly participants accessed the construct ‘intelligence’ when they were primed with either *professor*, or *Einstein*. Both primes resulted in quicker reaction times for intelligence congruent traits, but in the Einstein condition, when a *stupid* related trait was preceded with a self-concept word, reaction times were also quicker. The authors posited that participants in the Einstein condition performed a comparison to Einstein which led them to associate themselves more with traits related to stupidity. In a similar vein, Levy (1996) has used reaction time measures to test the accessibility of *wisdom* with elderly word primes. As part of her experiment, she primed participants with senility or wisdom related primes, and then had them read a story about a 73 year old lady called Margaret. Like the ‘Donald Story’ (Higgins et al., 1977, see Section 3.2.1), the story about Margaret could be interpreted in different ways. In line with Levy’s predictions, participants primed with senility viewed Margaret as more dependent, forgetful and senile than those in the wisdom-prime condition.

## 2.4 Perceptual Fluency and Mere Exposure

An interesting area of research, and one that both relates to much of the aforementioned research and which bears on this thesis, is perceptual fluency, the “The subjective experience of ease or speed in processing perceptual information”

(Oppenheimer & Frank, 2008, p 2). In Section 2.2 the work of Schwarz et al. (1991) and Gawronski and Bodenhausen (2005) was mentioned, and these are examples of this. This section expands on the implications of these and similar theories in relation to voice priming. It is important to note here that the usage of the word *fluency* here is different to that of everyday English, or in Linguistics. It is better thought of as ease, rather than as pertaining to any type of language ability or native-speaking abilities. If certain voices are perceived more fluently (with more ease) than others, then consequent effects on reaction times or evaluations might occur because of this ease (see Section 2.2). Before looking at some of these studies and the questions they raise, it is worth discussing Zajonc's 1968 pioneering paper on the phenomenon of mere exposure. Zajonc (1968) defined mere exposure as "A condition which just makes the given stimulus accessible to the individual's perception" (Zajonc, 1968, p 1). By having repeated brief exposures, an individual should then have an increased liking for the stimulus. This has been shown from a number of different perspectives and its robustness and reliability has been noted by other researchers (e.g., Bornstein & Dagostino, 1992). Bornstein and Dagostino (1992) showed such exposure effects with polygons and black and white faces; Russell (2003) looked at increased brand preference after brief product placement exposure; and Hilsenrat and Reiner (2011) looked at haptic preference for textures after subliminal exposure to rough and smooth surfaces.

It has been demonstrated too that when participants have repeated exposure to the same stimulus, they can potentially misattribute their feelings of familiarity, or fluency, for other attributes such as liking or clarity (Reber, Winkielman, & Schwarz, 1998). This has been demonstrated with ostensibly irrelevant, abstract stimuli such as font types. Song and Schwarz (2008) conducted a study that presented participants with a written exercise routine. In one condition, this was written in *Ariel* font, in the other the cursive *Brush* font. As the authors predicted, they found that participants felt the exercises would be both quicker and feel more fluent if they had read them in the Ariel font. They repeated this effect with a cooking recipe written in easy and non-easy font. Again, they found that participants estimated the dish would take less time to create and that they would be more willing to make it in the easy font condition. Huang, Song, and Bargh (2011) found that reading a market forecasting trend in a more readable font increased perceptual fluency, in turn influencing participants' perceptions as

to whether or not that trend was likely to continue. If we take the results of these various researchers, they lead to a prediction that a more familiar voice would lead to greater liking. This, in turn, should lead to some differences in reaction time, the direction of which would depend on the motivations of the participants. Whilst misattribution certainly factors into attitudes and judgements, it is possible that in many instances the increased exposure to something really does make it more likeable, and that this liking is not an artefact. The discussion section returns to the idea of perceptual fluency in relation to the results from this thesis.

## 2.5 Automaticity

Automaticity refers to control of one's internal psychological processes by external stimuli and events in one's immediate environment, often without knowledge or awareness of such control (Bargh & Williams, 2006, p 1)

The distinction between priming (See Section 2.2) and automaticity is not always clearly delineated. There are however practical differences that separate them, and their usage in the literature typically follows these practical differences. Automaticity is usually applied to actions - it is the effect, whereas priming typically refers to the procedure which may or may not produce those effects. If behaviour is to be called the action or reaction to something, then it should be acknowledged that behaviour does not exist independently of the situations and actors involved. Importantly however, these individual and situational influences may proceed without any conscious awareness, or even any cognitively motivated processing. Indeed they could operate from perception to behaviour without mediation. The literature presented on priming thus far presumes some type of mediating force. Automaticity does not require this. As such, it could be that any voice priming effects found in the present work happen directly from perception to behaviour, and this is something I remain open too. As an example, hearing an elderly voice could slow participants down regardless of their age, or views on ageing, simply because it activates a behavioural script. This is a theory of behaviour that has been forwarded mostly by John Bargh (Bargh, Chaiken, & Pratto, 1992; Bargh, 1994; Bargh & Chartrand, 1999; Bargh & Ferguson, 2000; Bargh et al., 2012). This view of behaviour is not universally accepted, particularly within cognitive

traditions. It is not, however, an isolated view. Eminent Social Psychologist Walter Mischel, in his chapter in 'The Automaticity of Everyday life' (R. S. Wyer, 2007), starts by asking "Was the Cognitive Revolution Just a Detour on the Road to Behaviourism?" (Mischel, 1997, p 181). He was cautious on the subject, but nevertheless concluded that Bargh presented a convincing argument for the situational importance of behaviours, but that it was still too early to discount the cognitive revolution. What is perhaps more likely is a re-conciliation between the two, the type of which Bargh alludes to in Bargh and Ferguson (2000). This is encapsulated by the following quote from that paper:

Our position is that psychologists studying higher mental processes should continue the scientific study of conscious processes but at the same time give appropriate attention to the deterministic philosophy that must underlie such analysis (Bargh & Ferguson, 2000, p 940).

Almost three decades ago, Johnson and Hasher (1987) took stock of developments in psychology, particularly relating to memory and cognition, and the likely future directions in the field. They wrote that "There has also been a swing from interest in deliberate strategies to interest in automatic, unconscious (even mechanistic!) processes, reflecting an appreciation that certain situations...seem not to depend much on the products of strategic, effortful or reflective processes" (M. K. Johnson & Hasher, 1987, p 655). Going forward to the next decade, John Bargh's 'The Four Horsemen of Automaticity' (1994) brimmed with reviews of studies that have indeed followed in that direction, and in many cases ventured much further in automaticity research, much of it by Bargh himself. Bargh identified three distinct subsets of automaticity: preconscious, postconscious and goal dependent. Pertinent to my work is his assertion that preconscious automaticity "includes chronically accessible trait construct influences on social perception, because they occur without intention and even uncontrollably..." (Bargh, 1994, p 4). The existence and effects of pre-conscious automaticity have been shown numerously (see for a review Dijksterhuis & Bargh, 2001). It is important here to note that more recent research has demonstrated that whilst the activation of certain constructs proceeds automatically, the evaluation of the stimuli is not necessarily goal independent. This was nicely demonstrated by Sherman, Rose, Koch, Presson, and Chassin (2003) looking at the activation of smoking constructs. When smokers in their study were experiencing withdrawal of

nicotine, they more positively evaluated smoking related stimuli.

Förster, Liberman, and Friedman (2009), in their chapter on priming, wrote that it is now generally accepted that exposure to social traits and characteristics, and stereotypes which contain these traits, can lead to behavioural assimilation to these traits, and that this has been shown abundantly. The natural progression of such an argument is that everyone should be assimilating, across time, place and situation, both behaviourally and linguistically. In other words, social classes would fall, race would cease to be important, dialects would die out and there would be an adoption of a sort of ‘new world order’ of homogeneity. This is clearly not the case, likely in part because of contrast and divergence in behavioural priming. For all the wealth of research showing prime-behaviour congruence (see for example, Bargh et al., 1996; Levy, 1996; Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000; Kawakami, Young, & Dovidio, 2002; Hundhammer & Mussweiler, 2012), there is a large body of evidence showing either an absence of assimilation, or behavioural contrast.

As described for spreading activation and priming processes in Sections 2.1 and 2.2, different researchers have proposed different mechanisms for contrast effects. There are motivational accounts, for example the motivation to overcome stereotype threat (where an individual feels threatened that they will conform to a typically negative group stereotype) (Spencer, Fein, Wolfe, Fong, & Duinn, 1998), and motivation to avoid death thoughts (Arndt et al., 2002). There are contrast and comparison accounts, where individuals compare, and in some cases contrast themselves with particular concepts and stereotypes such as in extreme comparisons of hostility with Hitler (Herr, 1986), or intelligence with Einstein (Dijksterhuis et al., 1998). There are affiliation accounts, for example by Lakin and Chartrand (2003), who showed that mimicry was significantly decreased when there was no goal to affiliate, and Yabar, Johnston, Miles, and Peace (2006) who showed that behavioural mimicry was mediated by group membership. Johnston (2002) demonstrated too that mimicry (previously assumed to be automatic) could be avoided if the to-be-mimicked target had a stigma related to task behaviour. In her study, the stigma was obesity and the task was the consumption of ice cream. Participants ate less in the presence of an obese confederate, compared to when that confederate had a task-neutral stigma

such as a birthmark. Goal dependency also affects the direction of priming. Strahan, Spencer, and Zanna (2002) have shown that drink consumption after being primed with thirst related concepts was dependent on participants' initial thirst levels. It is beyond the scope of this thesis to test for all of these different mechanisms, this research is highlighted because each of these accounts could be partly, solely, or not at all responsible for any effects found. Which of these seems most reasonable is highlighted in the discussion sections for each experiment.

To finish this section, it may be clarifying to highlight the following quote from Robert Zajonc, written in 1980. He wrote that:

Somehow we have come to believe, tautologically, to be sure, that if a decision has been made, then a cognitive process must have preceded it. Yet there is no evidence that this is indeed so. In fact, for most decisions, it is extremely difficult to demonstrate that there has actually been any prior cognitive process whatsoever (Zajonc, 1980, p 155).

The literature presented so far should highlight that whilst there is much research to support this view, the past few decades has yielded work that suggests it is too extreme a position. There are circumstances when an automatic account, bypassing any cognitive mediation, is not sufficient. Whether or not it is required is expanded on in the overall discussion. Moving closer to the experiments in this thesis, these findings are presented to reflect on which of the two accounts, automatic or mediated, are more suited to the effects of voice on behaviour, but in the full knowledge that these debates are ongoing and that the primary aim of this thesis is to show a stimulus-response relationship between voices and individual behaviour, and following from that, that models of language change and variation may need some expansion.

## Chapter 3

# Perceiving Others

People's outputs (evaluations, impressions, memories) are shaped and guided by their knowledge and pre-existing beliefs about the social world (Macrae & Bodenhausen, 2001).

Moving from the macro topic of priming, we now move down a level to the particular constructs utilized in this thesis. There exist a number of ways to prime with voices, but this thesis presupposes that listeners hearing individual voices have already-stored social information about similar voices, and it is this social information that is expected to affect their behaviour through the processes described so far. The voices, then, serve as a proxy for category membership, where the categories should be easily recognisable to the participants. So in the funnelling of the literature review towards the experiments, this section describes the types of social information people have, how we categorize others, and how a voice could activate this knowledge to enable priming. Gordon Allport more than fifty years ago wrote that "The human mind must think with the aid of categories...we cannot possibly avoid this process. Orderly living depends on it" (Allport, 1954, p 20). This is a view that has largely persisted in work on categorical person perception (see Macrae & Bodenhausen, 2001; Fiske, 2005). It is also the hinge upon which much of this thesis rests, no more so than for schematic processing and stereotyping which is expanded on in the following sections.

Save for the few of us that choose to live solitary lives, and those that do so out of no choice of their own, we humans exist in groups: groups of work colleagues, groups of ethnicities, groups of sexual orientations, groups of like-minded hobbyists and so on. For some time, writers and researchers have noted that not only

does our static environment have an impact on our behaviours, but so too do the interpersonal and social environments within which we operate. The stoic doctrines purported that ‘The wise man will not live in solitude, for he is naturally made for society and action’ (Stoic Doctrines, 7.123). This view has persisted:

‘Because our evolutionary heritage provides us with genetic material open to forces and influences from the physical environment, we also require a social environment for brain development and for the acquisition of skills such as speech and written communication. As physical and social creatures, human beings throughout their existence as a species have formed into groups’ (Russet, Starr, & Kinsella, 2010, p 51)

In this section I introduce some theories and models that researchers have put forward to help us explain how we perceive, categorize, respond to and store information about others. The nature of this thesis is not to offer support for one or other of these theories, but as has been explicitly stated, to provide some evidence that voices can impact behaviour over and above speech content and form. That said, looking at theories of social categorization helps lay the groundwork for understanding both the rationale for the current experiments, and for the discussion more generally.

### 3.1 Social Representations

“Theories of mental representation are inherently metaphorical” (R. S. Wyer, 2007, p 287).

In the following sections, and indeed throughout other areas of this thesis, certain knowledge structures are invoked, namely stereotypes and to a lesser degree schemas. It should become clear through reading that the work on stereotypes and schemas refer not to particular structures in the brain, but a convenient shorthand that serves a theoretic utility. Gordon Allport, writing on the history of social psychology originally in 1954 and reprinted in the 1984 edition of *Handbook of Social Psychology*, wrote of social psychologists at the time that they “tend to hold their postulated units in a tentative, heuristic way - as mere constructs - and not regard them as externally fixed in the person or in the species” (Allport, 1985,



p 34). More simply put, psychologists use these terms as a convenience, albeit a convenience with theoretical and practical sensibilities. They are not to be taken as concrete entities (see also Smith and Queller (2001)). Examples of this view can be found in, for example, Fiske and Taylor (1991) and Macrae and Bodenhausen (2001). Fiske and Linville have written that, “Believing in the cognitive contents of the black box is partly an act of faith and partly a theoretical convenience...at this most fundamental level, like religion, the basic premise (schema) is essentially unresearchable” (Fiske & Linville, 1980, p 552). In a more current review article, ‘Schema Theory Revisited’, McVee, Dunsmore, and Gavelek (2005) noted the resilience of the theory, and in a similar vein to Fiske and Linville and Neisser (1976), spoke of its utility. The same will be true of stereotypes, which are covered in the next section. In what follows I describe some different accounts of how stereotyping works, the types of moderating factors that influence their activation and application, and how they pertain to priming. In the remainder of the thesis the usage of the term stereotype should be taken as a shorthand that serves some utility in describing fairly automatic social categorization procedures that pertain to the priming experiments. Much of the work on stereotyping bears direct relevance to the present work, with the crucial difference that I do not set out to test a particular theory of stereotyping, nor settle any existing arguments in the stereotyping literature. What is referred to where stereotypes are invoked might better be thought of as *some knowledge structure about a group*, and the same is true of schemas.

## 3.2 Schemas and Stereotypes

### 3.2.1 Schematic Thinking

We create meaning from our surroundings, and in making sense of the world we develop mental short cuts, heuristics, and stereotypes to help untangle the complexity. A schema is a cognitive theory of organized prior knowledge. It is abstracted from specific encounters and instances, and available as a guide for new information (see Fiske & Linville, 1980). Schema and schemata became popular topics in the 1980s (see for examples S. L. Bem, 1981; Fiske & Linville, 1980; Carver, Ganellen, Froming, & Chambers, 1983; Hintzman, 1986) and much contemporary research presupposes that there are organized systems of prior

knowledge which we use to guide our processing of new information. Dovidio and Gaertner have written that, “essential for efficient functioning is the ability to quickly and effectively sort out different objects” (Dovidio & Gaertner, 2010, p 1089 ). Voice is such an object and should be subject to the same rules as other types of social information. When we hear a voice, a particular schema, which could be the category for which the voice is a proxy or the immediate contextual schema, serves to direct and filter the incoming information so as to make the task of interpreting the voice, or the situation, an easier task. I conceptualise schemas as maps. Just as maps show the general characteristics of an area, and the main features one might find there, the schema performs this role for individuals and groups. The maps are not however created by an individual observer sketching out the characteristics they observe. They are the result of many subjective encounters with the same stimuli (see Markus & Zajonc, 1985). This can happen either individually (such as when a child quickly develops a schema for mammal, which contains feature and spatial information) but also collectively, such as when a nation develops a schema for themselves and other nations. Macrae and Bodenhausen (2001) provide a more thorough treatment of schemas, the various metaphors different researchers use to describe socio-cognitive functioning and unresolved issues related to categorization.

Certain individual schemas are discussed in relation to the current experiments, such as in Section 5.9 on gender schemas. Neisser (1976) emphasizes the anticipatory function of schemas, insofar as they exist in order to facilitate the processing of incoming stimuli. When a certain type of stimulus, such as a smile, is very quickly perceived as cynical by one person, and shy by another, it is not unreasonable to assume that some anticipatory mechanism was ready for it. The speed by which these immediate affective judgements are perceived make it unlikely, in Neisser’s view, that the perceptions are built up, piece by piece, by the perceiver. He further writes that it acts both as a format, and a plan (Neisser, 1976, p 55). To give an analogy, when we see a crocodile in a field, we format the information to conclude that we are indeed seeing a crocodile in a field. If we see the same crocodile inside a museum, we may conclude that it is an artistic exhibition, likely artificial. However, the planning aspect also tells us that when we see the crocodile outside, we should be making plans to avoid harm, possibly fleeing. In the museum, we might plan instead to inspect the crocodile more closely, or perhaps

to photograph it. In the context of voices, as will be used in this thesis, we can highlight this process using the Asian voice as an example. When a listener hears a distinctly Asian sounding voice, one's Asian schema (and/or context schema) would then direct the listener's attention to Asian relevant and related features. This could be a stereotype, or perhaps a relevant heuristic. Once the listener has brought attention to bear on the voice, and the schema and/or stereotype has become active, further connotations with the voice and its speaker or social group become active through a spreading activation processes as mentioned in Section 2.1.

It is important when dealing with schematic explanations to maintain focus on the situational context, and how one's participants construe the situation. Higgins et al. (1977) conducted a classic study showing that when participants were pre-exposed to trait information, it affected how they interpreted events in a following, ostensibly unrelated story. The participants were exposed to pairs of traits, such as *cautious* or *fearful*, prior to reading a story about a man named Donald. The story spoke of Donald's climbing and adventure pursuits. Those participants who had prior exposure to the cautious type traits viewed his behaviour as more cautious, and vice versa those who were primed with risky traits. Srull and Wyer (1979) followed this research again using the Donald story, this time priming participants with *hostile* or *benevolent* traits. Again, prior exposure to particular traits affected subsequent judgements of Donald. Describing a possible process, Herr writes that once the category has been activated, it serves "as a filter through which ongoing events are screened" (Herr, 1986, p 1114). This is in line with Schwarz's (2007) proposition that highly accessible events affect the interpretation of the current situation. When hearing a voice for the first time, whatever version of the stereotype, or exemplar, a listener activates will serve as this filter. In the experiments that follow, whether or not they activate a stereotype, or a particular exemplar, cannot be determined. If a participant activates an exemplar of their grandfather, who happens to be old, slow, and forgetful, then this would share many attributes with the prevalent elderly stereotype shared by many others. Exactly what aspect of the elderly stereotype an individual activates is not tested in this thesis. The prediction is, however, that what is activated should differ for younger and older participants, with the caveat that even within these groups certain individuals would activate exemplars quite unique to themselves.

### 3.2.2 Stereotypes

Stereotypes and the normative practices they purport to capture are not objective facts but social constructs that people can disagree and negotiate over” (Hogg, 2010, p 1116).

A stereotype is, technically, a particular type of schema pertaining to social groups (Hilton & von Hippel, 1996; Dovidio & Gaertner, 2010). In practice and popular usage it is however more often thought of as a biased and unfair representation of whatever group it is being applied to. This is not the type of usage referred to in this thesis, and popular accounts of stereotyping in the media do little to help clarify the distinction. In the introductory chapter to Macrae et al.’s *Stereotypes and Stereotyping*, Stangor and Schaller (1996) give an account of the different approaches to studying stereotypes, and how they differ with respect to underlying assumptions of representation and mechanism. The definition used for this thesis is much the same as Correll et al., who wrote that stereotypes are “category-based generalizations that link category members to typical attributes” (Correll, Judd, Park, & Wittenbrink, 2010, p 46). Like a schema, a stereotype not only contains information, but can affect how incoming information is processed. When the motivation to process is low, or one is mentally fatigued, incoming social information is more likely to be perceived in line with the stereotype, rather than as a situation or person specific behaviour (Spencer et al., 1998; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004).

There is an important distinction in stereotype research which is maintained as we move through the experiments of this thesis. It is the distinction between stereotype activation, and stereotype application (see for examples Fiske & Neuberg, 1990; Gilbert & Hixon, 1991; Kunda & Spencer, 2003; Krieglmeier & Sherman, 2012). The experiments in this thesis were designed to activate some type of category based trait associations (the particulars of which are outside the remit of this thesis but include stereotypes), and then assess how that information was applied to a specific task. Research has shown that people can automatically activate and infer traits from encounters with others (e.g., D. E. Carlston & Skowronski, 1994), and it has more recently been proposed that these trait activations may serve a functional purpose to interact with, or to avoid, the other

person (Crawford, McCarthy, Kjaerstad, & Skowronski, 2013). We may easily speculate how this might also relate to voices, such that certain traits are activated from voices which then serve as a functional cue for task behaviour. This point is expanded upon in the discussion.

Krieglmeyer and Sherman (2012) provide a good overview of these two processes - activation and application. Of stereotype activation they note that what is activated depends on factors such as the prototypicality of the group member serving as the cue, and the learning experience of the perceiver. Regarding prototypicality, this is a recurring theme not only in research directly on stereotypes, but also on priming. Prototypicality as a mediator of stereotype activation can be seen in Kahneman's representativeness heuristic research (see Kahneman & Frederick, 2002), Dijksterhuis's work on contrast priming effects (Dijksterhuis & van Knippenberg, 1998), Hummert's work on elderly and young stereotypes (Hummert, 1990), Livingston and Brewer's work on racial stereotyping (Livingston & Brewer, 2002), and Mussweiler's prototypical stereotype behaviours in relation to physical movement (Mussweiler, 2006). As an example, Livingston and Brewer (2002) found that the negative effects typically found in response to racial priming with black faces were weaker when the black faces were less prototypical, which is to say, less black looking. Taken together these different researchers indicate that the higher a stimulus is in prototypicality, the greater the possible priming effects. In the experiments in this thesis, then, the voices used were deliberately prototypical for their category. The gay voice was easily identified as gay, the Asian voice as Asian, the female voice as female and so forth. Prototypicality and categorization is perhaps best known through the work of Eleanor Rosch (Rosch, 1978), who like Fiske in her discussion of schemas and Kahneman in his discussion of heuristics, emphasizes the role of cognitive economy in categorization. These are just some of the works within psychology that reference prototypicality as a moderating factor in stereotype activation.

The second part of activation concerns the experience of the perceiver, and in this respect I work within the framework outlined by Neisser, who stated "Perception itself depends on the skill of the perceiver - on what he knows in advance" (Neisser, 1976, p 13). This episodic account (see e.g., Tulving, 1972; Hintzman, 1986) (also named exemplar account), concerns the social history of

the individual. This thesis relies on particular voices activating some associated knowledge of the categories for which the voices act as proxy. Hearing a Chinese voice in an experiment, for example, should bring to mind past experiences of similar voices for my participants, the people who spoke them, the situations they were in and so forth. Further, once this has been achieved, it is expected that other traits and ideas associated with those memories also become more accessible, through the process of spreading activation. Much space will not be devoted to the theories on episodic memory other than to outline some of the main features of it and how they pertain to person categorization and stereotyping. Logan (1988) offered a trace account of episodic memory, where every encounter leaves a trace, even if it is a near replication of a previous event. In his account, he refers frequently to *automaticity*, which is discussed in Section 2.5. Other researchers have proposed that the effect is cumulative, such that near-identical repetitions strengthen the trace. For further discussion of these two competing hypotheses, and some existing debates in memory research, Hintzman (2011) is a good reference.

During the encoding stage, the attention paid to the stimulus moderates the ease of future recovery such that divided attention results in reduced recall ability (Jennings & Jacoby, 1993; Iidaka, Anderson, Kapur, Cabeza, & Craik, 2000). In relation to schemas and stereotypes, this episodic memory encompasses the past encounters people have had with others, and helps not only to categorize individuals but also to prepare one for future encounters with similar individuals. As such, the individual encounters the participants have had with either real people from the different social categories, or media exposure to those categories, will at some level make them more prepared for future encounters with these categories. Even single instances have been shown to affect consequent tasks, such as in spelling tasks (Jennings & Jacoby, 1993), pronunciation (Goldinger, 1998) and picture naming (Carroll, Byrne, & Kirsner, 1985). These episodes could be eating chicken feet in a Chinese restaurant, being stuck behind an elderly driver on a free-way, or being refused a room in a hotel for being a gay couple, to relate some examples to the voices used in this thesis. These are easily recalled because certain facts about the encounters, such as heightened emotion, disgust, impatience and so forth, can pique attention which helps encode the episode more concretely. There is not yet a definitive answer to how memory works, let alone whether there are particular sub-types of memory that operate according to different strategies.

Of the field, Hintzman notes that “Unlike the hard sciences such as physics, the psychology of memory has no generally accepted conceptual framework to guide it” (Hintzman, 2011, p 255).

### 3.3 Level of Explanation

‘The special features of social phenomena make it premature to build models of social behaviour at the same level of detail as is possible with purely cognitive research’ (Fiske & Linville, 1980, p 549).

In the prelude of this thesis, it was expressed that my own particular interest is in showing that voices can affect subsequent behaviour. Moreover, it is to a greater degree the effects I am interested in, just as B.F. Skinner was more interested in pecking behaviour than he was in what type of pigeon food produced better results, or whether pigeons with low motivational states were apathetic to levers. In this sense my concern is somewhat behaviourist, but only in interest, not philosophically. One cannot give a full account of priming in relation to social knowledge without understanding that the creation of a social situation, and the behaviours contingent on that situation, exist in a perceptual loop in the mind of the perceiver. As far as socially motivated behaviour goes, mental constructs must play their part. How these are specified, and to what purpose, should be guided by the research question. Much of the research described thus far has been to show that there are processes intrinsic to a person, or at the least shaped by their own individual histories, that affect how they will respond to social information. The schema, discussed more in Section 3.2.1, is a clear example of a cognitive account of mental functioning. Methodological issues aside, a glance at the number of times the word ‘cognitive’ appears in *Personality and Social Psychology*, for example, shows an ever increasing use of the term. I mention this not out of protest, but because we need not always appeal to cognition for its own sake, particularly when that appeal is overly speculative. To explain why the pigeon pecks the lever, it can sometimes be enough to know that the pigeon does this because it receives food. Similarly, to know that a person will work harder when promised a greater reward, it can be enough to know that this reward will produce the behaviour. To link this back to my own work, if it is found that a person judges a photograph as more gay looking on account of just hearing a gay voice, we can treat the gay

voice as a driver of behaviour, rather than a more mysterious cognitive construct.

Fiske and Linville (1980) called upon social psychologists to remember that social phenomena are qualitatively different to the theoretical domains of cognitive and other branches of psychology. The behaviour of the retina in response to flashed light, or the limitations of short term memory in a laboratory number task, are not akin to feeling stereotyped, or feeling dislike for a member of an outgroup. Because of these differences, the theories used in social psychology (and other branches of psychology) are often those with utility, rather than specificity. The same is true for the theories and frameworks that support this thesis. They serve some utility, and they are a heuristic.

James Dinsmoor, writing in *The Philosophical Legacy of Behaviourism* (Dinsmoor, 1999), had much to say not only on the widespread influence of behaviourism, but also widespread misconceptions of what it was and continues to be. He also mentioned that it was Ulric Neisser's *Cognitive Psychology* that first brought the modern idea of cognitive psychology to the forefront, and indeed it was Neisser's subsequent book *Cognition and Reality* (1976) that inspired me to think more clearly about the differences (and similarities) between a behaviourist and a cognitive approach. He writes from a personal perspective of having lived through the transition from behaviourism to cognitive psychology, and of the bandwagonism of psychologists claiming cognitive approaches. One of the main differentiating features of a cognitive approach is to appeal to mediating constructs, however as Dinsmoor (1999) noted, many behaviourists had been doing this. Edward C. Tolman, the eminent behaviourist, even spoke of cognitive maps and representations (Tolman, 1948). Dinsmoor concluded his foreword in the book with the following caustic remark about the perils of careless usage:

As far as I have been able to determine, the word cognitive serves primarily as a magic wand that transforms mice into horses, a pumpkin into a stagecoach and Cinderella's rags into a gown for the ball. It is like a vial of holy water that one can sprinkle over one's manuscript (Dinsmoor, 1999, p 5).

I certainly do not want to be guilty of this in this thesis, and for this very reason have purposefully been careful when using the term *cognition*, and constructs relating to it.



## Chapter 4

# Language and Behaviour

In the prelude, I stated that my interest is in the interaction between non-linguistic behaviour and speech. Whilst this thesis concentrates primarily on voices, there has been a wealth of literature to inform us on how language interacts with behaviour. Much of what we have learned about the language and behaviour interaction is pertinent to this thesis and is discussed in this chapter. Although the research discussed thus far was primarily concerned with physical behaviour, it is necessary to widen our definition of behaviour for the following discussion to include linguistic behaviour. How does one's language both influence, and become influenced by, the language of others? This becomes particularly important if indeed voice can be shown to prime behaviour, as it would suggest that current models of language variation, which make no reference to automatic priming, need updated. I outline some of the theories of style variation below so that, after a presentation of the results of the thesis, hypotheses can be proposed for future work addressing how speech might be affected, over time, by priming.

### 4.1 Style, Convergence and Divergence

Hymes (1974) spoke of a speech community as being composed of different styles. Style, for him, was a way of speaking. Ervin-Tripp defined style as “co-occurrent changes at various levels of linguistic structure within one language” (Ervin-Tripp, 1972, p 235), and Bell defined style as “what an individual speaker does with a language in relation to other people” (Bell, 2001, p 141). Bell went on further to state that style is individual in nature, person centred, and not largely preoccupied

with structures and systems. Whatever approach taken to investigating style, they share in common a belief that an individual speaker is not constrained by one particular way of speaking. If style is affected by those around us, then how we categorise those in our social environment is significant and is likely to feed back into our own linguistic style.

If it is possible that we might adapt our speech to suit a task, and not just our situation and interlocutors, then research on how individuals from the same category vary their speech is an important starting point. Early work on intra-speaker variation (see for examples Labov, 1972; Bell, 1984) saw changes in speech style as essentially reactive (for a further discussion see Schilling-Estes, 2008). William Labov looked at speech on a casual/formal continuum, showing that our language behaviour can change in response to the formality of the environment, and the formality of our interlocutors. A limitation of this approach was the assumption that attention correlates neatly with formality. By way of a personal example, as a speaker of Standard Scottish English, when shifting into Glaswegian slang I am required, at least subjectively, to pay more attention to my speech than when speaking more formally. Similar ‘performance’ issues have been highlighted elsewhere, from Nikolas Coupland’s Cardiff study (Coupland, 1980), to Natalie Schilling-Estes’ focus on performance of the Ocracoke brogue (Schilling-Estes, 1998). It also raises the question that Bell (1984) raised on his own work, on the ratio of social to stylistic variation, and how to account for variation when style variations outweigh social differentiation.

Bell’s (1984) audience design model was however similarly reactive in approach, assuming that speakers react to their audience, and manage their speech accordingly. He demonstrated this in his investigation into New Zealand newsreaders - the same newsreaders in the same studio varied their delivery depending on the severity of the story and also their intended audience. Psychology and linguistics have a history of cooperation and Howard Giles’ Communication Accommodation Theory is the product of one such exchange - see Giles (1973) and for a more recent discussion, Giles and Ogay (2007). The theory combines insights from linguistics such as speech rate, with ideas from psychology such as liking and reciprocity. The theory holds that speakers converge, on dimensions of phonology, speech rate, pausing and so forth, in order to reduce

social difference. Conversely, they may also choose divergent linguistic behaviour, in order to create social and psychological difference (see Giles, Coupland, & Coupland, 1991). As an example, I might choose (willingly or not - but we come to this question later) to phonetically converge with working class adolescents in Glasgow for approval, but diverge from them in other situations, such as when holidaying abroad, in order to place distance between my and their identity.

Not well accounted for by formality or audience is variation that occurs within a given situation or group. That is to say, when the audience remains the same, and the situation remains the same, variation still happens (for more work on this area see Schilling-Estes, 2006). Coupland noted the tension between the traditional large-scale empiricist approach to language variation, and the newer ethnographic approaches that place speaker agency firmly at their core and between structure and agency, staticism and dynamism (Coupland, 2010). Central to much of the work in this area is individual identity. Mendoza-Denton (2008), Campbell-Kibler (2006), Eckert (2000), Bucholtz and Hall (2005), and Coupland (2001) all discuss language variation in relation to one's identity. Indeed, Bucholtz and Hall (2005) open their paper with the following remark: "Identity is the product rather than the source of linguistic and other semiotic practices and therefore is a social and cultural rather than primarily internal psychological phenomenon". If identity is central to, created by, and performed with linguistic variation, then what it indexes and creates must also vary. If our behaviour can change as a function of the voice we listen to, as is hypothesized in this thesis, then it is reasonable to suggest that this may operate also in the reverse such that we may change our linguistic behaviour to suit particular tasks, in response to past experiences of using and hearing particular voice styles in particular situations. This last point is discussed further in Section 4.5.

## 4.2 Lexical and Syntactic Influences

I wrote in the previous section that there was a wealth of research showing how our language can be influenced by external forces. The way we structure our language is one area where there are behavioural consequences of the type I am interested in. Lowrey (1998) found that syntactic complexity hindered

a message's persuasiveness for an imaginary breakfast cereal called Bran-New. When negative constructions were used (compare *Bran-New is preservative free* to *Bran-New also contains no preservatives*), or sentences started with conjunctions (compare *Bran-New is widely available* to *Because Bran-New is widely available*), participants were less persuaded by a message to want to buy the cereal. Chebat, Gelinas-Chebat, Hombourger, and Woodside (2003) have also looked at message simplicity and complexity in relation to advertising persuasion. In their study, participants were presented with strong and weak arguments about a skincare product, in a high and low readability condition differing by syntactic complexity. In the high readability condition, part of the advert read as 'It hydrates the skin and leaves no oil on it'. In the low readability condition, this read 'It offers an incomparable tonus, annihilates the dehydrating effects of showering, without leaving any oily film contrary to other brands on the market'. They reported that message simplicity was the more persuasive of the two conditions. Cartwright (1949), in one of the earliest studies on lexical priming, looked at the efforts of the U.S. to get citizens to buy war bonds during World War II. When 'buy war bonds' was changed to 'buy an extra \$100 war bond today', sales of war bonds rose. E. J. Johnson, Hershey, Meszaros, and Kunreuther (1993) looked at car insurance purchase decisions and found that if a higher priced, fully comprehensive insurance was the default offer then people were more likely to believe it was in their best interests to have such a policy. If the default was a cheaper, less comprehensive policy, the more expensive policy was not considered to be the best purchase.

In the experiments of this thesis, one can assume that participants who are either not being paid, or are being paid meagre sums of money, may not expend a great deal of effort on the tasks. They are unlikely to studiously weigh up the statements presented to them, or to carefully study the ambiguous stimuli to come to a conclusion about the age of person or where music is from. This allows the power of the default, conceptualized here as a combination of the particular voice they are hearing and the associated concepts from the category for which it is a proxy, to enter into the decision making process.

E. J. Johnson and Goldstein (2003), in their now famous study on organ donation rates, highlighted this power of default. If one has to opt-in to donate an organ, donation rates are considerably lower than if one has to opt-out. I

mention this because what is proposed in this thesis is akin to the voices in the experiments acting as a type of default and this is set out more more concretely as an anchor in Section 6.2. In Experiment 1, participants are presented with statements, rather than questions - for example, 'The elderly are wiser'. Here too there are defaults: the default in this statement is affirmative, and if participants have no strong opinion on the matter one way or another, would, based on research into norms and defaults, be more likely to agree with this statement than not. This thesis investigates whether voices can affect this process. Research has also shown that the default option, of choosing whatever is presented to you, is the clearest decision making channel and profits from the laziness, mindlessness and decision paralysis of humans (see for example Iyengar & Lepper, 2000; Langer, 1989; Schwarz, 2004). As Gilovich and Griffin have written: "When one cannot be bothered to figure out the right choice, forgets to figure out the right choice, or cannot discern, despite considerable effort, what the right choice might be, the default becomes the choice" (Gilovich & Griffin, 2010, p 558). When we combine default options with the opt-in/opt-out paradigm, the effects can be powerful.

Kahneman and Tversky's work on framing (Tversky & Kahneman, 1981; Kahneman & Tversky, 1984; Tversky & Kahneman, 1986) provides yet more evidence that the same sentiment, puzzle, or statement expressed in different ways can lead to behavioural differences. Tversky and Kahneman (1981) conducted an experiment where people were asked to choose their preferred program of action for the response to a mythical Asian disease, framed either in terms of gains or in losses. The majority of participants favoured the option with a guarantee of saved lives if the outcomes were framed in losses, but conversely, chose the option with the highest guarantee of deaths when the decision was framed in gains. The authors concluded from this, and much research thereafter, that decisions framed in gains prime conservative decisions, whereas decisions framed in losses prime risk taking. Bazerman (1983), looking at negotiation strategies in industry, found that outcomes framed in terms of gains are more favourable to outcomes framed by loss. Much work has been done too on the framing of political and social discourse (see for examples Schön, 1979; Keller-Cohen & Gordon, 2003; Musolff, 2006). I mention these political and social discourse researches as an example because in Experiment 1A, I try to frame the argument not by lexical items, but by voices. Voices should frame the situation and outcomes just like the lexical and syntactic

methods discussed above. A gay voice should frame the statements and stimuli being presented, particularly those where gay themes are in the statement or participants are asked to judge the sexuality of a person, and this should happen too with the other voices used as category proxies. The challenge, however, is the wealth of extra material embedded in a voice that is not necessarily available with words and grammar.

I want to conclude this section by highlighting research that demonstrates that A) the same message can be evaluated differently if presented differently, and B) that people, unless they consciously assess their choices, often act blindly. Levin and Gaeth (1988) found that people evaluate *75% lean* beef more positively than *25% fat* beef, and Linville, Fischer, Fischhoff, and Pryor (1993) found that students evaluate condoms with a *90% protection* rate more positively than ones with a *10% failure* rate. Although these experiments varied numbers and words to achieve the pay-off, it raises the question of whether different voices, linked to social categories with all of their connotations and expectations, might act in similar ways.

### 4.3 Speech and Non-Linguistic Behaviour

I have discussed thus far research demonstrating that language, the way it is put together, and the syntactical and lexical choices we use, can affect behaviour. This could be behaviour as choice, as belief or even physically represented in movement. I now turn to an aspect of language which is of central importance to this study, that of speech. A rich repository of research exists documenting the relationships between a person's speech, and aspects of their social environment (see for example Bucholtz & Hall, 2005; Stuart-Smith, Timmins, & Tweedie, 2007; Mendoza-Denton, 2008; Eckert, 2008) and their personality (see for example Chen, Benet-Martínez, & Ng, 2013; Chen & Bond, 2010; Campbell-Kibler, 2006; Apple, Streeeter, & Krauss, 1979; Allport & Cantril, 1934). Speech, according to various definitions, involves the act of speaking, of conveying some message or other.

#### 4.3.1 Speech, Environment and Behaviour.

In talking about speech environment, I am referring to the immediate external context in which one's speech is being used. A person could be in a bar, a

courtroom, in class, being reprimanded, being congratulated, speaking to children or telling a secret. The speech being used in these situations is likely to differ from the person's standard speech, which is the variety they typically use between friends, family and colleagues and is identified by largely following grammatical and phonetic norms, and it is this difference that is of interest in this section. There are certain features that alert us to these situations. Laryngeal hypertension, for example, is associated with aggressive speech (Laver, 1994). Our ability to hear this hypertension allows us to make the crucial decision on whether a man in the street who shouts 'Come over here a second' is looking for help, or a fight. The care taken to articulate, varying on what Lindblom (1990) called the *hyper-hypo articulation continuum*, can signal whether the speaker's environment is formal or informal. This allows the speaker to portray him or herself in a certain way, but also allows the listener to infer the degree of formality needed on the occasion. Bell (1992) wrote how newsreaders varied their speech along linguistic and paralinguistic dimensions, such as prosody and pitch, to suit the gravity of their story. As such, when we listen to the news, even if we tune in mid-sentence, we have a sense of whether the story is jovial or macabre. One may also betray their anxiety through speech, such as by quickening speech articulation, reducing articulatory forcefulness and using more frequent *ums* and *ers*. When speaking to children, adults may use child-directed speech, characterised by a slower rate of speaking, restricted vocabulary and extra stress on key words (Dockrell & Messer, 1999).

#### 4.4 Voice and Personality

The fundamental quality of one's voice, the phonetic patterns of speech, the relative smoothness of articulation, the length and build of the sentences, the character and range of the vocabulary, the stylistic consistency of the words used, the readiness with which words respond to the requirements of the social environment, in particular the suitability of one's language to the language habits of the person addressed: all these are so many complex indicators of the personality (Sapir, 1933, p 508).

Sapir's lengthy sentence well summarises this section. Moving away from situation specific speech, we also find relationships between a person's personality

and their speech. The extent to which this relationship is real, or imagined, is almost incidental for present purposes. Personality only becomes meaningful, in most situations, when there is a comparison between two personalities. Jean-Paul Sartre expressed this in his essay *The Humanism of Existentialism* (1965) when he wrote that man “can not be anything (in the sense we say that someone is witty or nasty or jealous) unless others recognize it as such” (Sartre, 1965, p 51). One would only know they were dull if others were less so, and what counts as a friendly personality in one culture could be seen quite coldly in another. Even if physiological voice differences were to be found between two groups, the strength of any finding would have to compete with the imagined link, which may be quite different. As Giles and Billings (2004) noted in the Handbook of Applied Linguistics, earlier research looking to empirically link voice with personality traits was somewhat fruitless, but the research into listener stereotypes of a voice was both convincing and abundant. As such, when the participants in the current experiments hear these voices, any associated traits and stereotypes of these voices should be available to them. The stereotypes referred to here are not those which were described previously, linked to social categories, but more to the voice characteristics themselves and whether listeners believe the voice sounds stereotypically friendly or shy and so forth. Even if those stereotypes have no real relationship to the actual personality traits of the speakers, automatic associations should be active for the participants.

Chen and Bond (2010), in the opening for their 2010 work on personality and language, wrote “The issue of whether personality changes as a function of language is controversial” (Chen & Bond, 2010, p 1). Reviewing what has been already discussed in the previous sections, their statement is somewhat surprising. After all, if so many other aspects of one’s life are open to influence, why not one’s personality? Particularly in light of both the automaticity and attitude research of the past few decades, the idea that personality may differ as a function of *language variety* should seem less controversial. Following on from the work of McCrae, Yik, Trapnell, Bond, and Paulhus (1998), which showed that Hong Kong Chinese bilinguals (English and Cantonese) scored lower on extroversion and openness than North American participants, Chen and Bond investigated whether or not ratings on the ‘Big Five’ personality traits (Costa & McCrae, 1992) were influenced by, in some part, the language used. Administering questionnaires



in both Chinese and English, they found quantifiable differences in allocated personality traits, dependent on language variety. The authors controlled for cultural background using compound bilinguals, and concluded that their findings showed cross-language differences related to dimensions of personality, and that this was perceived both by speaker and listener. This indicates that over and above individual voice stereotypes as mentioned in the previous paragraph, listeners hold trait associations at a social and cultural level for voices too, so that a Chinese voice or a gay voice should have its own trait associations for listeners.

Luna and Peracchio (2001) have looked at the perception of personality as a function of language on advertising messages to bilingual audiences. They showed the same video clip of advertisements to the same group of people, but at 6 month intervals. They found that, for example, the same person could be seen as either risky and assertive, or confused and passive, depending on whether the actor spoke Spanish or English in the advertisement. If certain language varieties make people appear more assertive or risky, then certain voices should do the same. A woman with an older sounding voice might be perceived to be more calming than one who was younger, a gay man more passive than a heterosexual man and so forth. Over and above second guessing a person's motives or sociocultural history, simple dimensions of personality are evidently entwined with voices, and spontaneous trait inferences could be made from these voices (for work on STIs see Crawford, Skowronski, & Stiff, 2007; Crawford et al., 2013).

Demonstrations such as those just mentioned, showing that language and personality have a behavioural link, are yet additional areas of research that make the overall hypothesis for this thesis more propitious. If people come to have not only expectations about a speaker's social background and their cultural values as discussed in this chapter, but also the types of personality that may be associated with that speaker, then from a social psychology perspective it makes sense to question whether the supposed personality of the speaker would translate into convergence or divergence priming in the listener.

## 4.5 A Gap in Sociolinguistics

As previously stated, the aim in this thesis is to explore whether, and in what situations, voices might affect our behaviour, and secondly whether the presence of voice generally results in behavioural differences in a task. In the process of writing this thesis it occurred to me however that if the aims were achieved, it would have implications for sociolinguistic theory. I had been focussed on a proposed speech to behaviour mechanism, but it hadn't occurred to me that given enough exposure, this could operate in reverse, so that over time our behaviours could affect our speech. This is the focus of this section.

In this chapter I have discussed some of the sociolinguistic theories that are offered in explanation for language change, both at the individual level, and at a group level, both in a temporary snapshot, and also across time. Comprehensive accounts of these processes, along with a wealth of examples of the processes, can be found in Sali Tagliamonte's 'Variationist Sociolinguistics: Change, Observation, Interpretation' (Tagliamonte, 2011) and 'Sociolinguistics: Method and Interpretation' (Milroy & Gordon, 2002). For long-term language shifts, there are a number of ways in which the linguistic and phonetic characteristics of speech can change. Whilst there are non-social reasons, such as the principle of least effort, much language change and variation is contact related. An example is when a group's mobility increases, thus increasing inter-group linguistic contact which allows members from different groups to eventually share features that were previously distinct. An example of this is the spread of Estuary English in the south of England, which has features of both received pronunciation (RP) and the London accent, and can be heard within a hundred mile radius from London, replacing more traditional county specific accents.

At the individual, temporary contact level, change is purported to happen for reasons of mutual intelligibility and to facilitate liking (Giles & Ogay, 2007; Shepard, Giles, & Le Poire, 2001). These temporary speech changes can be deliberate and controlled, or unintentional and automatic. At the deliberate level, this can be seen when vernacular speakers attempt to speak more formally in the presence of superiors or people of higher status, or answer the telephone in their *phone voice*. At a more automatic level, accent changes can be the result of perceptual feedback, such that hearing a higher vowel prompts a speaker to

raise their own vowels, or increase their tempo and so forth. In between the deliberate and the automatic is where, I believe, sociolinguistics has the most to say about individual speech variation. One particular line of research looks at the construction of identities through speech variation (Bucholtz & Hall, 2005; Campbell-Kibler, 2006; Stuart-Smith et al., 2007; Eckert, 2008; Mendoza-Denton, 2008). These researchers have argued that the focus must be on individual and group identities, and that these social factors are primary in language variation processes. As a result of this combined research, linguists now acknowledge that people can speak in one manner in one situation, and another manner elsewhere, because they have more than one identity across their social spaces.

Some of the approaches to individual language variation have been briefly summarised. In short, it is typically assumed that if one changes one's accent in the presence of another, it is to be better understood, better liked, or because environmental forces such as noise have influenced the speech production. If, however, it can be shown that different voices that act as proxies for social categories can lead to different behavioural outcomes, then it raises the question of whether *some* voice styles, and voices, might contribute to behavioural reinforcement. By this I mean that if individuals come to associate a particular type of voice, or style of speaking, with a certain behaviour or attitude, then it follows that they would use this knowledge in the future when attempting to produce similar behaviours or achieve similar aims.

Our social landscape is rich in category variation, each of these categories having their own richly layered stereotypes. The LGBT community, the elderly, the youth, different groups of immigrants, and businessmen are just some of the social categories for which most of us have readily accessible stereotypes and ideas about. Each of these categories have certain features in their speech that are specific to them, and it is these features that typically help the listener, along with other non-linguistic features, to categorize the speaker. People store this linguistic information with category and context information, and so when those categories and contexts present themselves in future, this information can be used as a helpful heuristic. Further than this, though, I speculate that people store information about what particular voice style and tone facilitated interaction with certain groups. By way of an example, in Experiment 1 of this thesis I investigate

whether certain voices can affect people's answers to socially sensitive statements. I want to know whether people will give more sympathetic answers to certain voices. If it can be shown that this is the case, then coupled with past research on how speech relates to personality (Allport & Cantril, 1934; Chen & Bond, 2010; Chen et al., 2013), we might speculate that individuals could vary their speech to achieve a specific purpose. This would not to be more liked, or to be better understood, but to gain from the listener a particular opinion or behaviour - to manipulate them, to put it more starkly - although certainly not anywhere near the degree Skinner might have enjoyed seeing. When we call our banks to complain, we use a different voice than when we are not complaining. Similarly, if we use a particular voice when interacting with the elderly, this may have subtle effects on our speech that we are unaware of. This is an idea which is expanded on in the general discussion in Chapter 8.

## Part II

# The Experiments

## 4.6 Introduction to the Experiments

Because psychology is about people, it cannot shirk the responsibility of dealing with fundamental questions about human nature...A seminal psychological theory can change the beliefs of a whole society...This can only happen, however, if the theory has something to say about what people do in real, culturally significant situations (Neisser, 1976, p 2).

Ulric Neisser wrote the above sentiment in his 1976 classic *Cognition and Reality*, expressing the view that too much of psychology had become detached from its subject matter: people. At that time, he claimed the field was lacking in ecological validity, that is, a foundation in reality. Experiments were being conducted that looked at behaviours in laboratories so far removed from the everyday experiences of people, that any claims of real-world relevance were hard to sustain. In a more recent paper, Mortensen and Cialdini (2010) remind us that social psychology is expected to be relevant to the real world. Laboratory work, they argue, helps to pinpoint mechanisms, and theory can help generate research ideas, but only by working with meaningful paradigms and getting out of the lab will we understand the strength and spread of these social effects (Mortensen & Cialdini, 2010). They continue to state that often the variables which psychologists try so hard to control are the ones that carry most of the weight in day to day life. Like Neisser, they point out the multiple interactions between environment and person.

The aim in designing these experiments, then, was very much guided by the above sentiments. My educational background is business, and in business school the majority of what one learns has real world, practical applications. I want to be able to show that any effects found in this thesis could reasonably extend into probable real-world situations, although these would be necessarily constrained. The first experiments, referred to as the *Voice and Values* experiments, are perhaps the most 'ecologically valid': the experiment replicates the common scenario of listening to, and responding to, a disembodied voice, from banking call centres to market research surveys and more. The second group of experiments, on reaction times, are less grounded in day-to-day scenarios, but offer a controlled way of showing that voices can produce behavioural effects.

## 4.7 A Note on Baselines

For all of the studies presented, there will be no ‘baselines’ as they are traditionally conceived. Sidman, in his instructional text ‘Tactics of Scientific Research’, wrote “The baselines themselves will constitute parameters of the phenomenon in question...” (Sidman, 1960, p 317). The phenomenon in question throughout this thesis is the effect of voice on behaviour, and thus baselines will necessarily involve some aspect of voice as a parameter. The exception to this is the silent conditions, which allow me to compare the impact not only of specific voices, but of voices overall. Where comparison levels have been set in the statistical analyses, they have been done so either to provide a clear contrast between certain conditions or to ease the interpretation of the data. It will be clear from the methods sections of each experiment what the comparison levels are, why they are so, and how to interpret the results from them.

## 4.8 Statistical Techniques

This section provides details on the types of statistical tests used in this thesis and the rationale for using them. It is applicable to all of the data analyses that follow, and can be treated as both an overview and introduction to these analyses. Each individual experiment has its own section on how its data were treated, and for each analysis there is a description of the variables tested and other relevant considerations.

The primary tool of analysis in this thesis is mixed model analysis (Baayen, Davidson, & Bates, 2008). For discrete data, logistic mixed effects models are used, and for continuous data, linear mixed effects models. The `lmer()` function from the `lme4` package (Bates & Maechler., 2008) was the primary tool used to run the models. The `lme4` (linear mixed effects 4) package was designed to allow the fitting of linear mixed effects models. Using the `lmer` command, the user can specify a two-sided argument for regression with both fixed effects and random effects. It can be further specified to operate on discrete data by using the *binomial* command. Where individual questions are analysed, general linear models with the `glm()` function are used. There is a growing trend in both psychology and linguistics to use mixed effects modelling for the way in which it deals with

potentially random effects. The main impetus for using this approach over a traditional ANOVA model is the ability to gain more control of subject and item variation. That is to say that rather than dealing with the mean of the response variable, it deals with both the mean *and* the variance for each response. This is particularly appealing when using stimuli that might reasonably be presumed to be assessed differently for different participants, such as ambiguous photographs or opinionated statements. There is also a growing trend to use mixed models in EEG research, where variation between and within participants is much larger than typical items subjected to ANOVA. Baayen, Davidson and Bates(2008) ran a study comparing traditional approaches (including ANOVA) and mixed effects approaches, concluding that the mixed model was more reliable. Judd, Westfall, and Kenny (2012) defined random effects thus: ‘random factors are factors whose levels are sampled from some larger population of levels across which the researcher wishes to generalize, whereas fixed factors are those whose levels are exhaustive’. A further benefit to these types of models is that they are less afflicted by missing data and unbalanced designs (Bagiella, Sloan, & Heitjan, 2000). If a particular participant has a number of missing cases, rather than excluding them as is typical in ANOVA, a mixed effects model will use what data is made available to it. Thus more data is preserved, which should in turn strengthen the predictions.

In the context of the experiments and according to Judd’s definition (above), the participants used for this thesis certainly qualify as random: they are largely (but not exclusively) university students, under the age of thirty, from Canterbury New Zealand. As with most psychological studies, the aim is to generalize from this sample, albeit with important caveats (See Section 8.3).

Both Baayen et al. (2008) and Judd et al. (2012) highlight that participants, and stimuli, are seldom random, instead being typically grouped and pre-selected for particular reasons. One result of this is that “Because the literature is filled with designs where stimuli should be treated as random but are not, we as a field are probably faced with many Type 1 errors, leading to persistent failures to replicate effects when different experimenters use different experimental stimuli” (Judd et al., 2012, p 65). In their 2008 paper, Baayen et al. (2008) introduced a method that allows participants and items (or anything that sensibly warrants being called a random effect) to be entered as crossed, independent random effects.



It also allows events that occur during an experiment to be taken into account. For all of the analyses, individual participants are entered as random effects. In Experiment 3, a reaction time paradigm, the individual trials and visual stimuli items were entered as a random effects, to take into account fatigue, learning effects and prior visuals, all of which are known to affect performance.

The output of lme4 models provides coefficient estimates, standard errors,  $t$ -values or  $z$ -values. The intercept for the logistic regression models is the log odds of the comparison level(s) producing the estimate for the dependent variable. For the linear regression models,  $p$  values were calculated using the LmerTest package in R, which runs the lme4 models through a Kenward-Rogers approximation test to calculate degrees of freedom and resultant  $p$  values. In Experiment 3, which looks at reaction times, the restricted cubic spline (rcs) function was used, which allows trials to be modelled non-linearly. Model comparisons were used to determine the optimal number of knots for the non-linearity, and are specified for each analysis.

I have chosen not to present any standardized effect sizes for my data. First, it is not clear what methods should be used to calculate a standardized effect size, such as R Squared, when using mixed effects modelling. There is ongoing debate on relevant internet forums, in both psychology and statistics, that suggest different and often opposite methods. In much of the literature that uses linear and logistic mixed effects models a standardized effect score such as R squared is not reported (see Nakagawa & Schielzeth, 2013). Baguley (2009) argues for presenting simple effect sizes (in this case, the coefficients) over standardized effects sizes, particularly when the original units of analysis are meaningful. In the experiments utilizing linear regressions that follow, the coefficients are easily understood with respect to their intercept, which is the value of a given comparison level. This is made especially easier by the use of the plotLmer in the function in the LanguageR (Baayen, 2013) package, which plots the estimates onto a probability scale. Baguley goes on to say that the “simple effect size retains more information about the context of the data than standardized effect size” (Baguley, 2009, p 611). A further reason is that the experiments in this thesis use a combination of continuous and categorical data in the analyses, and standardized effects scores are not typically applied to categorical data. In addition to the above, the goal of this thesis is not to predict actual outcomes, where effect sizes would be of greater importance, such

as when predicting whether or not a given medicine results in a particular outcome. Further, it is not to provide corroborative evidence for a known relationship between voice and behaviour. Although the statistical models used here are designed to predict outcomes based on the data, my interest is in uncovering whether there are underlying connections between variables, and a combination of t-values and p-values provide an indication of this.

## Chapter 5

# Experiment 1: Voice and Social Values

### 5.1 Introduction and hypothesis

The overarching hypothesis to this thesis is that voices have the ability influence people's behaviour. Consistent with this macro hypothesis, it is proposed that when people are presented with statements designed to assess their beliefs and values, the voice that presents this statement will act to influence their answers either through assimilation (strong hypothesis) or by contrast (weak hypothesis). This experiment was designed to test whether participants in an online setting, who listened to pre-recorded statements on social themes, would be influenced in their answers by the voice making the statements. I chose to focus on four social themes for which a speaker could be recorded whose voice could be identified as belonging to the group related to the theme: ageing and the elderly, gay issues, women's issues, and ethnic minority issues. As discussed previously, one's speech characteristics can often identify the group(s) that a speaker belongs to, what is referred to as the socio-indexicality of speech (Podesva, Roberts, & Campbell-Kibler, 2006; Bucholtz & Hall, 2005; Eckert, 2008). As such, speakers were recorded who could act in some way as a proxy for a particular social group.

This chapter is split into component sections. Following a short introduction, I will cover the original hypothesis and analyses for Experiment 1A, the replication Experiment 1B, the reanalysis of both 1A and 1B, the comparison with the

non-speech Experiment 1C, ending with a discussion. This should allow the reader to understand the rationale for the original design, and consequently the change which led to the second analysis. The original hypothesis was that when the social theme of the statement, such as gay issues, matched the voice making the statement (a gay male), this would lead to more supportive answers than when the social theme and voice were unmatched. This proposal rested on the idea that participants would, at some level, wish to converge with a speaker as predicted by a category-assimilation account, or simply answer with an idea that is more recent to them as a result of the voice, as predicted by a heuristics and biases account. In short, they were expected to answer positively towards social minorities and avoid answers which could be inferred as offensive to those groups when listening to a voice from the target group. Finally, a silent condition of the experiment is presented where participants answered text only statements, and the results of this are combined with the previous two experiments to show which voices drive the effects more strongly.

## 5.2 Self Presentation

People's attitudes are evidently susceptible to information that is coincident with the expression of those attitudes. This much is uncontroversial. I aimed to demonstrate in the experiments that follow that people's reported beliefs could be influenced by voices. The instability of attitudes has been noted by many authors (see for examples Mischel, 1968; D. J. Bem, 1972; Forgas, 1992; Ferguson & Bargh, 2007; Dasgupta & Greenwald, 2001). If the attitudes are indeed as unstable as these different researchers have argued, then they should be open to experimental manipulation. The tendency for people to present themselves in a manner befitting their audience or their desired identity is at the heart of this instability. By having voices presented aurally, rather than having a physically present speaker, this weakens the potential for self-presentation effects to occur. It also allows the effect of a voice acting as a category proxy to be isolated from the physical features of a person. The advent of the Implicit Association Test (IAT) (Greenwald et al., 1998) allowed researchers to gauge people's attitudes in spite of their self-presentation motivations. It should be clear from the research presented on priming, however, that the attitudes measured in the IAT are themselves open

to external influences, and indeed this was nicely demonstrated by Dasgupta and Greenwald (2001).

## 5.3 Experiment 1A

In this experiment, an online survey was administered to participants in which they heard a selection of statements read aloud. The survey was recorded by five different speakers, and participants were randomly allocated to one of these surveys, identical except for the pre-recorded speaker. After hearing the statement, they were asked to agree or disagree with the statement. I investigate whether the responses given across the five voice conditions differ.

### 5.3.1 Participants

Participants were 245 undergraduate students at the University of Canterbury, Christchurch New Zealand, who volunteered to participate in a language study. It was administered online using the Qualtrics survey platform, and participants were randomly allocated to one of five experimental conditions differing only by the voice presenting it. Course administrators from multiple departments emailed students to ask if they would like to participate. As such, the participants were drawn from a wide variety of subject areas. Demographic information was collected and is given below. No pre-selection was made for determining which students would be accepted and, as such, any student who completed the survey was included for data inspection. There were originally 275 submissions, so that 30 participants were not included. Data were removed where participants had abandoned the survey before the final statement and where participants had entered blank or meaningless answers but continued through to the final statement and pressed submit. Also removed were any duplicates where participants had submitted the survey twice (IP address entries and timestamps were checked). After these exclusions, the make-up of participants was 130 (52%) females and 115 males, with an age range from 18 - 50 years old, and median age of 20 years. New Zealand European was chosen by 84% of participants for their ethnicity, and 91% of participants identified as heterosexual. The number of participants for each voice condition was: 49 for the young heterosexual male voice, 42 for the female voice, 63 for the Asian male voice, 42 for the gay male voice and 49 for the older male voice.

### 5.3.2 Statements

For each of the social themes (age, sex, ethnicity, sexual orientation), a list of socially pertinent statements was prepared and, following input from colleagues, narrowed down to a list of twenty-two statements overall. These statements were designed to test the original hypothesis that a voice belonging to a target group would prime supportive answers for that group. The statements are shown in Appendix A. For some of the statements, supporting the statement (and therefore the group) entailed showing less support for another group. An example would be ‘The elderly are wiser’, where agreement would indicate that the participant believed that wisdom was not a quality of youth. Other statements could be construed very differently by the same macro-demographic group. For example, the statement ‘I am comfortable with same sex marriage’ might appear to be one that gay people would be predicted to agree with more than not. However, many gay people have no interest in joining this historically religious arrangement, whilst others fight vehemently for the right to do so. For this reason, simple agreement with the statement should not be taken as a sign of support for that group, which is highly subjective for many of the statements. The questions were sent to three non-involved research assistants at the New Zealand Institute of Language, Brain and Behaviour, who were asked to code whether or not a *yes* or *no* answer would be seen as supportive. Their answers were tallied and if a majority rated a particular answer as supportive, this was used for the final coding in the analysis. It is important to highlight again that supportive simply refers to support for the particular social group that the question was designed to target, and does not indicate that the participant is a supportive person more generally.

### 5.3.3 Voices

The experimental conditions varied by the voice that presented the statements. There were five conditions, one for each of the different voices. To record the statements, voices were chosen based on the simple criteria that the speakers truly belonged to the social group they were to represent (such that the gay male voice was indeed recorded by a gay male) and that both the researcher, and friends and colleagues of the researcher, could accurately identify the speaker as belonging to

the intended social group. To further ensure that these category memberships were identifiable from the speech excerpts, a ratings survey was administered for the voices. Twenty undergraduate raters from both sexes were recruited via email at the University of Canterbury. They were asked the following questions for each of the voices.

1. Please select whether the speaker is male or female
2. Please estimate the age of the speaker
3. From the following options (heterosexual, gay, other), please select an option you think describes the speaker
4. Please guess the ethnicity of the speaker (e.g. New Zealander, American, Korean, etc)

The ratings results can be seen in the brackets in Table 5.1. Four of the voices represent the particular social groups for which the statements were designed to target - the fifth voice, a young heterosexual New Zealand male, was used primarily as a comparison. The term *comparison* is used here to denote a statistical comparison, where the young heterosexual male differs from the other voices through different category membership.

Table 5.1: Exp 1A: Speaker information

Social Group	Speaker	Speaker demographics and ratings
Age	Older male	Heterosexual (75%) NZ (90%) male (100%) in 50s (average age 48.9)
Ethnicity	Asian male	Heterosexual (90%) Chinese (75%) male (100%) student in 20s (average age 27)
Gay	Young gay male	Gay (70%) NZ (85%) male (100%) student in 20s (average age 24)
Sex	Female	Heterosexual (95%) NZ (95%) female (100%) student in 30s (average age 36)
Control	Younger male	Heterosexual (75%) NZ (95%) male (100%) student in 20s (average age 30.1)

### 5.3.4 Stimuli Creation

All of the voices were recorded using Audacity, an open-source sound recording and editing software package, on a laptop computer using an inexpensive microphone

in a quiet room. It was important to ensure that the clarity, duration and loudness were as close as possible as all of these are known to affect perceptual fluency (Schwarz, 2004). The past contact participants would have with these voices was expected to be quite different, and this is one of the underlying theories of why differences in task behaviour might occur. The recordings were clear enough to detect individual speech differences, such that that listeners could detect that they were listening to a New Zealand female or young gay male, and were all of equal recording quality as assessed by file size and audio inspection.

The first speaker to record (Female 1) was asked to read in a formal and clear manner. The remaining speakers listened to approximately one minute of this recording, and were asked to maintain a similar reading pace, volume, and formality of style. Where necessary, re-recordings were made to ensure that tempo and amplitude were comparable across voices. To achieve similar amplitude, the speakers were asked to read their name, date of birth, and address into the microphone, and where their voice showed significantly higher or lower amplitude on the waveform than the first speaker, the microphone was moved closer or further to the speaker to calibrate this. Audacity was used to inspect the recordings and to remove any unexpected noises such as a sneeze. The recordings are of roughly equal file size (a proxy for recording quality) and content, excepting the voice of the speaker. No further manipulation was performed on the recordings. Qualtrics survey software was used to create and present the survey. The audio statements were uploaded as MP3 files into the Qualtrics library, and embedded in the survey using JavaScript. The order of the upload was the same for each voice condition, but random overall. Along with each audio statement was the option to either agree or disagree with that statement by clicking into a box. After all the statements had been answered, participants were asked for the following information at the end of the survey: their age, sex, sexuality, and ethnicity. Participants worked through each statement in a fixed-order sequence. At the end, they were asked to provide their thoughts on the voice they had just heard (such as comments on ethnicity, sexuality, age and general comments) into a text box. Finally, they had the option to submit all of their answers. This was only possible if all of the questions had been answered. A screen-shot of how the questionnaire was presented is shown below in Figure 5.1. There is a play button, no accompanying question text, and then a forced choice answer format.





Figure 5.1: Example layout of the questionnaire

### 5.3.5 Results: Social Supportiveness

Table 5.2 below shows the mean values of social theme support. Answers that are either supportive or unsupportive are mutually exclusive - thus 77% supportive also entails 33% unsupportive. The mean values indicate that overall, people are more supportive than not towards the various social themes. Regardless of the voice presenting the statements, this support never drops below 50%. Questions relating to the elderly also garner more support than those relating to other social themes.

Table 5.2: Exp 1A: Mean support values

Social Support		
Social Theme	Voice Condition	Mean Support Value
Age support	Heterosexual male voice	0.68
	Asian male voice	0.72
	Female voice	0.77
	Gay male voice	0.77
	Older male voice	0.73
Female support	Younger male voice	0.55
	Asian male voice	0.56
	Female voice	0.61
	Gay male voice	0.65
	Older male voice	0.55
Ethnicity support	Younger male voice	0.54
	Asian male voice	0.53
	Female voice	0.62
	Gay male voice	0.61
	Older male voice	0.54
Gay support	Younger male voice	0.59
	Asian male voice	0.63
	Female voice	0.61
	Gay male voice	0.70
	Older male voice	0.62

A more thorough introduction to the analyses used, and the reasons for their use, was given in Section 4.8. A logistic mixed effects regression was hand-fit using a backwards stepwise elimination. Voice condition, social theme, and participant sex were entered as fixed effects, and participant and questions were entered as

random effects, because for both of these predictors, we can reasonably expect significant inter-subject and inter-item variation. Neither participant sex nor social theme were significant main effects. Interactions between social theme and voice condition, and social theme and participant sex, were not significant predictors. The female voice was set as the comparison level for voice condition as hers was the only female voice, allowing the male voices to be compared more easily. The choice of the baseline (or comparison level) is more of a convenience which allows us to see the effect of a particular item of interest against other items. It does not change the predicted probabilities for answering one way or the other, or the model estimates more generally. For social theme, the baseline was set to alphabetically to *age*, and there was no reason to relevel this.

Table 5.3 gives the predicted support values based on the mixed model analysis. This differs from the actual responses as given in Table 5.2. Compared to the female voice condition, all of the male voice conditions except the gay male voice condition returned significantly lower probabilities of supportive answers. The positive or minus coefficient values indicate whether or not each social theme, or voice condition, attracted more or less support compared to the comparison level which is included as the intercept. Here, this means that all of the voice conditions were associated with less supportive answers than the female voice condition. The female voice condition and gay male voice condition were not significantly different to each other. The model's predictions are shown graphically in Figure 5.2. Figure 5.3 gives an indication of the how participants listening to the different voices responded to the different social themes. This interaction plot was calculated with mean values of support. Participants listening to the control voice, the younger heterosexual male, were associated with the lowest probability of socially supportive answers. No fixed effects or interactions were found for participant sex, age, ethnicity, and sexual orientation.

Table 5.3: Mixed model output - modelling predicted social support by voice condition and social theme

	Estimate	S.E.	<i>z</i> value	<i>p</i> value	
(Intercept)	0.89	0.20	4.29	<.001	***
Asian male voice	-0.36	0.11	-3.09	.001	**
Gay male voice	-0.07	0.13	-0.58	.556	
Older male voice	-0.34	0.12	-2.77	.005	**
Younger male voice	-0.49	0.12	-3.93	<.001	***

The results from this analysis did not then support the assimilation hypothesis of voice and category matching, either for significantly more supportive or significantly less supportive answers for those categories in which the voice acted as proxy. The macro hypothesis, that voices would influence the responses, is however consistent with these results.

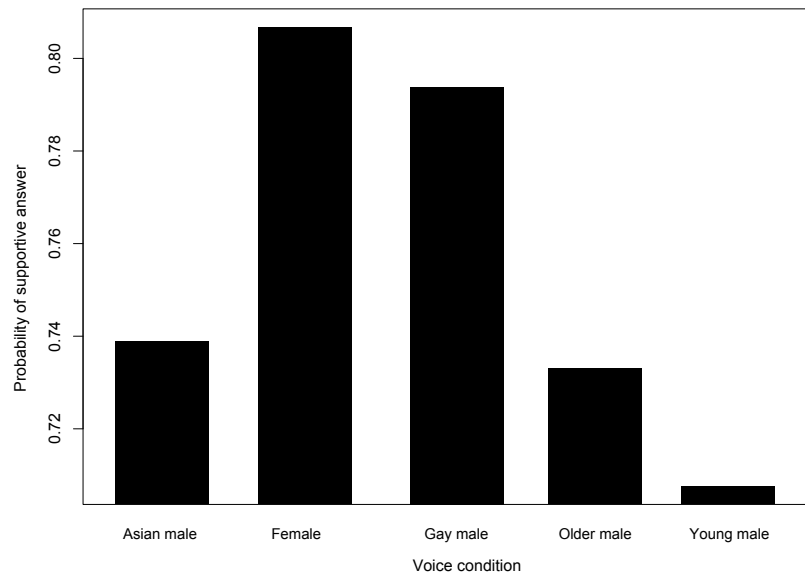


Figure 5.2: Exp 1A: Predicted support for each voice condition

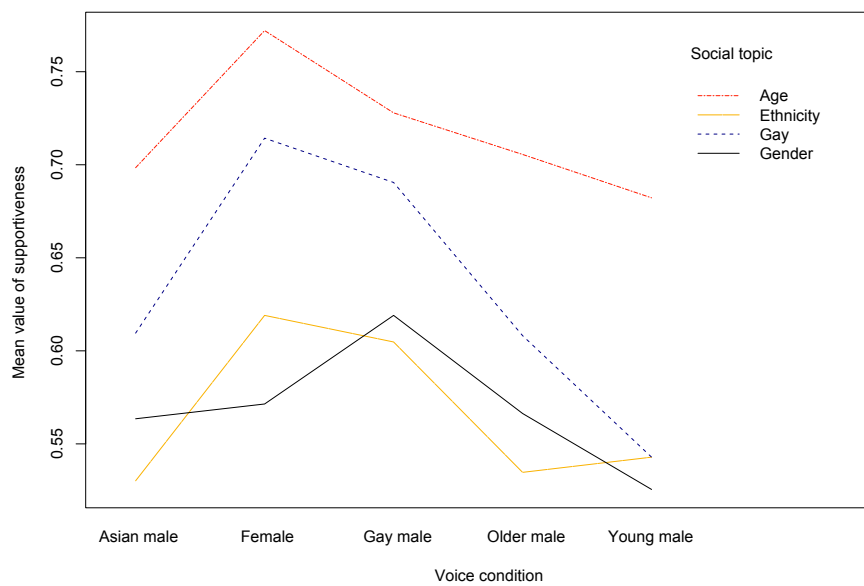


Figure 5.3: Interaction between voice condition and social theme using mean values

## 5.4 Experiment 1B

In Experiment 1A, the predictions of the model showed that the female voice was associated with more predicted supportive answers than male voices, and significantly so when those males were heterosexual. Only one female voice was used in Experiment 1A because voice and social target matching was expected. That is to say, that a voice belonging to a social group would prime increased or decreased support for issues experienced by that group, depending on how listeners construed the voice. With only one female voice, however, it could not be said whether this was due to the speaker being female, or another idiosyncratic feature of her voice. To try to answer this question, a follow-up experiment was run, which was designed to mirror Experiment 1A except that the voices were three New Zealand males and three New Zealand females. Using this design, a clear comparison can be drawn between the effect of male and female voices, allowing us to confirm or refute the voice sex effect seen in Experiment 1A.

### 5.4.1 1B Hypothesis

The hypothesis for this experiment was that participants in female voice conditions would be more socially supportive - that is to say, answer more positively to the different social topics - than participants listening to the male voice conditions. This was directly influenced by the results of Experiment 1A in which the female voice was associated with more supportive answers than the male voice conditions.

### 5.4.2 Participants and Procedure

189 participants initially submitted a response, and as with Experiment 1A, exclusions were made for unfinished, duplicate, or overly quick responses that indicated a lack of engagement with the task. After these 16 exclusions, 173 participants were included for analysis. They were predominantly students drawn from the University of Canterbury in response to emails and notifications posted by course administrators from a wide pool of subjects. 95 participants (55%) were female and the median age of the participants was 24 years. 88.6% of participants identified as heterosexual, and 76% as New Zealand European. For this experiment, three young heterosexual males and three heterosexual females were recorded. Their voices were recorded in the same manner as the previous experiment. Two of these females were young students, and for consistency, the third was the female from

Experiment 1A, who was thirty eight at the time. She was re-used to look at whether her voice would pattern similarly to the other females, or show significant differences in levels of support. All speakers spoke standard New Zealand English as judged by a New Zealand phonetician, with no voice degradation or speech impediments. Their demographic information is shown below in Table 5.4. As with the voices from the first experiment, emails were sent to undergraduates asking them to rate the voices. They were asked:

1. Please select whether the speaker is male or female
2. Please guess the age of the speaker
3. From the following options (heterosexual, gay, other), please select an option you think describes the speaker
4. Please guess the ethnicity of the speaker (e.g. New Zealander, American, Korean, etc)

Table 5.4 provides ratings data for the voices. The information inside the brackets indicates how participants rated the voices on whatever quality precedes the brackets.

Table 5.4: Exp 1B: Speaker information

Speaker	Demographic and Ratings Data
Female 1	heterosexual (90%) NZ (85%) female (100%) student in late 30s (average age 36)
Female 2	heterosexual (80%) NZ (90%) female (100%) student in 20s (average age 26)
Female 3	heterosexual (85%) NZ (100%) female (100%) student in 20s (average age 28)
Male 1	heterosexual (85%) NZ (100%) male (95%) student in 20s (average age 23)
Male 2	heterosexual (90%) NZ (100%) male (100%) student in 20s (average age 27)
Male 3	heterosexual (95%) NZ (95%) male (100%) student in early 30s (average age 27)

### 5.4.3 Qualitative Comments

After completion of the online surveys, participants were asked to give their thoughts on the voice that they had just heard. This is distinct from the pre-survey ratings task (see Table 5.4), which was conducted before the survey went live. At the end of the survey, participants were asked ‘Please comment on the voice that presented this survey, e.g. age, sex, sexuality or whatever else you wish to comment on’. Because the response box was text format, participants were able to enter ‘twentyish’ or ‘probably gay, unsure’ in the box. These comments were not designed to be used in any subsequent analysis, but to give me an indication of any noticeable features of the voices that these participants evaluated. Had any of the voices received numerous comments such as ‘unintelligible’ or ‘horrible voice’, they would not have been used and other voices recorded in their place. Excepting some humorous comments, such as ‘I wanted to throw my computer out the window listening to this voice’, and ‘She sounded dead’, most of the comments were appropriate for the voice. That is to say that nobody commented that the younger male sounded like an older female, or that the Asian male sounded like a New Zealand male. In hindsight, having a fixed answer format would have been more useful for further analyses, particularly as it became clear that affect may have played more of a role for participants’ responses than category activation. This is expanded upon in the discussion chapter. There are simply too many features of a voice, and affective reactions arising from these features, than could be tested for experimentally. That would be a separate thesis entirely, and as such these comments should be treated as preliminary indicators of what features listeners

most report on.

#### 5.4.4 Results

Table 5.5 below shows the mean supportiveness values for each of the social themes, broken down this time by voice sex rather than individual voices.

Table 5.5: Exp 1B: Mean support values

Social Support		
Social Theme	Voice Condition	Mean Support Values
Age support	Male voices	0.72
	Female voices	0.75
Ethnicity support	Male voices	0.65
	Female voices	0.67
Gay support	Male voices	0.68
	Female voices	0.66
Female support	Male voices	0.55
	Female voices	0.60

The statistical analyses were conducted using mixed effects logistic regression models (see Section 4.8). The independent variables were voice sex and participant sex, and the dependent variable was a binary response, 0 indicating not supportive, and 1 indicating supportive. Participants, individual questions, and voice condition were entered as random effects. Social topic was tested for but was not significant as a sole predictor or when interacted with the other fixed effects. The model's predictions are shown in Table 5.6. There was no statistically significant difference between the answers given in the male and female voice conditions. This is shown in the Table 5.6 for clarity. There was no interaction between voice sex and participant sex. Female participants were significantly more supportive than male participants. A plot of the mean support values showing the interaction between participant sex and social theme is shown in Figure 5.4, to give an indication of how male and female participants differed across social themes. This is not based on a statistical model, but rather a plot of the mean values.

Table 5.6: Mixed model output - modelling predicted social support by voice sex

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	0.67	0.22	3.04	.002	**
Female voices	0.14	0.14	1.00	.316	
Female participants	0.33	0.15	2.12	.033	*



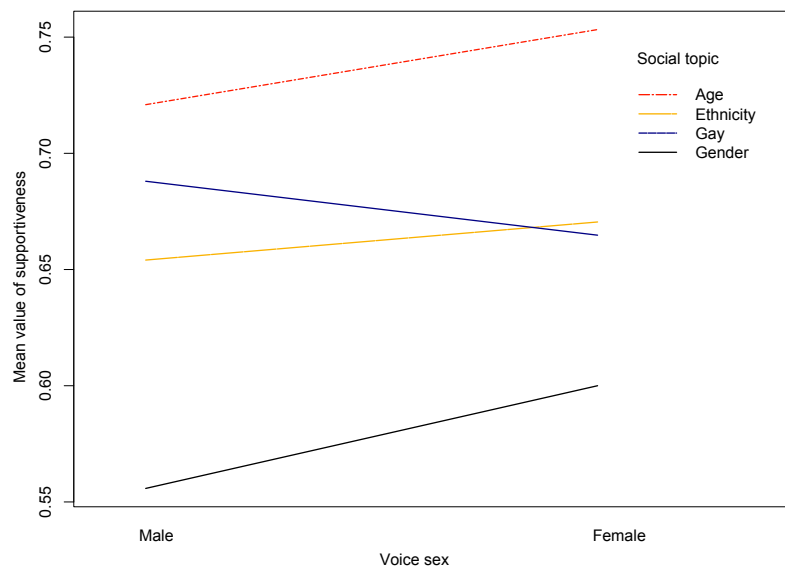


Figure 5.4: Interaction between participant sex and social themes

The interpretation of the results from Experiment 1A was therefore not supported by the follow-up. When the responses were analysed as supportive or unsupportive, the sex of the voice made no difference in this instantiation of the experiment. A rather clear candidate reason for this is the ambiguity of some of the statements, as was alluded to in subsection 5.3.2. This is another design limitation which is expanded upon further in the thesis discussion. Further, having a binary yes/no answer format could have obscured more nuanced answering. These results led me to go back to the data to investigate if there was a better way of representing the data, which is the focus of the next section.

## 5.5 Re-analysis

Having demonstrated that participants from the three male and three female voice conditions showed no statistically different levels of predicted support, I now focus on whether simple ‘agreement’ with the statements can shed light on the variation between voice conditions. As discussed previously, agreement and supportiveness are not the same. Comparing simple agreement to the coding that was used for supportiveness gives us a 55% similarity score (that is to say that out of the 22 statements, twelve would have been coded as ‘unsupportive’ to the particular social group if participants had answered affirmatively in 1A). It would be completely possible, too, for different researchers to code the answers quite differently, and still believe that they indicated a form of support. For example, one of the statements was ‘Women still face the glass ceiling in organisations’. A ‘yes’ answer was coded to be supportive; women do still face the glass ceiling as a cursory glance at multinational boardroom statistics indicates. However someone could answer ‘yes’ to this question and do so not because they think equal opportunities in the workplace has arrived, but because they interpreted the question to mean women should not face a ceiling, or perhaps even that they want to believe that women do not, even if statistics indicate otherwise. Another example is the question ‘It is reasonable to expect people to work until they are 70’. A ‘yes’ answer was coded as unsupportive towards the elderly here. It has been a topic of interest in the media, largely in a negative light. Someone could answer ‘yes’ to this, however, out of a firm belief that nobody should force retirement on another, and if the elderly are capable and willing, they should work until whatever age they choose. I would fully agree that this is also a supportive answer. It is important to remember that this experiment, and indeed this thesis, is not looking explicitly at social attitudes and discrimination. These questions could be coded any number of ways. However, as a result of the ambiguity of what constitutes support, the re-analysis focusses on agreement versus disagreement, which is not subject to the problems of ambiguity.

### 5.5.1 Experiment 1A Re-analysed

In this reanalysis, the same five voices from Experiment 1A were used. Individual voice condition was again entered as the independent variable, and response as the dependent variable, using a logistic mixed effects regression model. Individual statements and participants were entered as random effects. Interactions between

sex of participant and sex of voice were tested for, but yielded no significant results. The female voice was set as the comparison level here, in line with Experiment 1B where voice sex is the experimental variable. As the female voice was the only female from Experiment 1A, setting her voice as the comparison level here seemed logical. Social theme was tested for both as a fixed effect and an interaction and was not a significant predictor. Participant age, sex, and ethnicity were also tested for as both main effects and as interactions, and were not significant predictors.

### 5.5.2 1A Re-analysis Hypothesis

The hypothesis for this re-analysis was that participants in the female voices conditions would have higher predicted probabilities of agreement to the statements than those in the male voice conditions. This was based both on the results of the first instantiation of Experiment 1A, and also research into sex differences and persuasion (see Section 5.8).

Table 5.7: Mixed model output - modelling predicted agreement by voice condition

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	0.51	0.22	2.28	.022	*
Younger male	-0.35	0.10	-3.31	<.001	***
Asian male	-0.06	0.10	-0.62	.535	
Gay male	-0.05	0.11	-0.45	.652	
Older male	-0.16	0.10	-1.56	.117	

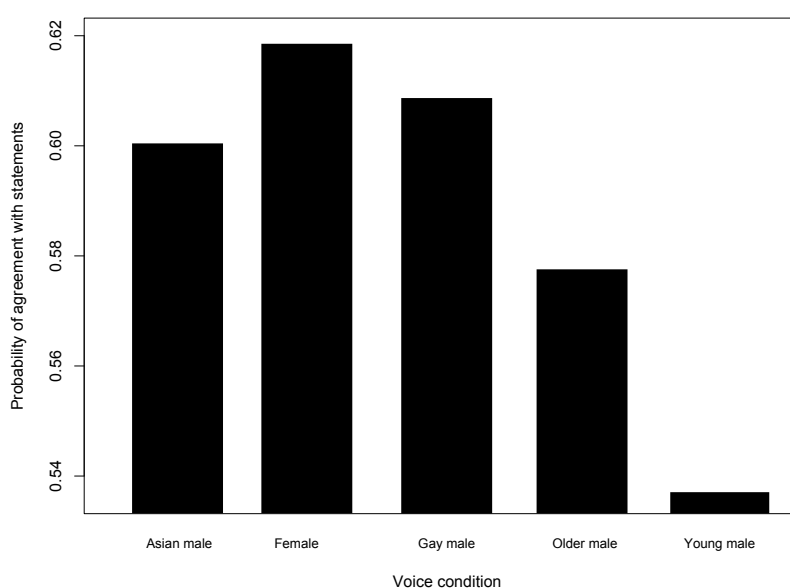


Figure 5.5: Predicted probability of statement agreement by voice condition

Table 5.7 shows that compared to the female voice, participants in the young heterosexual male voice condition had a significantly lower predicted probability of agreement with the statements. The other male voices also had lower probabilities of agreement, but none significantly so. The rationale for testing a sex-of-voice effect remained. It is worth noting too that whilst the younger heterosexual male is significantly different to the gay male condition, he was only rated as 5% more heterosexual sounding than the gay male, and indeed the older male was rated as less heterosexual than the Asian male. If some dimension of masculinity or heterosexuality is at play, this does not seem to be explicit. With Bayes theorem in mind, however, we might expect ratings of whether or not someone is gay or heterosexual or Asian to be swayed by the prior probabilities based on population proportions, and so I'm reluctant to use the ratings data to help support any particular interpretation.

### 5.5.3 Experiment 1B Re-analysed

The same six voices from Experiment 1B were used in this reanalysis. All of the voices were student aged, with the exception of Female 1 who was reused to look at whether there was something unique to her voice, or whether the results for her voice would pattern similarly to other females (see Table 5.7). Voice sex was used as the independent variable here, although individual voice conditions are presented graphically to show how they pattern in comparison to other voices of the same and opposite sex. The dependent variable was agreement or disagreement with the social statements. Individual participants and statements were again entered as random effects. Interactions between sex of participant and sex of voice were tested for, but yielded no significant results. The female voice was set as the comparison level here.

Table 5.8: Mixed model output - modelling predicted agreement by voice sex

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	0.61	0.23	2.60	.009	**
Voice sex = male	-0.26	0.08	3.34	<.001	***

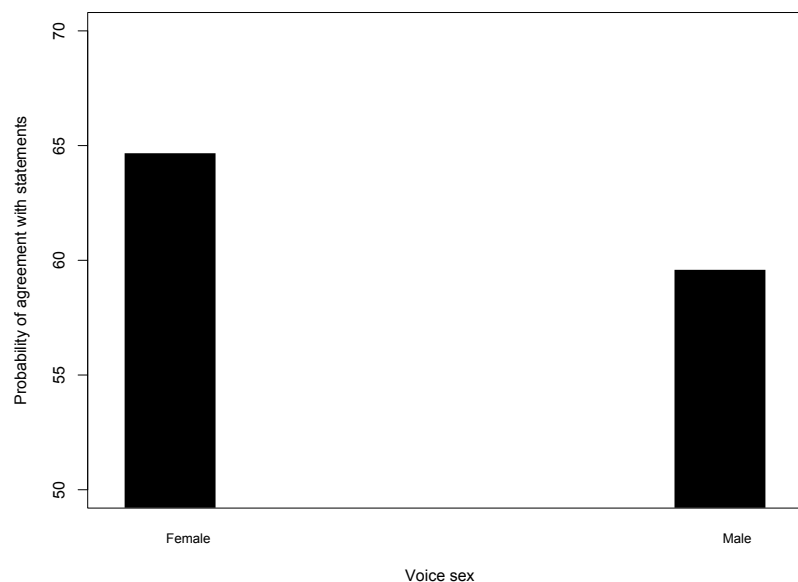


Figure 5.6: Predicted probability of statement agreement by voice sex

Table 5.8 shows that when participants heard statements in a male voice, they were significantly more likely to disagree than when they heard them in a female voice ( $p < .001$ ). Figure 5.6 represents this graphically by voice sex, and Figure 5.7 plots the individual voice condition results. As can be seen in Figure 5.7,

participants in the female voice conditions were predicted to agree more to the statements than participants in the male voice conditions. I also checked whether Female 1 (as used in Experiment 1A) patterned similarly to the other two female voices, running the same model using only the three female voice conditions. There was no statistical difference between the responses in these female voice conditions ( $p=.472$ ).

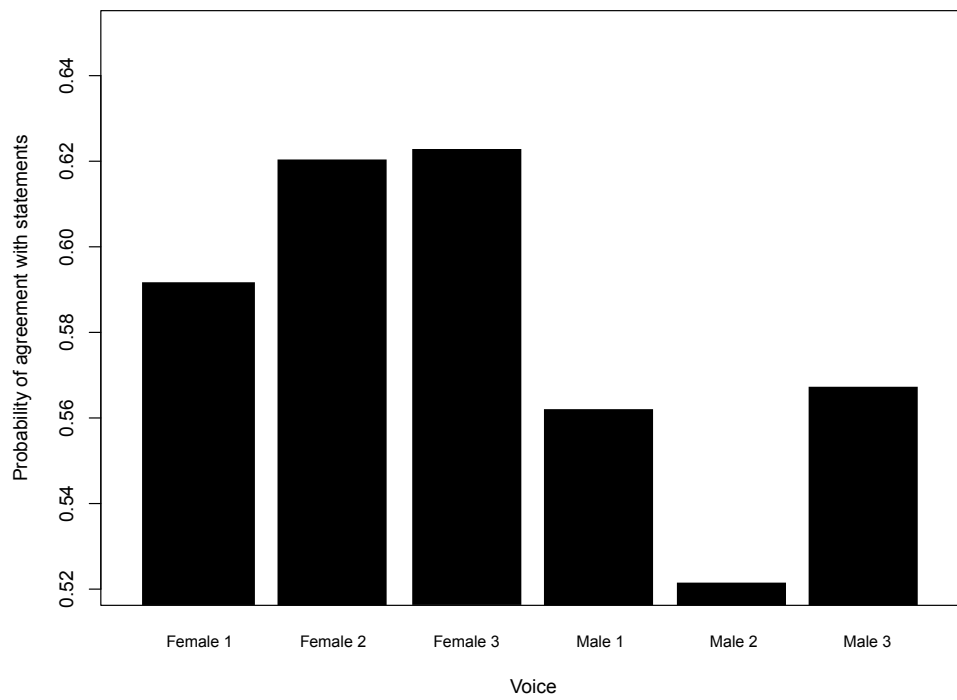


Figure 5.7: Predicted probability of statement agreement by individual voices

### 5.5.4 Experiments 1A and 1B Combined

The results from both experiments were combined to check whether or not the voice sex effect would remain with the addition of the male voices from Experiment 1A. Voice sex was entered as the independent variable, and a binary agree/disagree response as the dependent variable (where 0 indicates disagree, and 1 indicates agree here). Participants and individual questions were entered as random effects. Social theme and participant sex were tested for both as fixed effects and interactions, and were not significant predictors for responses. Table 5.9 below shows that when a male voice presented the statements, participants' predicted probability of agreement was lower. This overall effect of voice sex is plotted in Figure 5.8. This strengthens the two separate analyses, showing that two groups of participants, drawn at different times and exposed to different voices, patterned similarly to each other. When individual voices are entered as predictors in the model, we can see wider gaps between, for example, Female 2 and the Younger male. The predictions from the model for all of the voices is shown in Figure 5.9.

Table 5.9: Mixed model output- modelling predicted agreement by voice sex

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	0.57	0.22	2.61	.009	**
Voice sex = male	-0.22	0.05	-3.98	<.001	***

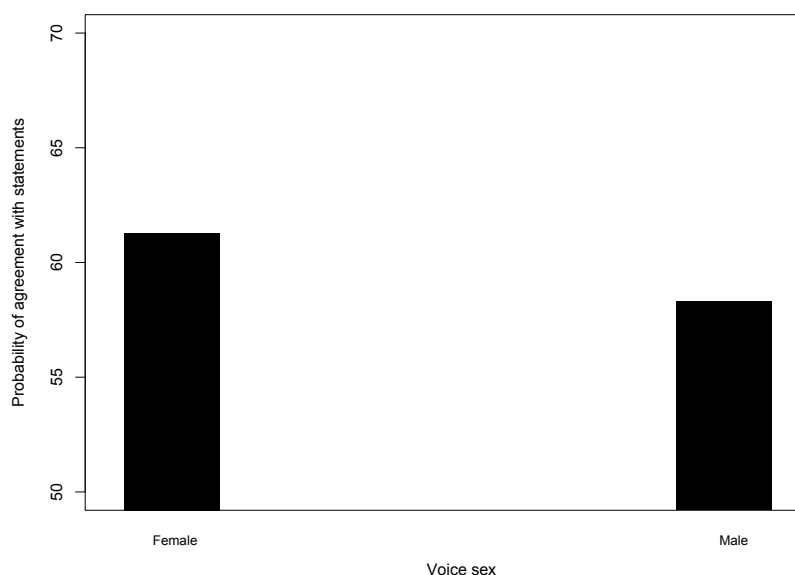


Figure 5.8: Predicted probability of statement agreement by voice sex

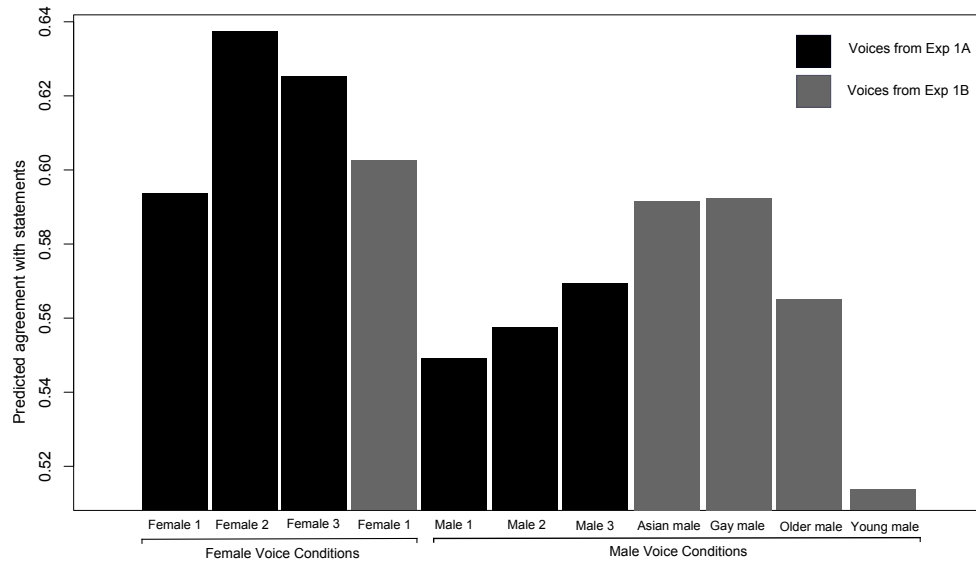


Figure 5.9: Predicted probability of statement agreement by individual voices



## 5.6 Experiment 1C: Silent Condition

Up to this point, I have argued that participants in male and female voice conditions attracted significantly different predicted rates of agreement and disagreement to the audio statements. In the mixed effects regression analyses presented, either an individual voice or voice sex was entered as the independent variable, and one of their levels set as the comparison level - that is to say, the predicted rates of agreement/disagreement were in relation to either whichever voice was the comparison level, or whichever voice sex was set as the comparison level. This presents a difficulty in claiming which voice, or sex of voice, primes participants away from what might be considered a 'baseline'. By baseline, I do not mean their *real* opinion (see Section 4.7). However, if they were presented with the statements in text format, as is typical for a survey, then the responses to this could be used to indicate what particular voice or voice sex acts most differently to a text only version, and this is the focus of the following section. Because of the issues with the coding in the supportiveness versus non-supportiveness analysis, and the lack of any clear effect of voice sex or voice and social category matching, I look at the silent condition with respect to agreement and disagreement.

### 5.6.1 Procedure

The materials and procedure for this experiment and analysis differ from the previous experiments only in that the statements are presented in text-only format. There is no audio in this experiment. Other features, such as the statements and the method of analysis remain the same. 87 participants originally submitted responses, and as with the previous analyses conducted, cases were removed where they were incomplete, had duplicate submissions, or had completed the task suspiciously quickly. After the 12 exclusions, 75 Participants remained for analysis. They were predominately (84 %) undergraduate students from the University of Canterbury, recruited through course coordinators. Their average age was 22 years, and 73% were female. 95% of them rated themselves as heterosexual and 89% as New Zealanders. A screen-shot of the text-only presentation is shown below in Figure 5.10.

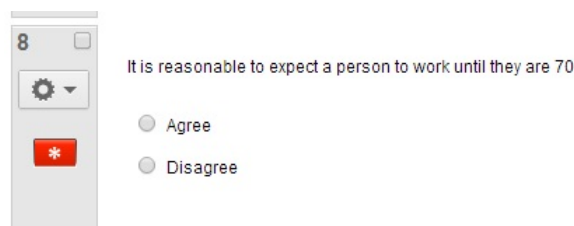


Figure 5.10: Predicted probability of statement agreement by mode of presentation

## 5.6.2 Results

The silent condition was combined with the analyses from 1A and 1B. The model below has mode of presentation (male voices, female voices, and no voice) as the independent variable. Participants and statements were entered as random effects. Social theme and participant sex were not significant as fixed effects or as interactions. *No voice* is set as the comparison level. The dependent variable is the predicted probability of agreement with the statements.

Table 5.10: Mixed model output - modelling predicted agreement by mode of presentation

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	0.48	0.23	2.03	.041	*
Voice sex = Female	-0.02	0.07	-0.26	0.79	
Voice sex = Male	-0.24	0.06	-3.56	<.001	***
Female participants	0.104	0.04	2.16	.030	*

Table 5.10 shows the predictions of the model, where we can see that the male voice condition was predicted to return significantly less agreement than the silent condition. There was no difference between the female voice and silent condition. These predictions are shown graphically in Figure 5.11. When the participants from the three experiments were combined - a total of 494 participants- , the sex of the participant was also significant, with male participants predicted to agree more than female participants.

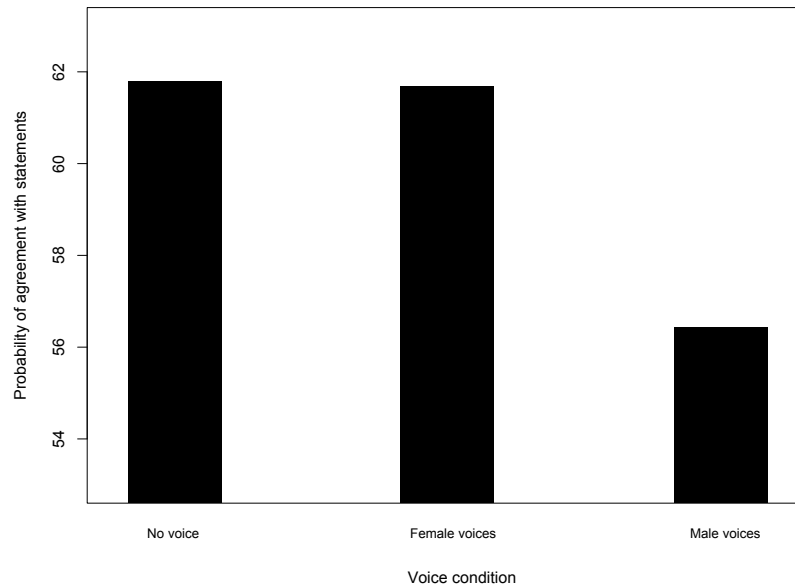


Figure 5.11: Predicted probability of statement agreement by mode of presentation

Finally, I checked again to see how participant sex affected the supportiveness towards the social topics. Whilst there was no overall effect, there was a highly significant effect of female participants giving less supportive answers to female relate statements ( $p < .001$ ). This is shown in 5.12 below.

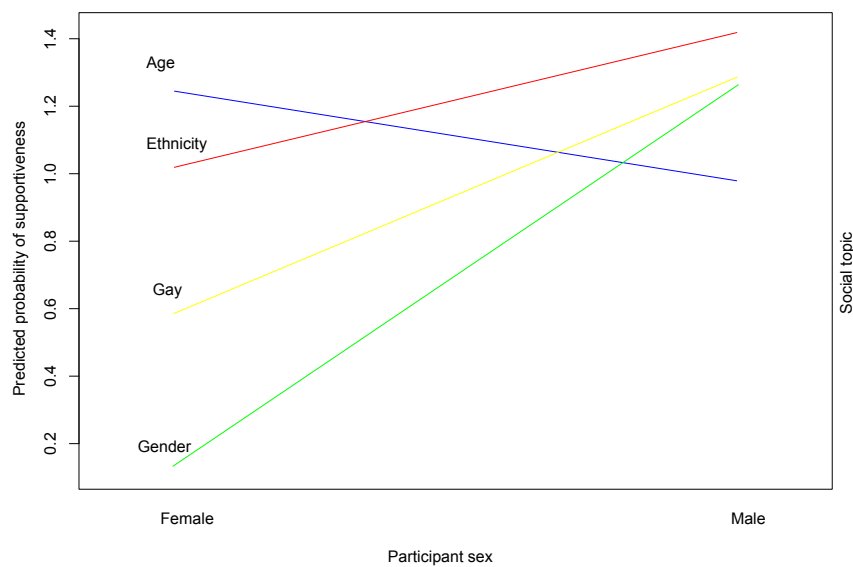


Figure 5.12: Predicted probability of supportive answers by participant sex

## 5.7 Experiment 1 Discussion

### 5.7.1 Summary

In Experiment 1A, the female voice was associated with both more socially supportive answers, and also with more agreement to the statements. In Experiment 1B, there was no difference in social support between the three male and three female voices but the difference between predicted agreement remained. One reason that may account for this is stimuli awareness. This was not a timed task, and participants were faced with often provocative statements that most people would take some time to think about. We know that people typically construe a situation by accessible constructs. As a result of this, the gay male voice could make people view the whole survey as more gay-themed than, for example, elderly-themed, or an Asian voice more focussed on ethnicity and so forth. If participants were aware that the voice they were listening to was having an effect on their answers to questions, then research such as Lombardi et al. (1987) and Strack and Schwarz (1993) has shown that participants may resist using this information to form a valenced judgement.

Whilst there was no effect upon the relative supportiveness of the answers in the replication, there was however an effect of voice sex on levels of agreement. As in Experiment 1A, participants in the female voice conditions in Experiment 1B had higher probabilities of agreeing with the statements. Finally, when a silent condition of the experiment was conducted and analysed in conjunction with 1A and 1B, it indicated that it was the male voices that primed more disagreement when compared to a silent, text-only condition, rather than female voices priming more agreement. The participants in the male voice conditions in the experiment had higher levels of predicted disagreement than those in the female voice conditions.

The agreement or disagreement of the participants to socially relevant statements in these experiments appears to have been affected more by the sex of the voice that presented the statements than any between-voice variations. It is therefore with sex and gender that the discussion of plausible explanations for this begins, with the proviso that these are more to help with future direction than to interpret the current results.

## 5.8 Evaluating Men and Women

In Experiment 1A the sex of voice was a secondary consideration, and in Experiment 1B the exclusive focus. This section presents research that helps with some possible interpretations of these results. If differences arise between male and female voice conditions, then listeners may be accessing different schemas or stereotypes for these voices, and the histories they associate with such voices. The research discussed in the following few paragraphs seeks to highlight these possible differences in order to understand why and how they might inform participants' behaviour.

In Figure 5.4 (indicating the social supportiveness of participants in Experiment 1A) we can see that female participants were the least supportive to female topics. This is consistent with the work of Brunel and Nelson (2003), who showed that in an advertising context, females were more positive towards other-oriented charities (for example disabled charities) than to those with an explicit female focus. When we look at the silent condition in Experiment 1C, we again see that female participants had the least predicted support for female related issues.

Hilary Lips in, in her 1988 text book 'Sex and Gender: an introduction' wrote that not only were the sexes viewed differently, but that males were seen as superior to women (Lips, 1988). Broverman, Vogel, Clarkson, and Rosenkrantz (1972) developed a sex-role survey in response to what they saw as a 'prescriptive' approach to the study of sex-role variation, such as the California Psychological Inventory (Gough, 1956) that based masculinity-femininity scales on sex appropriate behaviours of the first half of the 19th century. Broverman et al. (1972) were, at the time, well placed to update the existing views as they were in the midst of what is commonly referred to as the *second wave* of the feminist movement, particularly in the United States. They sought to study the traits that people assigned to males and females, and the strength of those traits. Their participants were asked to ascribe traits to a typical adult male, a typical adult female, and to themselves. The authors presented multiple findings from their study, and a few of the more relevant ones are selected here. Their results showed

that there was a clear separation of male and female sex role stereotypes, as shown both in previous literature and by common ‘wisdom’. That is to say, people did not view males and females as being particularly similar, contrary to some of the popular claims at the time. Their results also indicated that relative to males, women were viewed less favourably, writing:

Women are perceived as relatively less competent, less independent, less objective, and less logical than men; men are perceived as lacking interpersonal sensitivity, warmth, and expressiveness in comparison to women. Moreover, stereotypically masculine traits are more often perceived to be desirable than are stereotypically feminine characteristics (Broverman et al., 1972, p 75).

The relevance here concerns the seemingly *objective* versus *warmth* components of male and female stereotypes respectively. If men are seen as objective and cold, then this may, at some level, give statements read in a male voice(s) a more objective feeling, and those in a woman’s voice a more compassionate feeling. It could also, possibly in addition to or as a separate process, happen that it causes some mood change in the listener, and it is this mood change that results in task differences.

Eagly, Mladinic, and Otto (1991), taking the research of Broverman et al. (1972) and Rosenkrantz, Vogel, Bee, and Broverman (1968) as their starting point, sought to establish the veracity of the claim that men were evaluated more positively than women. The authors had, in Eagly and Mladinic (1989), found preliminary data suggesting that both in terms of attitudes and beliefs, it was in fact women who were evaluated more positively than men. In their 1991 study they sought to test this directly, with a between-subjects design (rather than the within-subjects design they had used previously). The target groups they wished to seek evaluative data for were: women, men, the Democratic Party and the Republican Party. Also included in their survey were questions about clowns, rapists, Europeans and teenagers, and other ‘filler’ categories. Eagly et al. (1991) found stronger results than in their 1989 study, again confirming that evaluations of women were more positive than those for men. Further, this was true for both male and female participants, or to put it another way, there was no observable in-group sex bias in their participants. This of course says nothing of the effects

of the stereotype, or the real life discrimination faced by women and other groups for whom stereotypes may be positive. The authors were careful to note that their findings were extracted from college students, who are younger and more well educated than the U.S. population as a whole. They were not, however, particularly liberal (as mentioned, the authors had included the Democrats and the Republicans as groups they sought evaluative ratings for). The researchers concluded by offering some reasons for why their findings contradicted those of earlier studies. These include improved methods, progress in society - there was a generational gap between their work and that of Broverman et al. (1972) and Rosenkrantz et al. (1968) -, and by looking at the macro *women* category, rather than to subtypes such as housewife. Deaux (1985), in this vein, suggested that there were more stereotypes for women than for men. How people respond to gendered information (or indeed process it as being gendered) is also variable, which is the topic for the following section on gender schema.

## 5.9 Gender Schema

There appears to be no other dichotomy in human experience with as many entities assimilated to it as the distinction between male and female (S. L. Bem, 1981, p 354).

I suggest that when hearing statements in a female or male voice, a quality that is typically assigned to the sex might also be assigned to the voice, in turn influencing how the speech is processed. I expand here on how that might occur. From a young age, children learn, and are taught, that there are differences between males and females (Fagot & Leinbach, 1989; Bussey & Bandura, 1999). Through processes of direct instruction, socialization, and cognitive development, children acquire a behavioural repertoire that they, and society, believe to be appropriate for their sex. These behaviours become their gender. Further than this, however, Bem in her influential 1981 paper proposed that this translates to how children, and latterly adults, process information. Bem argued that the child begins to use a *gender schema*, and that incoming information related to males and females is systematically processed by this gender schema. Schemas were discussed in section 3.2.1, but as a reminder, a schema serves as a guide to

information processing. Using information built up over the lifespan, it directs and filters information to make it more readily accessible, and relevant, to the task at hand. If the task at hand is to answer a question about one's attitudes and beliefs, then past experience in tailoring these attitudes to both sexes should influence how one reports them consequently.

To the extent that people observe men and women engaging in a division of labor, they regard them as psychologically different (Eagly, 2009, p 650).

Eagly (2009) has argued that we infer gender differences not from social prescription, but from observations of behaviours that over time we come to know as typical or not. She has noted that the abundance of research on gender differences and the opposing directions of many of the findings, not to mention number of theories, open it to the criticism of randomness and of there being no overarching effect. She has argued, though, that this cannot be true as in our daily lives, it is patently clear that gender remains influential in multiple ways, from the choices people make and how they interact with each other, to the institutions of society that still have huge gender ratio problems (Eagly, 2009). It is this system of influences, from the personal to the structural, that feeds into the gender schema as described above, which in turn should act as an aid to decision making under certain circumstances.

## 5.10 Persuasive appeal

One implication of the sex and gender differences highlighted in the previous two sections is that the persuasive appeal of women may be different to that of men. Carl Hovland was the progenitor of the systematic study of persuasion, and his research program looked specifically at how persuasion was influenced by the source of the message, the content of the message, and the audience for the message (Hovland & Weiss, 1951; Hovland, Janis, & Kelley, 1953; Hovland & Pritzker, 1957; Hovland, 1959). The result of this research indicated that messages are more persuasive when the source is credible and knowledgeable on the subject matter, but also that they appear not to be biased. Additionally, even weaker messages with less credible sources can, over time, lead to opinion change through what Hovland called the ' sleeper effect '. One avenue for further exploration here,



then, would be to have female and male voices rated on dimensions of credibility and trustworthiness, and then design questions that were designed more clearly to elicit persuasion, rather than simple agreement and disagreement.

Whilst experiments in this chapter are, to my knowledge, the first to look at voice priming of social attitudes, other researchers have in the past looked at gender differences in social persuasion, argumentation and agreement. An early study by Piliavin and Martin (1978) looked at argument and discussion styles of males and females, in groups composed of: all males, all females, and mixed male-female groups. Their paper reports a number of findings, but relevant to these results was that men disagreed more when discussing with other men, and showed more agreement when interacting with women. Carli (1989) extended this line of enquiry, looking at how two people with opposing views on a subject, such as alcohol limits, come to persuade each other. Taking the dyadic male:male, female:female and male:female paradigm, they confirmed Piliavin et al's work and also found that when interacting in mixed sex pairs, both sexes exhibit fewer stereotypic traits than when in same-sex pairs. They found that whilst positive social behaviour was linked to liking, this did not translate to greater persuasion and that men and women did not differ, overall, in influence strategy, contrary to what had been shown in previous research (which they suggest was because previous research had relied on self-reports). Carli (1989) does not offer a definitive explanation as to why influence is greater when females are present, instead positing that it could be because people expect women to be more friendly and social, and are therefore less inclined to be competitive, or because they expect to be competitive with males. Other researchers have however found that men are seen to be more credible and persuasive (Kenton, 1989), and so care should be taken not to over-interpret these results.

### **5.10.1 Voice Affect and Processing**

An important consideration, and one that is directly linked to persuasive appeal, is the affect of the voices. Affect was quantitatively untested in the current experiments, but work by Robinson and McArthur (1982) has shown that participants rated a pre-recorded male voice as more dominant, and paid closer attention to it. Pushing this reasoning to this thesis, it is possible that

closer attention ensured a more analytical approach to statement evaluation and disagreement, such that statements presented in a male voice were challenged more. Moreover, two of the male voices (Asian male and Older male) were more negatively commented on in the free-response feedback, and a third (Male 2) was neutrally judged at best. It is possible, then, that other participants listening to these voices also experienced no fondness for these voices, even if they did not comment directly on it. None of the female voices attracted particularly negative comments. It is unclear if this is because the participants simply paid more attention to the male voices, as in Robinson's (1982) study. A voice that produces some negative affect in listeners might also activate *system two* processing, the more analytical and systematic style of processing (see Stanovich & West, 2000). Chaiken (1980) has conducted work in this area, showing that information from unlikeable sources was challenged more when the topic was of low personal relevance, whereas information from likeable sources was heuristically processed. If this were so, then one would again expect the statements in the present experiments, many of which were of low personal relevance to the participants, to be challenged more, leading to greater disagreement, when voice sex was male.

In a similar vein, the mood that the voice puts participants in should also play a part, as predicted by mood(and feelings)-as-information models (see Schwarz & Clore, 1983; Schwarz, Bless, Wanke, & Winkielman, 2003; Schwarz, 2011). Mood can act as its own heuristic, whereby participants use information about their current mood to assess how they view a target. When in an unpleasant room, for example, they may report diminished life satisfaction, but when asked about their current living arrangements, use the room as a contrast and rate their living satisfactions are rather adequate (Schwarz, Strack, Kommer, & Wagner, 1987). A stimulus may lead to a certain mood, but how that mood is then used is task-specific. It is possible, for example, that the Asian male voice induced a bad mood in participants, owing to the increased attention needed to carefully listen. The listeners, however, then have to use that mood information, and this would vary across individuals, and also across the statements. When they are asked to rate whether the excerpt of Cantonese dialogue is pleasant, then, the bad mood might make them think "Well, it's more pleasant than the speaker that's been asking these statements", or alternatively, "No, I don't like any of these foreign sounding voices, please let me out of here". This would be an interesting avenue to

follow up in future experiments.

### 5.10.2 Conclusion

The main point to stress here is that voice differences arising from demographic distinctions of speakers, in this instance the sex of the speaker, can act as a behavioural prime. This was the hypothesis based on the priming literature reviewed in Chapter 2, and it was supported. Voice, like other types of stimuli including text, visual information, non-speech audio information, texture, and so forth, can affect behaviour in subtle ways. In the experiments described above, participants agreed less to statements read in a male voice than those in a female voice, or when they were text only. The current results do not support a simple category activation account- there was no voice and topic matching, and whilst the male voices prompted greater disagreement, this was not true of the female-topic questions. It does appear that something more subtle is at play, and I have offered voice affect, credibility, and persuasiveness as factors for further exploration based on some of the literature in this area.

To fully understand why male voices acted in this way, future experiments would have as part of their hypothesis claims about gender stereotyping, voice credibility and affect. These voices would have to be controlled for other social qualities, such as age, ethnicity, sexuality, and many more things that, due to the infancy of voice and priming work, are unknown to affect listeners. Pitch, breathiness, relative loudness (that is to say the loudness of the speaker, not the recording) and more phonetic features could also play their part. Much work would have to be done establishing the features which can act as primes before making conclusions about the effects, and as Stroebe and Strack (2014) have argued, theories would follow from multiple detections of similar phenomena (both direct and conceptual) rather than precede them.

Further, it needs to be stressed that these effects were not large - the difference in predicted probabilities ranged between .53 and .67. The suggestion is not that the male voice activated the concept of *objectivity* so strongly that it informed everything thereafter, or that the female voices made participants feel compassionate to the extent that they agreed to anything. Rather it is that cues were taken from the voices which subtly interacted with the task of the listener,

such that, were they to use an anchor to base their decision on, voice was an option available to them. Further, the participants used in this thesis are *WEIRD*, in the terminology proposed by Henrich, Heine, and Norenzayan (2010), expanded on in Section 8.3. That is to say, they are western, educated, industrial, rich and democratic. I would not predict these results to pattern similarly if conducted in other countries, particularly those with dramatically different social structures and sex and gender attitudes. In a country where women's opinions and rights to public debate are stifled, and men rule both the political and social agenda, then we might expect the reverse of this result. In a country where gay men are stigmatized and penalized, hearing an identifiably gay voice might provoke strong aversive feelings, such that participants in that country would strongly disagree with many of the statements simply because the voice was gay. There are many different reasons why certain voices in certain cultures would provoke strongly repellent, or pleasing, reactions. This would have to be considered for future researchers looking at similar questions to those presented in this thesis.

## Chapter 6

# Priming, Heuristics and Ambiguous Stimuli

### 6.1 Introduction

In the previous chapter, it was argued that voices have the ability to sway reported opinions in particular ways. The most robust of these findings was that male voices led to greater disagreement with socially targeted statements. The agreement given to statements recorded by the female voice were more similar to those given in a silent, text-only condition. It was the male voices, it appears, that drove the observed effects and although there were also smaller, more specific effects on different social themes, these were rarely significant or robust and were not elaborated on.

In this chapter, I investigate whether voices can act as something like a heuristic. Heuristics, in the sense that I use the term, comes from Tversky and Kahneman's (1974) seminal work on judgements and decision making. Heuristics, they wrote, "reduce the complex tasks of assessing probabilities and predicting values to simpler judgemental operations" (Tversky & Kahneman, 1974, p 1124). These rules of thumb, which come from an "intuitive, rapid, and automatic system", allow people to conserve mental resources in making a decision or evaluating incoming information. The lines between priming and heuristics are necessarily blurred by shared stimuli and outcomes, although they can be thought of as distinct processes too. To briefly clarify, when I speak of priming I mean the

incidental activation of knowledge structures (Bargh et al., 1996), which is the working definition of priming provided at the outset of this thesis. That is to say that the person who is primed is not necessarily aware, and certainly not actively thinking about, the stimuli responsible for a behavioural effect. Heuristics, on the other hand, are typically said to require attention if not awareness of their effects. In the case of anchoring, attention must be paid to an initial value. For the availability heuristic, attention must be paid to exemplars of similar events, and co-occurrences of incidents, in memory (Tversky & Kahneman, 1974). The experiment presented in this chapter is, like all of the experiments for this thesis, a priming experiment. The reviewed literature on priming and automaticity is directly relevant here, however a short discussion of heuristics should prove helpful.

In the literature review, I wrote how Susan Fiske and Shelly Taylor referred to people as *cognitive misers* in their 1984 Social Cognition book. In their 1991 edition, they changed their terminology to *motivated tactician*. These motivated tacticians are people who display more cognitive bias under certain conditions, biases which typically work favourably for the individuals in those conditions. This motivated tactician is not so much miserly, but resource efficient. Part of being a cognitive miser or motivated tactician is to conserve resources, to use mental short cuts by drawing on past and present cues, and to ignore all the available information, and instead focus upon what each individual considers salient. Whilst heuristic usage has historically been considered a detrimental bias, a number of researchers have argued that deviations from the normative decision making model, based on classical decision making theory, may in fact be the best strategy in certain situations (Macrae, Milne, & Bodenhausen, 1994; Stanovich & West, 2000; Gigerenzer & Gaissmaier, 2011). In Macrae et al's study (1994), participants were asked to look at a screen that would present personality traits with a person's name. Concurrently, they would also be required to monitor an audio passage which they would be then tested on later. In one condition, these traits were accompanied with a stereotype for which the traits were congruent, such as *skinhead*, and traits such as *aggression*. In the other condition the stereotype was absent. In their first study the stereotypic information was presented above the level of awareness, and in their second study, below the level of awareness. In their third study, the level of subliminality was increased, and participants were directed to prioritize the impression formation task. Across all three studies,

results supported their predictions and suggest that the presence of stereotypic information aided processing, thus freeing up resources for the tasks at hand. Further, across three studies, they showed this could be both unintentional, and proceed without awareness. The presence of stereotypes, a type of heuristic, certainly appears to be a useful strategy under such conditions.

Gigerenzer and Gaissmaier (2011) review much of this recent line of work, highlighting that when the required information to make a classically rational decision is not available, or when the decision maker does not possess the capabilities to do so, then heuristic usage may in fact be normative. They provide a nice quote from George Stiglitz, the economist, saying of the financial crash of 2008 “It simply wasn’t true that a world with almost perfect information was very similar to one in which there was perfect information” (Stiglitz, 2010, p 243). Quite distinct from the information perspective, there are social and moral situations where heuristics could be seen to be a superior form of decision making. An example would be automatically offering a seat on a bus to an elderly person (a social norm that Gigerenzer and Gaissamer call a social heuristic), rather than systematically calculating their age, fitness levels, your own tiredness levels and whether or not there are other spare seats in order to make your decision.

## 6.2 Anchoring, Accessibility and Availability

If the information we have about a target object is ambiguous, if the time or mental resources we have available are scarce, or if we are distracted while forming our judgement - in short, if we have to make judgements under suboptimal conditions, we are particularly likely to use what is at the top of our minds as the basis for our judgement (Strack & Mussweiler, 2003, p 80).

I focus here on two heuristics that relate to the current experiment - anchoring, and the availability heuristic. A comprehensive treatment of many more can be found in Daniel Kahneman’s ‘Thinking, fast and slow’ (Kahneman, 2011), which collates decades of work on heuristics and decision making. These two heuristics are part of a larger collection of meta-cognitive experiences that people use in their naive theories of memory (that is, what they think about their own acts

of remembering and information processing), of which a good discussion can be found in Schwarz (2004). On anchoring, Tversky and Kahneman (1974) wrote “In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer...adjustments are typically insufficient. That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring” (Tversky & Kahneman, 1974, p 1128). An example of anchoring they provided was asking a group of participants to guess whether or not Mt. Everest was more or less than 50,000 feet. When participants were asked to give a concrete numerical answer, it varied around the value initially offered. In day to day life, examples of anchoring can be seen in house purchasing, criminal sentencing, and product bargaining, where initially offered values affect final values. If a street vendor in Beijing suggests you pay 500  *yuan* for a fan, you are unlikely to offer 1000, and depending on the sensitivities of the vendor, probably will not get it for 50. Epley and Gilovich (2010), in a larger article about linking anchoring with persuasion research, stated that anchoring research has been in waves, and the current ‘third wave’ “considers anchoring in all of its everyday variety and examines its various moderators in these diverse contexts” (Epley & Gilovich, 2010, p 21). This third wave is in many ways similar to the third wave of sociolinguistic research (see Eckert, 2000), which has sought to study language in all of its everyday environments and contexts. It is indeed this area of sociolinguistics that partly inspires this thesis.

The second heuristic relevant here is the availability heuristic. Again, this originated primarily from the work of Tversky and Kahneman (1974). This heuristic is integral to the perceptual fluency and ease of recall research which is referred to in both Section 2.2 and 2.4. The availability heuristic is a bias which causes people to overestimate the likelihood of something based on the ease with which they can recall similar events. Further, when events co-occur, the connections between the associations are said to be strengthened. As Tversky and Kahneman wrote, “That associative bonds are strengthened by repetition is perhaps the oldest law of memory known to man...” (Tversky & Kahneman, 1973, p 208). In the experiments that follow, this associative strength is assumed to exist for the participants such that when they hear the voice, their past experiences of hearing this voice are also associated with the contexts and outcomes of hearing these voices. As an example, if they have compounded in memory the experience



of hearing a gay voice and, for example, liberal politics, this experience should make available not only gay themed concepts but also liberal politics or whatever other compounded concepts they have activated. When they hear an older voice, this may be perhaps compounded with grandparents or films that depict the elderly in a certain way.

Many of the day to day occurrences of the availability heuristic are trivial. There are instances when it is however more serious. The Mumps, Measles and Rubella (MMR) vaccine scandal is one such example, where the publication of a subsequently discredited medical paper that suggested colitis and autism could result from the vaccine resulted in mass panic and lower rates of essential vaccination. The media was criticized for giving undue prominence to the research, and even after it was retracted and further research showed that the MMR vaccine was safe, parents, driven largely by media fear and heuristic decision making, were often reluctant to inoculate their children for mumps, measles and rubella (Serpell & Green, 2006).

### **6.2.1 Non-numerical Anchoring**

Whilst anchoring, and the availability heuristic, have traditionally been studied with numerical values in mind (either by initial values, or by ignoring base-rates), researchers have also used more abstract types of anchors. Whether the specific term anchoring is used varies, but the underlying assumptions are similar. Strack and Mussweiler (2003) argue that the anchoring heuristic is a specific type of accessibility, one of the most powerful constructs in social psychology. They further note that, unlike numeric priming, their model of anchoring-assimilation happens through “increased accessibility of semantic information as a consequence of a presumed hypothesis test” (Strack & Mussweiler, 2003, p 83). Applying this model to my task, we would propose that participants have highly accessible social characteristics related to the voice they are hearing. Having been asked to judge something, they run a hypothesis test. In the case of being asked to judge whether a voice is gay, this hypothesis would simply be ‘Is this speaker gay?’. The active semantic information, which should vary from participant to participant but have enough overlap, is at the “top of their mind” (Strack & Mussweiler, 2003, p 80), which may lead them to end their hypothesis test and answer, more often than

not, that the voice is a gay male.

Chapman and Johnson (1999), in Experiment 1 of their paper, used non-traditional anchoring to investigate people's judgement of an apartment's value. In the same paper, they also mentioned random anchors, suggesting that the mere presence of the anchor could lead participants to consider its presence informative and thus use it in for target evaluation. Mussweiler and Strack (2000) showed that concepts related to the anchor become more accessible through the associative network. In their study, participants who were asked whether the annual mean temperature in Germany was higher or lower than 20 degrees Celsius were quicker to respond to summer related words in a following lexical decision task. Participants who were asked whether the mean temperature was higher or lower than 5 degrees Celsius were quicker to respond to winter related words. This indicates that there is a semantic component to anchoring, not just purely a numeric one as most studies have assumed.

Kahneman (2011) has written that anchoring can occur both from systematic and automatic processing. The systematic process he terms anchoring by adjustment, which involves mental effort and deliberation of an initial value. The more automatic, System 1 process, was initially just a hunch of his. This hunch, in light of much evidence since, appears to be correct and the automatic form of anchoring can be just as easily called priming. In his words, "suggestion and anchoring are both explained by the same automatic operation" (Kahneman, 2011, p 123).

If we have recently heard a voice, that voice is accessible in working memory, particularly if the voice or speaker are idiosyncratic or meaningful to the listener. As such, when making a decision where little or no useful information is available, it seems reasonable to assume that the voice might play a role in the decision. Aside from being recent, if a person is presented with an audio question that they must think about, the 'prime' is not entirely incidental. The voice has been paid attention to, and people likely try to deduce speaker information from the voice. That speaker information, then, becomes a type of anchor. As an example, if you were asked by a Chinese male to guess where music of ambiguous origin was from, would the speaker information you have deduced make it more likely you would

say the music was of Asian origin? This question, and others of a similar nature, are tested directly in this experiment.

## 6.3 Experiment 2: Ambiguous stimuli

### 6.4 Method

The questions which make up this experiment were included in the online questionnaires described in Experiments 1A and 1C, and so the data analysed comes from the same participants - 245 from 1A, and 75 from 1C, making a total of 320 responses for each question. The questions were interspersed throughout the social statements. Again, each listener heard the question in either one of the different voices, or read it in the text only condition. Again, as with Experiment 1A, the question themes were associated with particular social groups related to the different voices: elderly, gay, female and ethnic minority themes. I have chosen to treat it as a separate experiment for both practical and theoretical reasons. First and practically, the questions in this part are not statements to be agreed or disagreed to, but ones where values or answers are required. Participants were not expected to have any existing opinion on the stimuli, and so should be forming their answers *online*, so to speak, and using mental short-cuts. Secondly and more theoretically, the hypotheses were much more loosely formed and distinct from the hypotheses of Experiment 1.

### 6.5 Hypotheses and Questions

A challenge in this work is that is not built upon a strong body of continuing research, offering an opportunity to begin addressing this. There is, to my knowledge, no existing literature showing the effect of particular voices, in this case acting as social category proxies, as anchors or decision making cues. Therefore, the overall hypothesis to this thesis, and for each experiment therein, is that voice cues can have an effect upon individual behaviour. The overall hypothesis is non-directional, as stated in the introduction Section 1.3. That said, based on the category activation literature described previously, and as set out by Förster et al. (2009), one reasonable hypothesis is that voice-and-target assimilation might

occur, such that a gay voice would prime higher ratings of gayness, a female voice higher ratings of femaleness and so forth. Whilst there is ample research showing the many conditions in which assimilation does not happen, it is the most commonly described finding in the literature and can be taken as the primary hypothesis for these questions.

A paraphrased version of the questions for each social theme are given in their respective analyses. These serve to make the analyses clearer to follow. For the list of the questions and their original wording (which do not always correspond to how the values were transformed for analysis) please refer to Appendix B. A screen-shot of how the questions were presented is shown in Figure 6.1, in this case showing the question that asked participants to rate how heterosexual in appearance the male in the photograph was.

24

On a scale of 1-100, how confident are you that the male in the photograph is heterosexual - (0 = not at all, 100 = definitely)

Please indicate with your slider

0 10 20 30 40 50 60 70 80 90 100

Figure 6.1: Example presentation of ambiguous stimuli question

### 6.5.1 Data Analysis

In designing the questions in this section, it was apparent that some had a more strongly directional hypotheses than others. Further, the extent to which some of them were tapping into the same overall question, or were separate, was clearer for some question combinations than others. Because, however, assimilation was a primary hypothesis, it made sense to test whether or not the questions could be collapsed into a smaller number of variables. To standardize how the questions were combined, a principal components analysis was conducted in R, using the

*psych* package (Revelle, 2014). For each of the social themes, I ran the *fa.parallel* command from the *psych* package on the variables from each social theme. For each of the analyses that follow, I specify if they have combined questions, or whether they present single question analyses that do not form natural factors with other questions.

For questions with a binomial distribution, *family=binomial* was specified in the models. From the output of the general linear models, which were used for single question analyses, predictions were generated using the *predict* function in R, and it is these predictions that are plotted in the accompanying figures. For combined questions the *lmer* command was used (see Section 4.8). The magnitude of the effects can be estimated by looking at the coefficients and standard errors. The independent variable was always speaker, and the dependent variable was always response. Participant sex and age were tested as fixed effects, and are reported where found to be predictive. Participant ethnicity was not significant, likely due to very low numbers of non New-Zealanders. For consistency and clarity, the silent condition is entered as the comparison level each time, and post-hoc comparisons between voices are given where useful.

## 6.6 Results

### 6.6.1 Gay Themed Questions

- 1) On a scale of 0-100 (0 = definitely gay, 100 = definitely heterosexual) rate the following photograph
- 2) On a scale of 0-100 (0 = definitely gay, 100 = definitely heterosexual) rate the following audio clip
- 3) On a scale of 0-100% guess the percent of 1000 polled men that listed Lady Gaga among their favourite artists

The principal components analysis indicated that there were no underlying groupings to the answer values. This was further supported by looking at the values of the correlation matrix, which were low at: .08., .06 and .001. None of these questions were thus combined.

- **Question 1)** On a scale of 0-100 (0 = definitely gay, 100 = definitely heterosexual) rate the following photograph

Table 6.1: Linear regression output - predicted confidence of photograph being heterosexual male by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	54.61	2.37	22.95	<.001	***
Asian male voice	9.32	3.50	2.65	.008	**
Female voice	6.81	3.97	1.71	.087	.
Gay male voice	-2.90	4.00	-0.72	.468	
Older male voice	10.93	3.78	2.89	.004	**
Younger male voice	13.01	3.78	3.44	<.001	***

This question was analysed using a linear regression model. The independent variable was speaker, and the dependent variable was a response between 0 and 100. The silent condition was used as the comparison level. The predictions of the linear regression model can be seen in Table 6.1. The younger male, older male, and Asian male voice conditions all returned significantly different, and higher, probabilities of *heterosexual* ratings of the photograph. It was the gay male voice however that returned the only negative coefficient. The model was run again with the gay male voice as the comparison level, and it was significantly different to all of the other voice conditions, but not to the silent condition ( $p=.468$ ) That is to

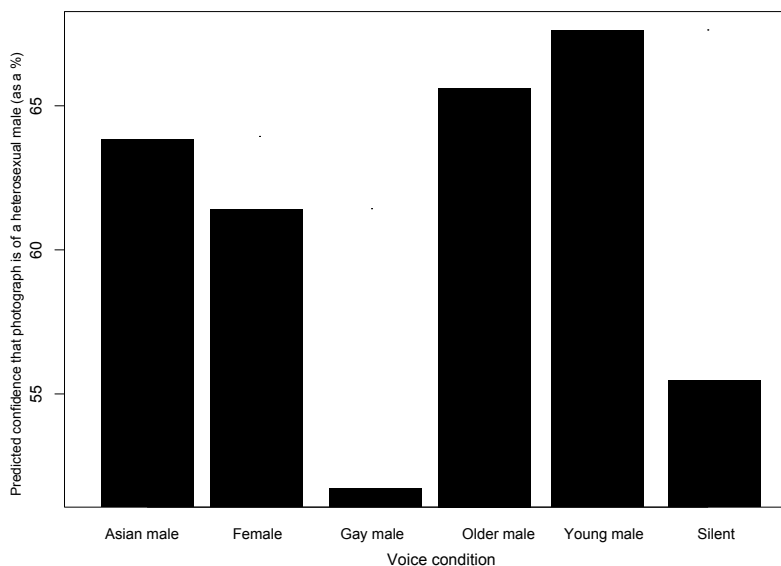


Figure 6.2: Linear regression plot: predicted confidence that the photograph is of a heterosexual male

say that hearing non-gay voices appeared to prime participants to rate the male in the photograph as more heterosexual looking. The model's predictions are shown graphically in Figure 6.2, where it is clear that the silent condition acts quite differently to the non-gay voice conditions. One interpretation of these results is that hearing a heterosexual voice primes participants to rate a photograph as more heterosexual, possibly as a result of participants' construing relevance from the voice, but that the effect is negated by hearing a gay male voice ask the question.

- **Question 2)** On a scale of 0-100 (0 = definitely gay, 100 = definitely heterosexual) rate the following audio clip

Table 6.2: Linear regression output - predicted confidence of audio being a heterosexual male by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	40.24	2.70	14.87	<.001	***
Asian male voice	-0.31	3.98	-0.08	.936	
Female voice	2.64	4.51	0.58	.559	
Gay male voice	-5.82	4.55	-1.28	.201	
Older male voice	-3.13	4.30	-0.72	.466	
Younger male voice	11.49	4.30	2.67	.007	**

This question was analysed using a linear regression model. The independent variable was speaker, and the dependent variable was a response between 0 and 100. The silent condition was entered as the comparison level. The predictions of the

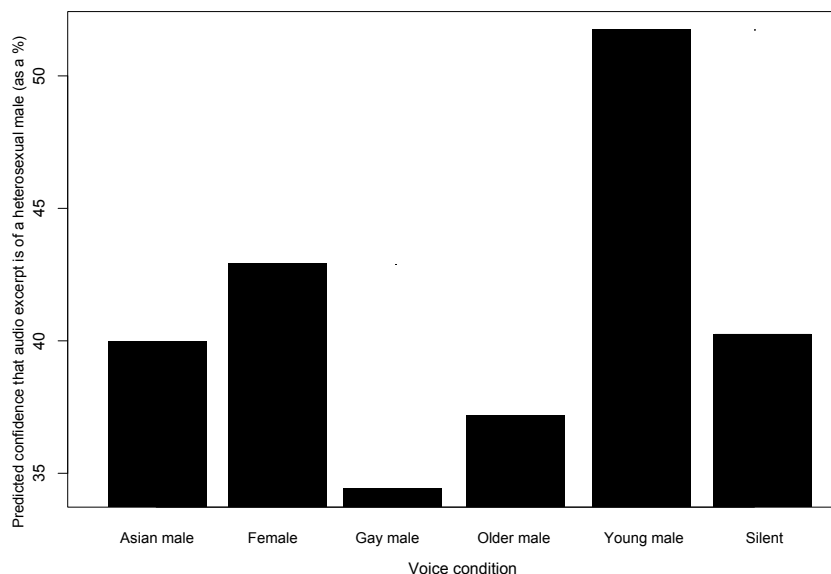


Figure 6.3: Linear regression plot: predicted confidence that the audio is of a heterosexual male

linear regression model can be seen in Table 6.2. Compared to the silent condition, the younger male voice condition returned a significantly higher probability of the audio being judged as heterosexual. Looking at the coefficients, we can see that the gay male condition had the lowest coefficient, and thus lowest predicted probability of participants judging the audio clip as heterosexual sounding. The model was run again using the gay male condition as the comparison level, and it differed significantly from the younger male voice condition ( $p < 0.01$ ). The model's predictions are shown graphically in Figure 6.3. Participants then were generally predicted to rate the audio clip as slightly gay sounding, but when hearing the young, heterosexual male voice were predicted to rate the excerpt as more heterosexual.

- Question 3) In a poll of 1000 men, estimate the percentage that listed Lady Gaga among their favourite artists)

Table 6.3: Linear regression output - predicted estimate of percentage of Gaga fans per 1000 men

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	13.86	1.85	7.48	<.001	***
Asian male voice	4.41	2.73	1.61	.107	
Female voice	15.22	3.09	4.92	<.001	***
Gay male voice	9.42	3.11	3.02	.002	**
Older male voice	6.21	2.94	2.10	.035	*
Younger male voice	2.11	2.94	0.71	.474	



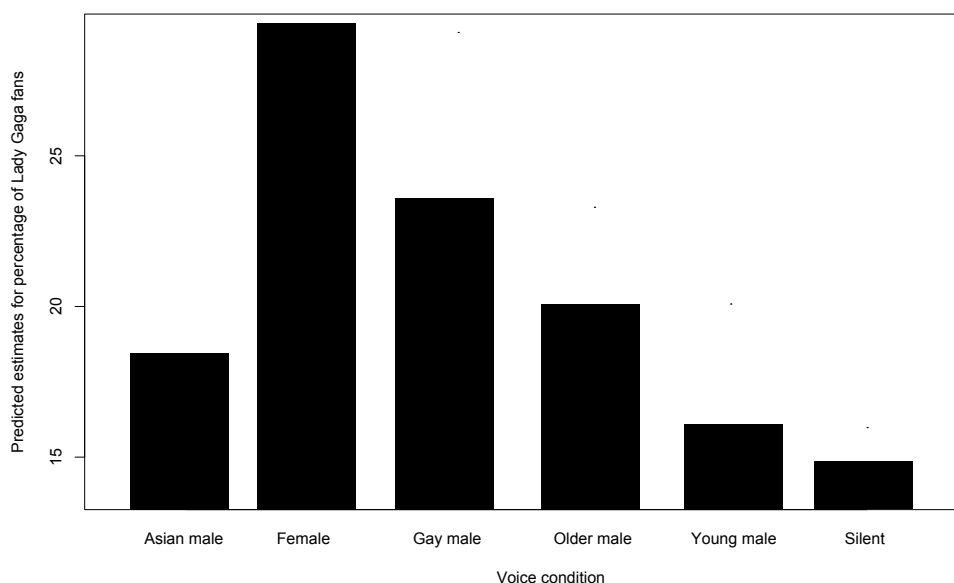


Figure 6.4: Linear regression plot: predicted estimates of male Lady Gaga fans

This question was analysed using a linear regression model. The independent variable was speaker, and the dependent variable was a response between 0 and 100. The silent condition was entered as the comparison level. The predictions of the linear regression model can be seen in Table 6.3. Compared to the silent condition, the female voice, gay male voice and older male voice returned higher predicted estimates of Lady Gaga fans. That is to say that participants listening to those voices were predicted to estimate *more* Lady Gaga fans in a poll of 1000 men. Looking at the coefficients, we can see that the female voice had the highest predicted estimates, and the silent condition the lowest estimate of non Gaga fans. The model's predictions are shown graphically in Figure 6.4. This question serves as reminder that these hypotheses were very loose. Finding that a female voice was associated with more predicted Lady Gaga answers could have just as easily been the hypothesis and entered alongside the gender themed questions. For this reason I am reluctant to over-interpret these predictions.

### 6.6.2 Gender Themed Questions

- 1) On a scale of 0-100 (0 = definitely female, 100 = definitely male) rate the following photograph
- 2) On a scale of 0-100 (0 = definitely female, 100 = definitely male) rate the following audio clip

- 3) On a scale of 0-100% guess the percent of crowd in the photograph that is male

A principal components analysis (PCA) for these three questions returned eigenvalues equal to or greater than 1 for questions 1 and 2. The correlation between the questions was low, at .11, but as they follow the same hypothesis direction, and use the same scale and values, they were combined based on these PCA results. Question 3 is presented individually and follows this analysis. The predictions of the model for questions 1 and 2 combined is shown in Table 6.4. Speaker was the independent variable, and response between 0 and 100 the dependent variable. Individual participants and questions were entered as random effects. There was no effect of participant sex or age in this combined analysis. A low rating on the 1-100 scale indicates more female ratings, and higher values more male ratings.

Table 6.4: Linear regression output - predicted estimates of more or less male weighted answers

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	61.53	2.55	24.07	<.001	***
Asian male voice	-21.95	4.34	-5.05	<.001	***
Female voice	-5.67	5.02	-1.12	.260	
Gay male voice	-0.43	5.07	-0.08	.931	
Older male voice	2.52	4.75	0.53	.595	
Younger male voice	10.16	4.75	2.13	.033	*

Compared to the silent condition, the Asian male voice and younger male were significantly different, but in opposite directions. Participants listening to the Asian male voice were predicted to answer with lower values roughly corresponding to more female ratings, and those listening to the younger male voice with significantly higher male ratings. The predictions are shown graphically in Figure 6.5. The analysis was run again with the female voice as the comparison level, and both the Asian male condition and the younger male condition were significantly different. Participants listening to the Asian male were again predicted to give lower ratings of maleness,  $p=.003$  and those listening to the younger male voice higher ratings of maleness  $p=.007$ .

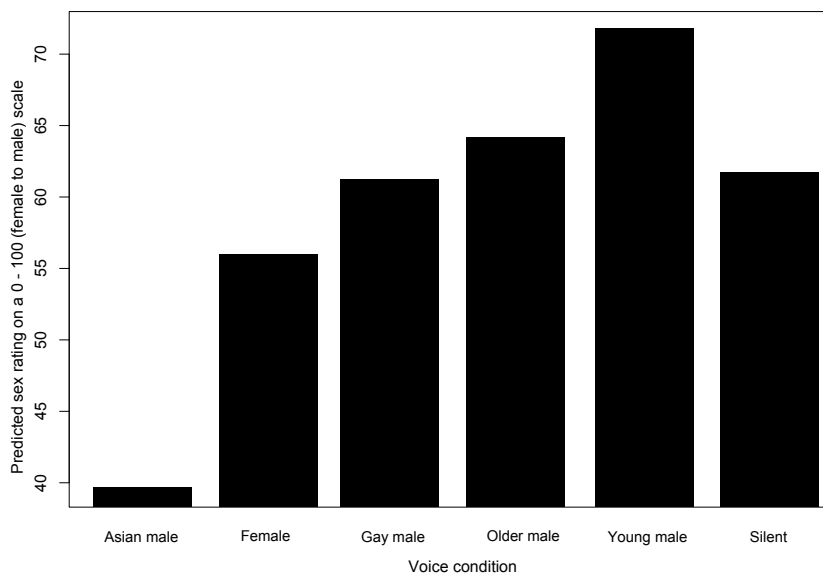


Figure 6.5: Linear regression plot: predicted value on female-male scale between 0 and 100

- Question 3) **Guess the percent of the crowd (photograph) that is male**

This question was analysed individually using a general linear model. The independent variable was voice condition and the dependent variable a response between 0 and 100, indicating the percent of the crowd the participants thought was male. There was no effect of participant sex or age. The predictions of the linear regression model can be seen in Table 6.5. All of the voice conditions were predicted to return higher estimates of males in the crowd compared to the silent condition. The predictions of the model are shown graphically in Figure 6.6, showing that participants in the silent condition were predicted to estimate a significantly lower percentage of males in the crowd. It should be noted, however, that the range of predicted probabilities was narrow, between 46 and 56 percent.

Table 6.5: Linear regression output - modelling predicted guess of the % of crowd that is male by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	43.89	2.78	15.78	<.001	***
Asian male voice	10.60	2.06	5.13	<.001	***
Female voice	6.47	2.23	2.90	.003	**
Gay male voice	10.47	2.23	4.67	<.001	***
Older male voice	7.48	2.14	3.48	<.001	***
Younger male voice	7.21	2.11	3.40	<.001	***

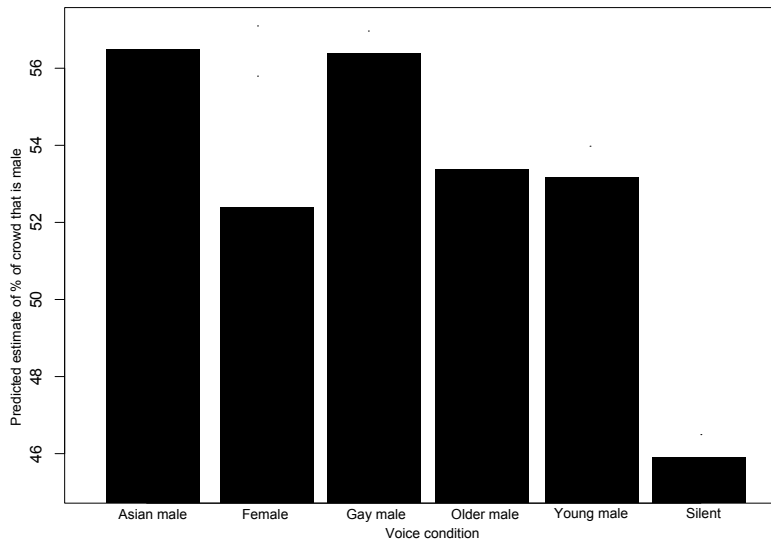


Figure 6.6: Linear regression plot: predicted guess of the % of crowd that is male

### 6.6.3 Ethnicity Themed Questions

- 1) Name a country that you associate with human rights (1 = Asian, 0 = non-Asian)
- 2) Name a language you find unpleasant to listen to (1 = Asian, 0 = non-Asian)
- 3) Guess the country of origin of the following excerpt of music (1 = Asian, 0 = non-Asian)

Although these three questions had the same hypothesis that hearing an Asian male would affect the answers differently to the other conditions, the reasons for this, and the direction of those effects, were not comparable. For the first two questions, there was no strong directional hypothesis: either spreading activation could encourage more Asian themed answers (Hypothesis 1), or participants might use a deliberate strategy where they avoid naming Asian countries because of priming awareness and the sensitive nature of feeling racist, or avoid cognitive discomfort afterwards (Hypothesis 2) (see Gawronski & Strack, 2012). For the third question, it was expected that the availability heuristic, and spreading activation, would lead to participants guessing that the music was more Asian sounding after hearing the Asian voice.

A PCA suggested that there was no underlying groupings to the question data, which was supported by low values on the correlation matrix between questions. A scatter plot of the answers also showed no clear patterning.

- 1) **Guess the country of origin of the following excerpt of music**

The answers to this question were free response. Any answer of a country that fell within the ASEAN+ 3 (see Section 1.2), was coded to 1, and all other answers to 0. The silent condition was entered as the comparison level. The independent variable was speaker, and the dependent variable was a binary response, 0 indicating non-Asian, and 1 indicating Asian. Participant sex and age were tested for but were not significant. The predictions of the logistic regression model, shown in Table 6.6, show that compared to a silent condition the Asian voice condition returned the highest predicted probability of *Asian origin* as the answer. This was in line with the working hypothesis. The female voice condition was also significantly differences in the opposite direction, with the lowest predicted probability of *Asian* responses. The model's predictions are shown graphically in Figure 6.7.

Table 6.6: Linear regression output - modelling predicted probability of replying *Asian* by voice condition

	Estimate	Std. Error	z value	p value	
(Intercept)	-0.46	0.23	-1.94	.051	.
Asian male voice	0.88	0.35	2.41	.015	*
Female voice	-1.33	0.50	-2.65	.007	**
Gay male voice	-0.30	0.41	-0.74	.452	
Older male voice	-0.26	0.38	-0.68	.496	
Younger male voice	-0.26	0.38	-0.68	.496	

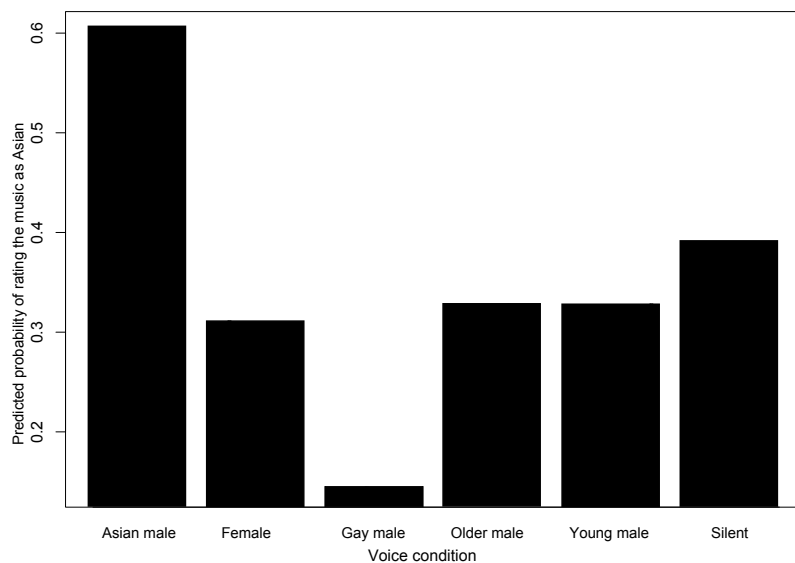


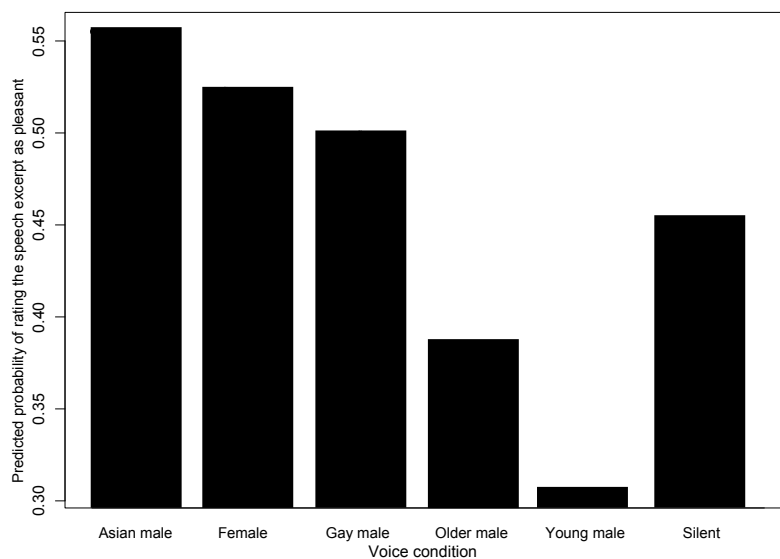
Figure 6.7: Linear regression plot: predicted probability of replying *Asian* by voice condition

- 2) **The following excerpt of speech is pleasant to listen to**

Table 6.7: Linear regression output - modelling predicted probability of agreement by voice condition

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value
(Intercept)	-0.18	0.23	-0.80	.420
Asian male voice	0.41	0.34	1.19	.232
Female voice	0.28	0.38	0.73	.465
Gay male voice	0.18	0.38	0.48	.628
Older male voice	-0.26	0.37	-0.72	.471
Younger male voice	-0.63	0.38	-1.63	.103

The silent condition was entered as the comparison level here. Speaker was the independent variable, and the dependent variable was a binary response, 0 indicating not pleasant, and 1 indicating pleasant. Participant sex and age were not significant. Looking at the predictions of the model in Table 6.7 we can see that compared to the silent condition none of the other voices had significantly higher or lower predicted probabilities of the speech being rated as pleasant. Looking at the plot of the model's predictions in Figure 6.8, there did appear to be a significant difference between voice conditions. The Asian voice was then re-entered as the comparison level in line with the question theme, and the younger male voice condition returned a significantly lower probability of *pleasant* answers ( $p=.009$ ). The Asian voice did not appear, then, to prime particularly higher ratings of pleasantness, but rather other voices, particularly the younger male voice, primed higher ratings of unpleasantness.

Figure 6.8: Linear regression plot: predicted probability of rating the speech as *pleasant* by voice condition

- 2) **Name a country that you associate with human rights violations**

The silent condition was entered as the comparison level here. Voice condition was the independent variable, and the dependent variable was a binary response, 0 indicating non-Asian, and 1 indicating an Asian country. Participant sex and age were not significant. Looking at the predictions of the model in Table 6.8 we can see that compared to the silent condition, both the female voice condition and the gay male voice condition returned significantly higher predicted *Asian* answers. The Asian male condition was the only one with a negative coefficient, indicating that participants who heard this voice were predicted to be the least likely to answer an Asian country, although this was similar to the predictions given for the silent condition. Looking at the plot of the model's predictions in Figure 6.9, there did appear to be a significant difference between voice conditions. The Asian voice was then re-entered as the comparison level in line with the question theme, and both the female and gay male conditions were predicted to have significantly more Asian answers,  $p=.018$  and  $p=.033$  respectively. One interpretation for this, although it is somewhat speculative and by no means the only account, is that when hearing western sounding voices, participants felt better about answering with non-Western countries, including those from Asia, for reasons of ingroup solidarity. For the Asian voice, they may have deliberately avoided naming Asian countries for reasons of stereotypical bias (after hearing the voice), and in the silent condition they had been primed by previous questions on ethnicity, and so like hearing the Asian voice, awareness of the possible prime could have moderated its effect.

Table 6.8: Linear regression output - modelling predicted probability of replying *Asian nations* by voice condition

	Estimate	Std. Error	<i>z</i> value	<i>p</i> value	
(Intercept)	-1.38	0.28	-4.80	<.001	***
Asian male voice	-0.06	0.43	-0.14	.888	
Female voice	1.00	0.42	2.34	.019	*
Gay male voice	0.90	0.42	2.09	.035	*
Older male voice	0.66	0.41	1.57	.114	
Younger male voice	0.02	0.45	0.05	.955	

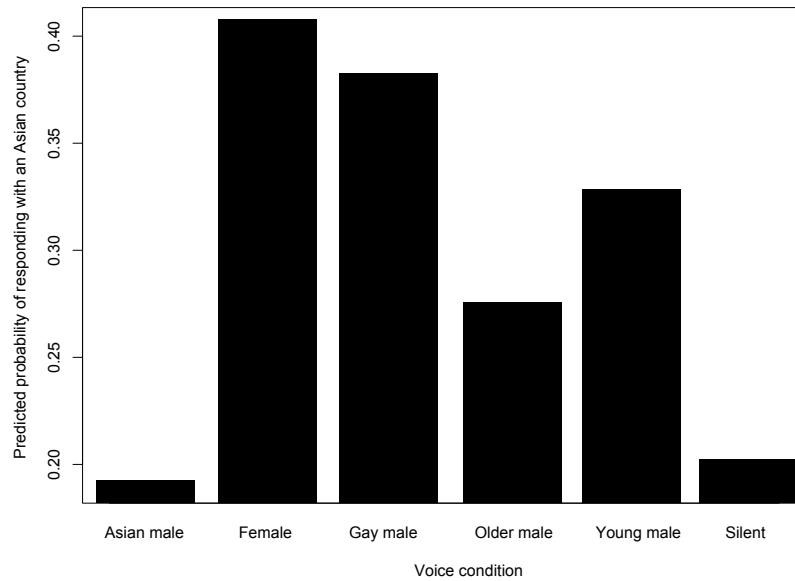


Figure 6.9: Linear regression plot: predicted probability of replying *Asian* by voice condition

#### 6.6.4 Age Themed Questions

- 1) Guess the age of the person in the photograph
- 2) What is the best age to be?
- 3) How many times per year does an average person visit their GP

A principal components analysis indicated there were no underlying groupings to the answer values in these questions. Correlations between questions were also particularly low - .09, .03 and -.03. None of the questions were thus combined for these analyses.

- 1) **What is the best age to be?**

Table 6.9: Linear regression output - predicted age value given to question by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	24.90	0.97	25.56	<.001	***
Asian male voice	1.99	1.43	1.39	.164	
Female voice	0.85	1.62	0.52	.599	
Gay male voice	1.65	1.63	1.00	.313	
Older male voice	5.09	1.55	3.28	.001	**
Younger male voice	0.70	1.55	0.45	.649	

This question was analysed using a general linear model. The independent variable was speaker, and the dependent variable was a response between 1 and 100. The silent condition was entered as the comparison level. The predictions of



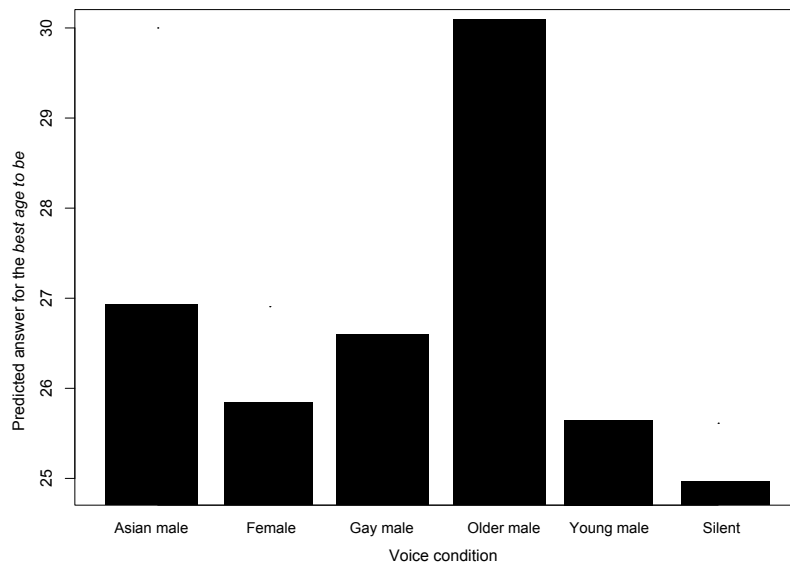


Figure 6.10: Linear regression plot: predicted age value given to question by voice condition

the model can be seen in Table 6.9. Compared to the silent condition, the older male voice condition was the only to show a significant difference, with participants predicted to give higher estimates for the best age to be. The model's predictions are shown graphically in Figure 6.10.

- 2) Please guess the age of the person in this photograph

Table 6.10: Linear regression output - predicted guess of age of male in the photograph by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	56.99	1.55	36.58	<.001	***
Asian male voice	-0.52	0.98	-0.52	.599	
Gay male voice	-0.94	1.13	-0.83	.404	
Female voice	-1.98	1.12	-1.77	.077	.
Older male voice	1.23	1.06	1.15	.204	
Younger male voice	-1.68	1.07	-1.56	.118	
Participant age	0.17	0.06	2.79	.005	**

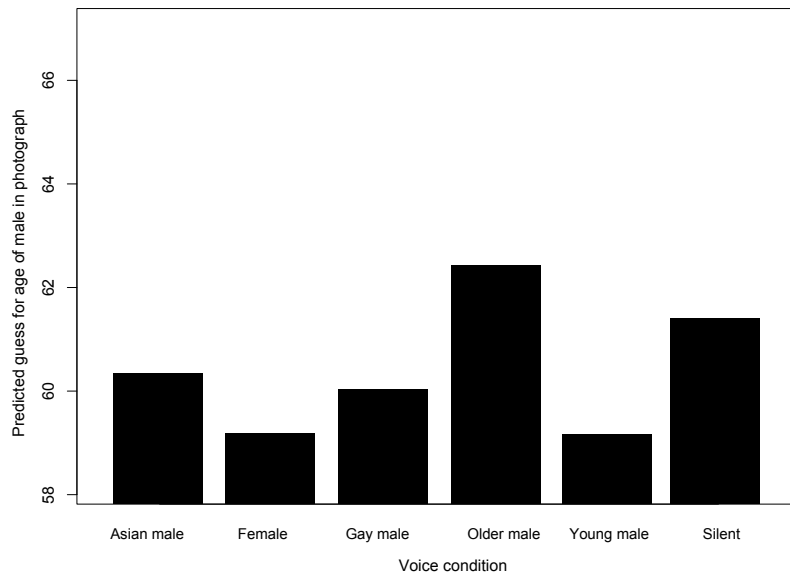


Figure 6.11: Linear regression plot: predicted guess for the age of male in the photograph by voice condition

This question was analysed using a general linear model. The independent variable was speaker, and the dependent variable was a response between 1 and 100 indicating the guessed age. The silent condition was used as the comparison level. The predictions of the linear regression model can be seen in Table 6.10. None of the voice conditions had significantly different predictions from the silent condition, but the older male voice was the only condition with a positive coefficient. The model was run again, with the older male condition entered as the comparison level, and returned a significantly different prediction between the older male condition and both the younger male voice ( $p=.013$ ) and the female voice ( $p=.008$ ). These predictions can be seen graphically in Figure 6.11. Further, the age of the participants was a significant main effect. As age increased, so too

did predicted estimates of the age of the photographed male. It must be noted however that the effect sizes, as indicated by narrow range of coefficient values and their standard deviations, was particularly small for this question.

- 3) **How many times does an average person visit their GP per year?**

Table 6.11: Linear regression output - predicted guess of yearly doctor visits by voice condition

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value	
(Intercept)	5.01	0.68	7.33	<.001	***
Asian male voice	-2.07	1.00	-2.06	.040	*
Gay male voice	-1.39	1.15	-1.21	.227	
Female voice	-2.52	1.14	-2.21	.027	*
Older male voice	-2.52	1.08	-2.32	.021	*
Younger male voice	-1.90	1.08	-1.74	.081	.

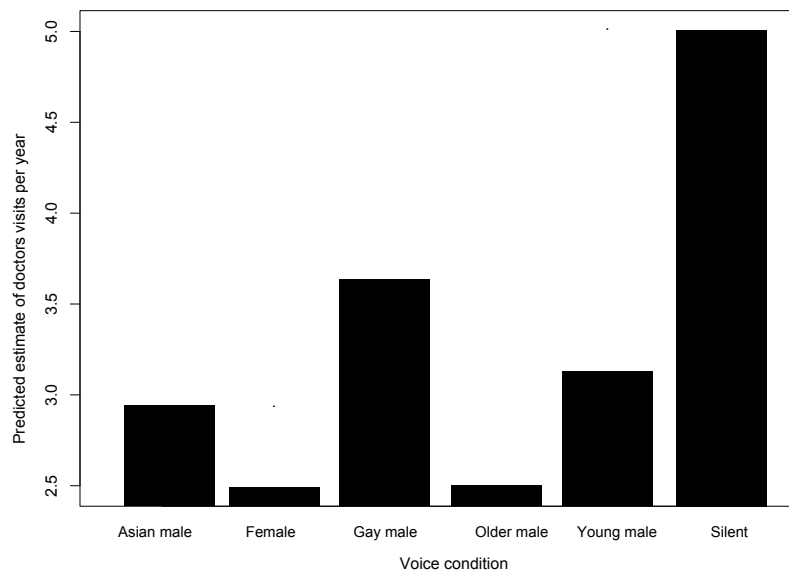


Figure 6.12: Linear regression plot: predicted guess for doctors visits per year by voice condition

This question was analysed using a general linear model. The independent variable was speaker, and the dependent variable was again a number between 0 and 100. The mean response was 3.5 visits per year. The silent condition was entered as the comparison level. The predictions of the linear regression model are presented in Table 6.11. Compared to the silent condition, the Asian male, female and older male voice conditions all returned significantly lower predicted values for yearly doctors visits. Looking at the model's predictions in Figure 6.12, it is clear that the silent condition here acts quite differently to the voice conditions, and that in absence of a voice, participants were predicted to give higher estimates for

doctors visits. Again, looking at both the coefficients and the standard errors, the effect sizes for this question were small.

### 6.6.5 Speaker Effect Overall

The questions presented above for this experiment were largely individually analysed. The effect sizes, as indicated by the coefficients, were often small and the directions of the effects over the questions was, as expected, often different. In order to assess whether voice was a warranted predictor, the responses from the linear models were combined (scale data questions) and a log likelihood comparison was conducted on two models: one with voice condition as an independent variable, and one without. The models had participant sex and age as the other fixed effects, and the dependent variable of response between 0 and 100. Both participant and individual question were entered as random effects. Comparing the model with voice condition, and the one without, gave a clear indication that having voice condition in the model explained significantly more of the variance,  $p=.003$ .

## 6.7 Summary

Before moving on to the discussion, this section summarizes the findings of the questions in this experiment. Because each question had hypotheses which were largely based on hunches, and some of them without a firm direction, I will not devote much space to interpreting them individually. For the gay themed questions asking participants to judge a visual and audio stimuli on aspects of perceived gayness, it was the younger male voice in both instances that diverged the most from the answers in the silent condition. However all of the heterosexual voices showed higher ratings of heterosexuality than both the gay voice and silent conditions. It is possible, then, that hearing anything heterosexual sounding (implicit as this may be) primed a heterosexual bias in answering. The question on Lady Gaga, however, showed that the female voice predicted the highest answers, which as I mentioned in that analysis, would be easily explained by invoking a singer-sex and voice-sex prime congruence, or a simple accessibility account. Neither explanation is supported more by this data, but it does highlight that these effects have no clear prior direction.

For the gender themed questions, there again appeared to be an effect from the younger male voice compared to the silent condition, whilst the answers in the female condition were not dissimilar to those in the silent condition. Rather unexpectedly, however, there was a strong effect from the Asian male condition, which returned significantly lower ratings of maleness for the audio and visual stimuli. The effect of the younger male would be, again, compatible with an associative network explanation, such that hearing a young, heterosexual sounding male prompted participants to rate stimuli as more male. This would not explain why the female condition did not prompt people to rate the stimuli as more female, or the Asian male condition as less male.

For the ethnicity questions, there were again some expected directions, and some unexpected. The question which asked participants to guess the origin of music showed a clear effect of the Asian male voice compared to the silent condition and other conditions. This question is a clear candidate for an availability and spreading activation account. They were presented with something ambiguous, asked to judge it, and used whatever tools at their disposal to help them, which in this case was an Asian sounding voice. The other two questions were different, and tapped into aspects of possible racism. The question that asked them to rate whether an excerpt of speech was pleasant showed that the younger male differed most from the other conditions. Participants in that condition were more likely to disagree and rate the speech as unpleasant, and those in the female condition more pleasant. This is akin to the results in Experiment 1 that showed participants in the younger male voice condition were more likely to disagree overall, and in the analysis of Experiment 1A, the least likely to support ethnicity themed questions.

For the age themed questions, no clear explanations present themselves. For the question that asked participants to state the *best age to be*, there was an effect of the older male condition compared to the silent condition, but the coefficients indicated that this was not a sizeable effect. When asked to rate the age of the photograph, participant age rather than any particular voice condition was a more significant predictor. Finally, for the question that asked about number of yearly GP visits, all of the voices appeared to differ from the silent condition. This indicates that questions about GP visits are less influenced by different voices, but

more so on hearing a person ask the question.

The results from each of these categories discount a simple category assimilation account. The only case where this evidently occurred was for the gay-related stimuli, where the heterosexual voice clearly primed participants away from rating the stimuli as more gay. The other stimuli interacted with the different voices in different ways. Like Experiment 1, a simple voice-topic matching account does not appear to be adequate here, for reasons expanded upon in the discussion.

## 6.8 Experiment 2 Discussion

In this discussion we return to one of the primary ideas presented in the introduction - voice as a heuristic. I stated at the outset that this is not a divergence from the overall automaticity argument. Automaticity, as it is used here and as discussed in the introduction, hinges on there being recently available stimuli to influence proceeding behaviour. Kahneman (2011) has argued that anchoring can be both a systematic process, and also a more automatic priming effect, where recently available stimuli can have effects on following decisions. I propose that the voices in this experiment acted as anchor primes. Anchors, because the nature of the task means attention has been paid to them, and primes because the participants were likely unaware of the effect the voice was having on that decision making process. As with Experiment 1A, the data showed that the voices participants have listened to are used in some way, and that they feature in the decision making process.

The anchoring and availability approach appears to be the most parsimonious explanation for these effects, and until further experiments are conducted exploring alternative predictions, one that is capable of handling different effect directions and magnitudes. This is not a full account, and alternative suggestions are proposed in the next few paragraphs. Shah and Oppenheimer (2008), in their paper on heuristics, mention a variety of reasons and processes offered in heuristics research. Pertinent to this experiment is the following remark: ‘The second form of this effort reduction principle suggests that decision makers access information that is easier to retrieve, either because it is computed quickly or has been made readily available through other means’. This ‘other means’, in this experiment

are the voices. Kahneman and Frederick (2002) speak of attribute substitution, where people will use easier cues to formulate an answer rather than more thought-provoking ones. Their theory supposes that once an intention to judge has been made, a search is initiated that seeks to find a reasonable value. This value could be quite specific, or rather vague, and would depend on the target and the perceivers existing knowledge (Kahneman & Frederick, 2002). The search can use readily known information, such as when estimating one's weight. In other instances, when the target is ambiguous and the information required to arrive at the correct judgement is not available, people may perform attribute substitution. Like the 'other means' that Shah and Oppenheimer spoke of, the substituted attributes in this experiment are the voices.

The information required to answer the questions given to the participants was either not accessible, or simply impossible given their circumstances. There is no way they could have known with certainty the age of the man in the photograph unless they personally knew my father. The search for the correct answer called upon their own knowledge of faces and ages, perhaps their own parents and so forth. But as the analysis showed, there were clear differences between the conditions. These differences were not always large, but in the realm of priming, that was not to be expected. In other words, their search did not stop with their own exemplars. At some level, the voice entered their evaluative process. As Gigerenzer and Gaissmaier (2011) have argued, using available environmental stimuli may not constitute a bad decision, or a detrimental bias. There may be good reasons to consider that the anchor, or prime, is relevant to the task. Schwarz (2004) puts this succinctly writing "People generally assume that the thoughts that come to mind while thinking about X are relevant to X - or why else would they come to mind at this point?" (Schwarz, 2004, p 342). Using the Asian male speaker as an example might help to illustrate this point. Participants might well conclude that owing to the rarity of a survey being presented using audio questions, and the even rarer possibility that a survey for native speakers of English would be recorded by a non-native speaker, that the Asian voice indicates an Asian bias to the research. It would be perfectly rational, in their minds, to think that if they are being asked to guess the origin of music, that it must be related to the voice that is asking. Being swayed by the voice does not seem so illogical when viewed this way.

A study conducted by Hay and Drager (2010) provides a language related example of how this process could work. In their experiment, they sought to investigate whether the presence of country-specific cues could affect participants' perception of speech. Their participants sat in a room with either a stuffed toy kangaroo or a kiwi bird, which was ostensibly unrelated to the task they were to complete. They then listened to sentences, recorded by a male speaker, that had been designed to include particular vowels. After hearing the sentences, the participants were asked to select a vowel from a synthesized vowel continuum that best matched the vowels they had heard the speaker produce. This continuum was designed at one end to sound more Australian, and at the other end, more New Zealand. The authors found some evidence for country-congruent priming, particularly driven by the female participants and participants from higher socio-economic background, where exposure to the different toys led to vowel selection differences between the two conditions. I suspect that the processes I have outlined above were partly responsible for this. Participants did not have sufficient information to accurately determine which vowels they had heard, and in their search, they appear to have used the stuffed toy as a type of anchor. The mere presence of the toy would have likely involved a spreading activation in the associative network. The increased availability of Kiwi or Australian traits, coupled with the desire to conserve mental energy and stop searching for the correct answer, is certainly a candidate for the effects they found and so too for the results presented in this chapter.

Spreading activation (see Collins & Loftus, 1975) is an integral part of this heuristics explanation, but for some questions more than others. It is certainly an incomplete explanation, and indeed it would be difficult to test this spreading activation account given how diverse it can be across individuals. Answering that a male in the photograph is younger than he really is could be spreading activation for one person, but affect related for another - this will need careful testing.

The question 'Please guess the country of origin of the following excerpt of music' serves as a good reference point to discuss this idea. The music was difficult to identify in that both the range of answers given in the experiment, and the difficulty of my Mandarin speaking colleagues to identify it, made it an ideal choice as an ambiguous prime. For some, or all, of the questions spreading activation



theory is a reasonable candidate for the effects (see Section 2.1). Using the Asian voice as an example to help illustrate this, Figure 6.13 shows one possible representation of this for a listener.

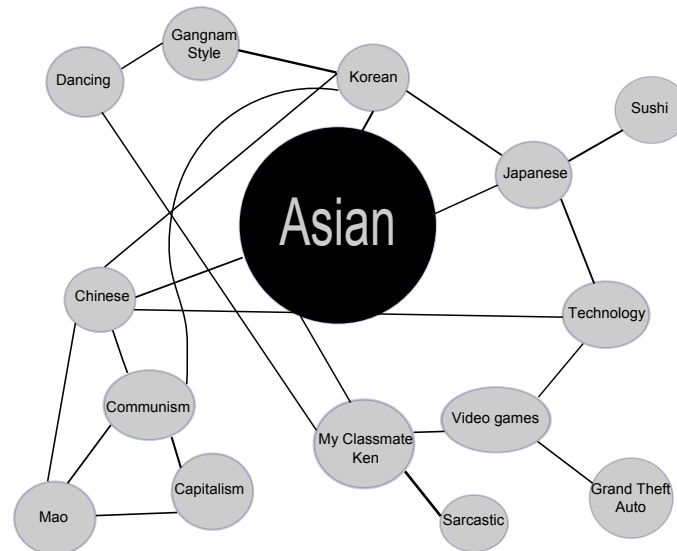


Figure 6.13: One possible spreading activation account for a listener hearing an Asian voice

When participants were asked to guess where this music was from, they had to use what cues they could to make their guess. Stimuli internal cues such as the rhythm, tone, musical key and such like are some of the cues that would inform the guesses. When the Asian male voice preceded this music excerpt, however, spreading activation would already have occurred, indeed from the very first question. Concepts related to Asia, and/or male, should have been active for these participants, and in their search for an answer to this ambiguous question, the activated concepts were highly available. As Cesario et al. (2006) have written, “When one perceives a social category member, presumably that category becomes activated along with all associated traits” (Cesario et al., 2006, p 895). Asian could have activated Chinese, which could be linked to tonality, which itself could be linked to unintelligibility. In every one of the questions presented above, spreading activation is a valid candidate towards an explanation. As another example, when asked to estimate their confidence that a photographed male was heterosexual, participants hearing any other heterosexual voice appear to have rated it as more heterosexual. If heterosexuality is something they had implicitly activated from the voice (which would have been aided by questions with a sexual orientation theme), then upon presentation of the photograph, selecting one of these active concepts

would conserve mental energy by allowing heuristic decision making to proceed. The participant need not expend too much mental energy by systematically listing points for and against the man's heterosexuality. Further, it might help them avoid simply providing a random guess because the task is too difficult. The suggestion (or the *anchor*, in Kahneman's terms) of heterosexuality from the voice may provide them with a foothold. Knowing which attitudes are directly activated is a difficult task. As Wittenbrink et al. (2001) have argued, attitudes are not a single entity stored in memory, waiting to be retrieved. More likely is that they are formed from a variety of discrepant memory contents, and which of these are accessed and combined to form any currently active attitude depend on what is made salient.

## 6.9 When category activation isn't enough

As mentioned at the outset of this chapter, whilst there is a lack of research on voice and stimuli priming, there is enough literature to invest in a category assimilation hypothesis, or in some cases one of contrast. The results presented above do not, however, support anything as simple as that. There are still obvious questions for which the results here offer no answers. That it was the Asian male voice that primed an increase in perceived femaleness for Kathleen Turner's voice, for example, is at odds with a simple category activation/assimilation account. I looked at an audio recording of the Asian speaker's voice acoustically using the phonetics program *praat*. It was higher in pitch than the other male voices, at around 600 hertz. The gay male's voice was around 550 hertz, the older male's voice around 470 hertz and the younger male's voice around 380 hertz. This is certainly something that could be followed up in the future, where it is tested experimentally whether or not laboratory controlled pitch can prime ratings of sex. Whilst spreading activation and anchoring accounts could be offered, it would require some manoeuvring along the lines of: higher pitched voices brought to mind past examples of higher pitched voices, and through the associative network, more examples of female speakers became active which consequently primed greater ratings of femaleness. 100% of participants rated the Asian male as sounding male, so any effect of pitch here would be implicit. A simple explanation would require a simple mechanism, and as researchers of priming and automatic behaviour have long noted, priming is anything but simple. A number

of researchers have made the point that when it comes to social traits, and their relationship to behaviour, precisely what is activated is not always well understood (see for examples Forehand, Deshpandé, & Reed, Americus, 2002; Dijksterhuis & van Knippenberg, 1998; Bohner & Chaiken, 1994; Wheeler & Berger, 2013). This is an ongoing concern, and neatly summed up by DeMarree and Loersch who write “Despite the diverse set of potential effects produced by any given prime, it is still not well understood when one particular outcome should be expected to occur over any other” (DeMarree & Loersch, 2009). To answer why the particular voices did particular things would almost require an individual study for each type of voice, and each type of prime. These are important questions that deserve future exploration.

I hope to have shown in this chapter that, like those experiments described in the previous chapter, listeners use voice information to aid their decisions for ostensibly unrelated tasks. It is by no means their main cue, as the coefficients indicate, and the fleeting nature of priming (Sedikides & Skowronski, 1991; Shanks et al., 2013) means that quite different effects could be found when replicated. What I would predict, however, is finding differences across conditions: that in a given circumstance where voices are present, individuals take information from those voices that, for reasons not yet well understood, feed into their judgements. In Experiment 1, the unrelated task was to agree or disagree with a socially themed statement. In this experiment, the unrelated task was to judge a picture, or audio, or to give a random guess to a question to which there was no obvious answer. At no point were participants asked to relate the voices to the task at hand, yet these results indicate that they do extract information from the voices for the task, either by construing relevance or by invoking similar concepts. In the opening sentence of this paragraph, I wrote *ostensibly unrelated*, and this point deserves emphasis. The voices, and the task demands, are only unrelated as far as rational, normative decision making goes. If we have learned anything from the priming and heuristics literature it is that we simply do not know what people consider relevant to the task. With the myriad associations and activations that occur in these situations, relevance may be something as abstract as the number on a spinning wheel (Tversky & Kahneman, 1974), or a voice asking you to guess the number of males in a crowd of people.

## Chapter 7

# Experiment 3: Voice and Reaction Times

### 7.1 Introduction

In order for a prime to effect a behavioural change based on group characteristics, it must activate some group category. In Chapter 5, it was argued that the accessibility of sex as a social category led to different behavioural outcomes based on the sex of the voice. In this chapter, 4 experiments are described which sought to establish whether the age category, or age stereotypes, could drive voice priming effects. The experiments tested whether or not hearing an elderly voice could slow participants down in a speeded reaction time task, consistent with prior research on age themed priming. Experiments 3A and 3B look at the age of voice, and then compare that to a silent condition in Experiment 3C. Experiment 3D then uses a local and non-local accent to test whether or not voice familiarity, rather than age, might be responsible. This was first tested with younger participants, then with older participants. Unlike the experiments in Chapter 5, where participants were able to consider their answers before making their choices, I wished to see if similar speech priming effects would occur much more quickly and automatically.

Kawakami et al. (2002) conducted an experiment that, like the present one, sought to affect the reaction times of participants by activating the elderly stereotype. They primed their participants by having them perform an age

categorization task on a series of photographs of men and women, half of which were old. They were then asked to perform a lexical decision task, where they responded either to country labels or to stereotypic traits about age categories. They found two main findings of interest. First, participants primed with the *elderly* category responded more slowly to the semantically neutral country labels. Second, participants primed with the *elderly* category reversed this effect (responded more quickly) when they were presented with elderly congruent traits. That is to say that the elderly prime produced elderly congruent behaviour when participants' attention was not drawn towards the elderly, but showed no effect when attention was brought to bear on the category. In the present experiment, the task is age-neutral (that is to say, does not require participants to explicitly think about or see any stimuli with an age theme) and so from their findings we would predict elderly congruent behaviour.

## 7.2 The Elderly Stereotype

The elderly stereotype is clearly not uni-faceted. It is instead a collection of different traits and associations - see Hummert (1990); Hummert, Shaner, and Garstka (1995) for reviews. As Hummert (1990) has stated, there is a superordinate category of *older adult*, and branching down from this are multiple positive and negative subcategories, although the negatives tend to outweigh the positives for the elderly stereotype. Further, and importantly for this chapter, the negative stereotypes are typically found when the group category is activated, and less so for idiosyncratic members of a group, such as the Dutch Queen Mother as a member of the *elderly* category (Dijksterhuis et al., 1998).

Stereotype duality was nicely encapsulated by Brewer and Lui (1984), who gave the examples of *senior citizen*, and *elder statesman* to highlight the discrepancies that can arise. Both denote elderly gentlemen, but have identifiably different sets of associations. Moreover, the content of stereotypes can not only vary within the superordinate category, as in the previous example, but between different groups holding the stereotype. Levy and Langer (1994) have also shown that the effect of the elderly stereotype was culturally conditioned. In a culture where the stereotype is predominantly negative, such as the USA, it hampered the performance of

elderly U.S. citizens on a range of memory recall tasks. However, this effect, called *stereotype threat*, can be overcome. Levy (1996) and Levy, Ashman, and Dror (2000) have conducted work showing that when the positive aspects of a stereotype are made salient task performance need not decline for elderly participants. Levy (1996) presented her elderly participants with 12 positive and negative primes (such as *astute* for positive, and *Alzheimer's* for negative). Positive stereotype priming improved her elderly participants' (60+ in age) scores in a memory task, compared to their scores before the priming intervention. The perspective from which the stereotype is seen, and the duality of its associations, are directly relevant to this study. Each individual participant could, based on their own history and experience with particular voices and groups, activate one of many dominant associations from the voices. The younger group might be more likely to activate the senior citizen version of the stereotype, thus hampering performance. The elderly participants may be more likely to activate the statesman version, either negating or reversing the effect. This is expanded upon in the discussion of age stereotype literature in the following paragraphs.

Whilst forgetfulness was not directly tested in this thesis, we can view forgetfulness as a type of impairment, and studies that have used memory performance use similar underlying assumptions about stereotype content and task performance. Dijksterhuis et al. (2000) looked at whether the amount of contact people have had with the elderly would affect how they perform in a memory task after being primed with the elderly stereotype. Their results showed that as hypothesised, after being primed, people who had reported higher levels of contact with elderly people performed worse in the task. Those participants who reported little contact with the elderly, however, did not show priming effects. A lack of contact with the elderly would also, however, be consistent with predictions from a *Perceptual Fluency* account, namely that less contact with a stimuli should retard processing in some way (Jacoby & Dallas, 1981). Common experience tells us that younger people generally spend more time with people of a similar age to themselves, and so too for older people. If the results were to show that older people performed worse in the task when hearing an older voice, this would offer further support to that prediction.

Hess, Hinson, and Statham (2004) have provided further evidence of elderly

stereotype priming, particularly on how negative and positive ageing stereotypes interact with these effects and whether implicit versus explicit cues act to constrain these effects. The researchers found again that elderly stereotypes do indeed hamper performance in memory recall tasks, but that this was only for the negative elderly stereotype, consistent with Levy (1996). In their negative prime condition, where participants had conducted a scrambled sentence task containing words such as *confused* and *senile*, participants' recall ability was worse than in the positive prime condition, which had words like *dignified* and *insightful*. Crucially, they did not find these results for younger participants, which they suggest indicates group-based expectations about performance. This is inconsistent with Bargh et al.'s walking study (1996), where young participants slowed their walking speed in response to elderly priming. It should be noted, however, that task differences as well as behavioural differences make parallels harder to draw. Further, the extent to which the findings should mirror that of Bargh et al. (1996) given that the context, materials, and paradigm were different is unclear. In Hess et al.'s experiment (2004), when participants were made aware of the primes by having the salient words highlighted in yellow to stand out from the other words, they were able to counteract its effects (that is to say, their performance was less hampered).

It should be highlighted again that behavioural contrast effects are common in priming studies. Herr (1986) was one of the first to look at how the extremity of exemplars can affect priming. His experiment showed that if a category exemplar was extreme enough, it would significantly predict a contrast effect in participants' judgements in an unrelated task, rather than the assimilative effects found for more moderate exemplars. His study primed participants with hostility - this could be moderate hostility (Alice Cooper, for example), or extreme hostility (Hitler and Ayatollah Khomeini).

Dijksterhuis and van Knippenberg (1998) conducted an experiment in which they tested whether or not the activation of a group stereotype, the elderly for example, has the same effect as activating a particularly salient exemplar from that category, in their case the Dutch Queen Mother. The authors argued that whilst it is clear that a direct stereotype-behaviour assimilation link may exist, there are times when the individual, who acts as the stimulus, is foregrounded over the stereotype. In their words "...the individual also constitutes the social stimulus.

Plainly put, seeing an individual sometimes also involves seeing the individual in front of the stereotype...” (Dijksterhuis & van Knippenberg, 1998, p 863). Taking direct inspiration from Bargh et al. (1996), they activated the elderly stereotype using a scrambled sentence task. Two conditions followed: one where participants were asked directly about the Queen Mother, and one where the questions were age-neutral. The researchers then timed how long it took for participants to leave the experimental location. They found that when primed with a specific exemplar of the elderly category, participants in fact walked more quickly from the experiment than in the neutral prime condition. This is, of course, contrary to the effect that Bargh et al. (1996) found when activating the macro *elderly* category. These studies are referenced because the voices used in this experiment were neither extreme (a child, or someone in their nineties for example) or idiosyncratic (speech impediment or unusual breathiness, for example). Any contrast effects found would then likely be due to participant/stereotype congruency rather than exemplar extremity.

### 7.2.1 Elderly Speech

I have discussed how exposure to pertinent information can prime human behaviour in expected directions, through the activation of stereotypes. Further, I have looked at how the elderly stereotype is generally a negative one, particularly in western cultures. Studies on priming such as those already discussed have used varying stimuli - words presented in isolation, scrambled sentences, images of popular soft drinks, and stuffed toys. In the present experiments, the aim was to establish whether a human voice that could accurately be assigned to a demographic group could act similarly to prime behaviour. An elderly voice was chosen because a number of previous studies have shown that individuals can, to varying degrees, guess a speaker’s age from hearing their voice alone, and as the studies in the previous section have shown, the elderly stereotype is one with a history of showing priming effects, particularly assimilation effects.

Allport and Cantril (1934), in one of the earliest studies to look at the information listeners receive from voice quality, tested the ability of listeners to guess speaker age from recordings. The average ages reported for their speakers were 25, 37 and 41, remarkably close to the real ages of 27, 36 and 51, although the



authors report a central tendency around the ages of 35 to 40. The authors also found that listeners were able at levels more than chance to guess the personality traits of a speaker from their voice, particularly in the dominant/passive pair. This was an early indication that not only are voices heard in terms of their broad characteristics, such as age, but also individual dimensions such as *miserable*. In this chapter, these individual personality traits of the voice would be involved in what particular version of a stereotype the listeners activate, and what they believe about the speaker they are hearing. This is likely to differ for older and younger participants. This thesis does not test what personality traits listeners extract from voices. Before doing that, it should first be established whether voices can indeed prime differences in task conditions, and from the directions of those effects, this is something that could be explored in future.

Braun (1996) asked two groups of listeners to judge speaker age based on 45 second recordings. The first group were trained phoneticians, and the second group university students. The estimations of both groups yielded strongly correlated scores between perceived age (PA) and chronological age (CA), with a mean difference in PA and CA of 5.9 years for the phoneticians and 6.5 years for the student group. Schötz (2006) as part of her PhD looked at the perception of age from voice stimuli, and again found that people are generally good at guessing, with the mean difference between PA and CA 7.9 years. She notes, however, that some people got it markedly wrong, and also as Allport and Cantril found, were reluctant to make extreme judgements of age in the upper or lower direction, resulting in older voices being more typically identified as younger, and vice versa. These studies indicate that we can reasonably expect people in an experimental condition to identify an older male voice and a younger male voice from audio recordings.

### **7.3 Experiment 3A: Younger Participants**

#### **Hypothesis**

In line with the previous hypotheses for Experiments 1 and 2, it was proposed that individual voices would have effects on task performance. In this specific experiment, the hypothesis was that when listening to an elderly sounding voice,

young participants would react more slowly in a reaction time test than when they were listening to a younger voice. This is in line with the assimilation effects typically found in category-activation and priming research.

### 7.3.1 Reaction Time Task

Participants took part in a visual reaction time test created using E-Prime software. This was conducted at the University of Canterbury. Participants saw images of different shapes and colours, and were asked to decide whether the accompanying word was an accurate description of the image. In conjunction with the visual information, they also heard pre-recorded speech throughout the experiment. The task chosen requires a certain amount of re-thinking, and whilst it is not difficult is certainly mentally draining. Participants come to expect the word to match either the shape or the colour, but when having to attend to both features are under processing load, and it is this processing load which is said to more easily allow priming effects (Bodenhausen & Lichtenstein, 1987; R. Wyer & Carlston, 1979; Kahneman & Frederick, 2002) The image below depicts the set-up.

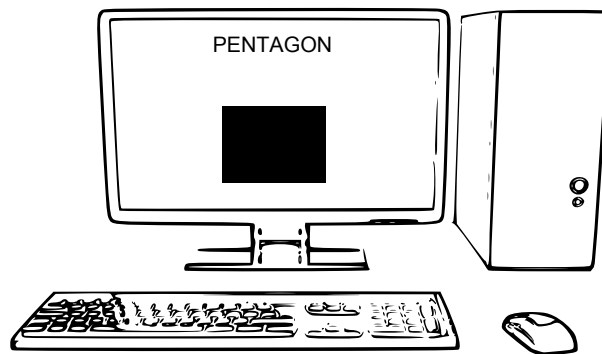


Figure 7.1: Exp 3: Experiment set-up

### 7.3.2 Audio Stimuli

A script was created, consisting of task instructions and task encouragement, for the speakers to record. Two speakers recorded task instructions and encouragement, under the pseudonyms of *Stephen Evans* and *David Jones*. The words themselves were not important for this task, but rather allowed audio to be inserted into the

reaction time experiment in a reasonably believable manner. They were recorded using an AKG boundary microphone in a quiet office. As with Experiment 1, readjustments to the microphone position and recording volume were made to compensate for significant differences in speaker amplitudes. The older male recorded first, and the younger male took two recording attempts to match the speed of the older male. The length of each audio clip is considered equal, inasmuch as length differences should not affect the participants' responses.

Below is the script that the speakers recorded and the participants heard. The script of Experiment 2 is in brackets, as it had slight changes in order to correspond to two different experiments and speakers, as the participants were told there were.

*'Experiment 1(2), recorded by Stephen Jones (David Evans). Please remember that once this experiment has begun it cannot be paused. For this reason we suggest putting your phone on silent or turning it off. The experiment should not take longer than 15 minutes. Colour Image Task: In this experiment, you will be presented with various images of different colours. If the word above the image is an accurate description of the image, you must press z. If the word above the image is not an accurate description of the image, then you must press m. In the first section, the computer will tell you if your responses are correct. Instructions for the following sections will follow. Remember, press z for correct and m for incorrect.'*

After the 40 practice trials, they heard:

*'Good work. In this next section the task will be a little more difficult. Also, the computer will no longer give you feedback. When you enter a response for each trial the next trial will begin immediately'*

After another 30 trials, they heard:

*'Now that you have completed the first trial, the computer will randomise the order of the images and begin another trial of roughly similar length'*

After another 30 trials, they heard:

*‘That was the end of trial 2. There is one more trial to go before the end of the experiment (Experiment 2). You may take a short break while the computer prepares for the final trial.’*

After another 30 trials, they heard:

*‘You have now completed all the trials of Experiment 1(2). Before starting Experiment 2 please take a break to stretch your legs and rest your eyes. When you return you will be asked to start Experiment 2. Please press space before taking a break (Thank you for your participation. Please inform the experimenter so that they can check it has been recorded and saved properly.)’*

### 7.3.3 Speaker Ratings

The speakers were chosen to broadly represent the categories of ‘older male’ and ‘younger male’ in New Zealand. To control for voice sex, only males were chosen to record as I had an older male readily available to record for me. For consistency, both males were New Zealand born, heterosexual, and unaware of our hypothesis. Speaker 1 was a 20 year old student, and Speaker 2 a 70 year old retiree. Their voices were pre-tested in a ratings experiment, where 20 mostly undergraduate male and female raters judged them on specified criteria. The average age of Speaker 1 was rated at 23 years old and Speaker 2 rated as 75 years old. Both speakers were judged exclusively male. Raters unanimously reported that Speaker 1 was from New Zealand, whilst 85% thought Speaker 2 was from New Zealand. Speaker 1 was judged as heterosexual by 90% of raters, and 100% rated Speaker 2 as heterosexual. Further to this, a number of raters offered qualitative comments. These were designed to indicate to me if there was anything particularly unusual about the voice, or whether a participant recognised the speaker, and are not used in the analysis.

### 7.3.4 Procedure

The experiment consisted of two halves. The first half had either the younger or older voice played at set points across the trials. The second half had whichever voice was not heard in the first half. This was counterbalanced for participants, so half heard the young voice first. Each half consisted of a practice phase of 40 trials, and then 3 blocks of 30 trials. In the practice phase for each part, participants received accuracy feedback to familiarize them with task expectations. In the following trial blocks, there was no accuracy feedback. Participants were instead instructed to ‘hurry up’ (text flashed on screen) if they took longer than 1 second to respond. One second was chosen as it is long enough to be considered a cognitive response rather than simply a motor reflex (see Ratcliff 1993) but not so long that it allows participants to slow right down and systematically respond to each stimulus presentation.

For each trial, participants saw a shape and word on screen. Thus for each part, participants saw a total of 130 trials of shape and word combinations (40 in the practice phase, and 3 sets of 30 thereafter). The possible shapes were: a circle, hexagon, square, triangle, trapezoid or oblong. The shapes were in one of the following colours: red, green, blue, pink, black, and yellow. The order of coloured shapes was random. Above the coloured shape was a word: the word either referred to the colour of the object, or to its shape. See Figure 7.1 for a graphical representation. Once participants had completed both parts of the experiment, they were debriefed and given a 10 dollar shopping voucher.

### 7.3.5 Participants

Participants were twenty University of Canterbury undergraduate students who had responded to campus-wide advertisements looking for research participants. They were all young females from New Zealand, with an age range between 18 and 26 years old, and a mean age of 21 years old. This demographic was targeted both because prior researchers in the department had reported better success at recruiting women, and also to control for participant variation in the analysis, particularly due to smaller participant numbers. They completed the experiment on a PC at the New Zealand Institute of Language, Brain and Behaviour.

### 7.3.6 Analysis

The data were first inspected to remove reaction times that were likely under the realm of normal cognitive ability, or too long to be confident they were concentrating on the task. This serves both as an attempt to report only reaction times that we can reasonably assume are task related, and also to boost the power of the statistical model. Without knowing exactly what each subject was doing for each stimulus response, it is uncertain whether the extreme values are invalid. In other words, the real reaction times of the processing under study overlap with the outliers. The approach taken here comes from Ratcliff, who advised researchers to reduce the effects of outliers without losing too much data of interest (Ratcliff, 1993). In the realm of attitudes and cognition, particularly IAT tests, 300 milliseconds and 3000 milliseconds are commonly used cut-off values (Greenwald, Banaji and Nosek, 2003). In stroop tasks, which are of a similar nature to this colour/shape task, other researchers have used cut-off values of 150 milliseconds for the lower limit, and 2000 milliseconds for the upper limit. 250 milliseconds is used here as a lower limit and 3000 milliseconds for the upper limit. This was done after discussion with a statistician who also looked at the spread and skew of the reaction time values. One participant was removed as the majority of her responses were around the 200 to 400 milliseconds mark (indicating that she might have been pressing without necessarily looking at the screen), along with 20 data points from other participants over the upper limit. 19 subjects and a total of 4920 trials (99% of the available trials) remained for statistical analysis. As the measure of interest here is the effect the voices have on reaction speed, and not on cognitive performance per se, there was no a priori reason to exclude practice trials from the analysis as they should also be open to priming effects, particularly as they are at the start of the experiment. Differences between practice and non-practice blocks were checked to ensure that these trials warranted inclusion.

The data were then plotted to inspect the distribution and skewness. As is typical for reaction time tasks, data were positively skewed. A log transformation was applied to correct for skew and to prepare data for the statistical analyses. A plot of the data with the outlier cut-offs applied is shown below Figure 7.2.

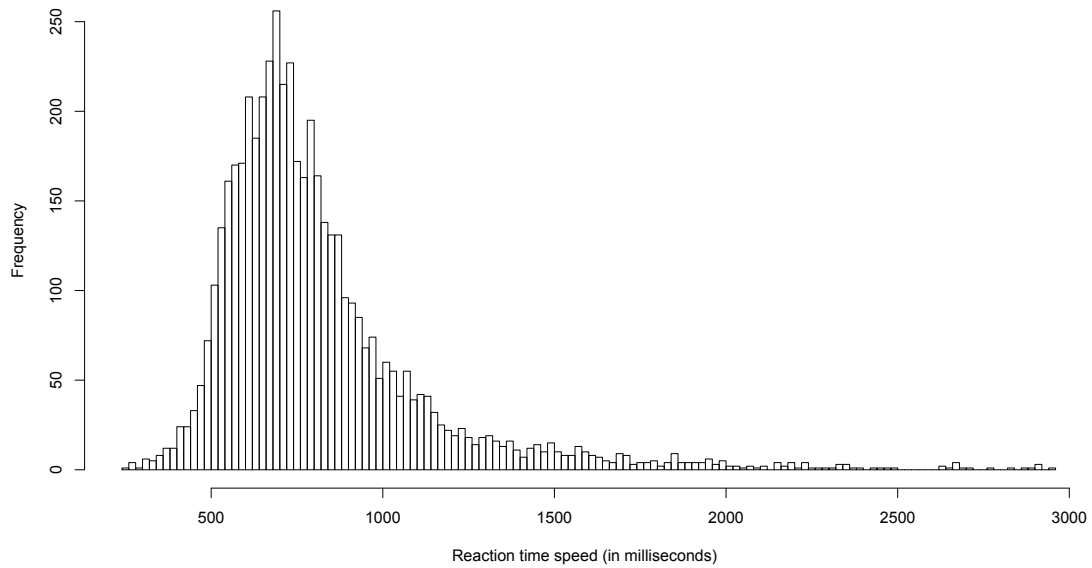


Figure 7.2: Spread of reaction times for younger participants

### 7.3.7 Results

The median reaction time for subject reaction times was 737 milliseconds for the non-practice trials, within the range of normal reaction times for this type of task. The median reaction times for each condition were 753 milliseconds for the older voice condition, and 722 milliseconds for the young voice condition. The median reaction time for the practice trials was 753 milliseconds. The mean accuracy score was 90% in the older voice condition, and 91% in the young voice condition. The mean accuracy in the practice trials was 92%.

The RT data was analysed using a linear mixed effects model. A fully fit model was first tested for with: voice age, trial number, and order of presentation as fixed effects, and: participants, voice age, and stimulus type as random effects. After pairwise model comparisons, the model that best fit the data had the fixed effects of voice age and trial number (using a 3-knot restricted cubic spline), the random effect of participant, and the dependent measure of reaction time. Significant differences in accuracy were tested for. There were no differences either between voices,  $p=.178$ , or between the practice and non-practice trials,  $p=.175$ . Further, this model was compared to an identical model without voice age as a predictor,

and the model with voice age was a significantly better fit,  $p < .001$ . The intercept is the log reaction time when the voice age is young.

Table 7.1: Mixed model output - modelling predicted reaction time by voice age and trial

	Estimate	S.E.	$t$ value	$p$ value	
(Intercept)	6.844	0.032	210.41	<.001	***
Older voice	0.035	0.008	4.27	<.001	***
Trials (1st point)	-0.002	0.00	-14.71	<.001	***
Trials (2nd point)	0.001	0.00	7.96	<.001	***

The predictions of the model are shown in Table 7.1. Participants during the older voice condition were predicted to be slower than when in the young voice condition, as indicated by the positive coefficient for *Older voice*. As they progressed through the trials, their reaction times decreased as is to be expected from learning effects.

Figure 7.3 plots the effects of the two voice conditions on reaction times. It has been back transformed from the log values, so the y-axis represents actual predicted reaction times in milliseconds, giving an indication of effect size. When listening to the older voice, participants had a predicted average RT of 760 milliseconds, compared to 735 milliseconds when listening to the young voice.



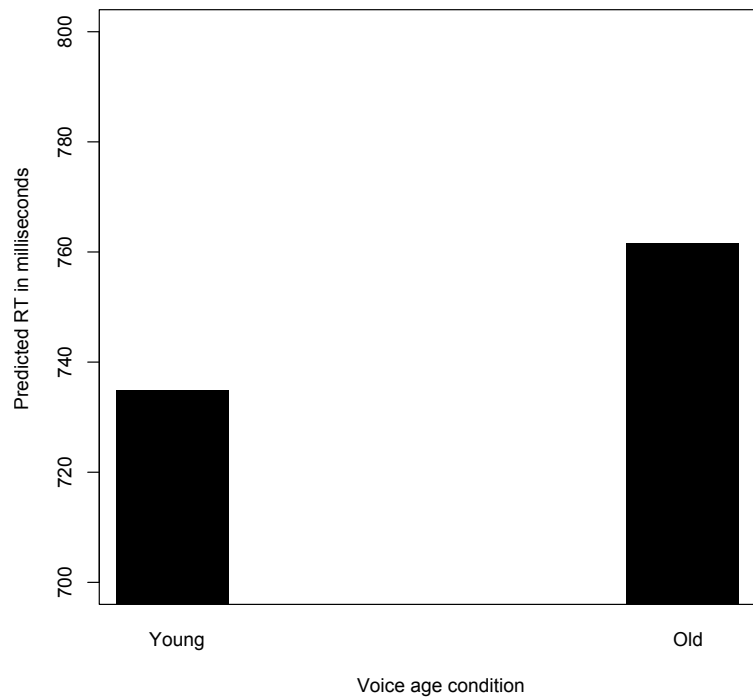


Figure 7.3: Mixed model plot: predicted reaction times by voice condition

In Figure 7.4 it is clear that participants became quicker at the task as they progressed through the trials, and then towards the end begin to show stable response times. Their accuracy, although not significant (likely due to the high mean accuracy rate of 91%) also improved. This is common for a task of this nature, where participants become accustomed to what is required of them.

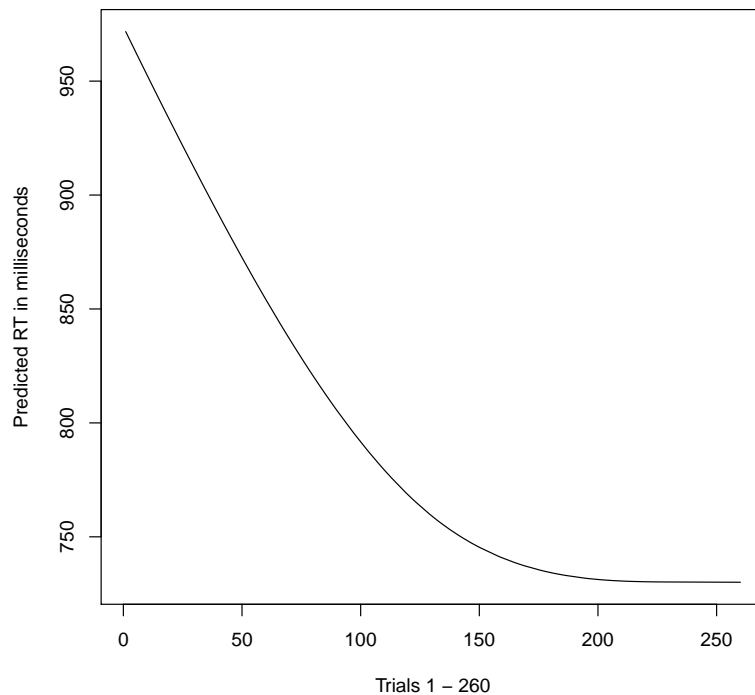


Figure 7.4: Mixed model plot: predicted reaction time by trial number

I also looked at whether comparing the first 130 trials for the two conditions would return any significant difference - that is, whether those participants hearing the young voice for the first 130 trials were quicker than those hearing the older voice. They were not ( $p=.208$ ). I then looked at whether there was a difference between the two practice blocks for the two conditions. Using the same model as above, the results confirmed a significant difference. For the two blocks of 40 practice trials, young participants hearing the younger voice were predicted to be faster than those hearing the older voice ( $p<.001$ ). The effect of voice condition, then, was stronger at the beginning of the blocks, particularly during the practice trials. As their speed increased into the non-practice trials, this effect was less.

To summarise, when 19 young participants completed an attention-consuming reaction time test, with an older and a younger voice used to provide audio instructions, their predicted reaction times were slower for the older voice condition compared to the young voice condition, regardless of the order of presentation. Thus far, the results are in line with the hypothesis presented at the beginning of this

chapter that the elderly voice would result in slower reaction times for younger participants.

## 7.4 Experiment 3B: Older Participants

This second instantiation of Experiment 3 investigates whether the effect of voice age would be the same for older participants as for younger participants. As outlined in the review of the literature, there is evidence both of category trait assimilation, and also of contrast. These effects can be a function of self-stereotype threats and goals, and also due to group versus exemplar priming. I wanted to look at whether the effect was driven by the stimulus itself, or was a product of both the stimulus and listener. Whether or not older participants patterned in the same direction as the young participants would help to narrow down the driving force of the priming effects.

### 7.4.1 Method and Participants

Experiment 3B is a replication of 3A, using identical stimuli and set-up, thus please see Experiment 3A for details on the method. In Experiment 3B, participants were 20 older New Zealanders, 14 females and 6 males. The mean age was 71 years, with a range between 63 and 79 years. These participants were recruited from researchers working with elderly speech at the University of Canterbury's Department of Communication Disorders. They had no recorded speech or cognitive impairments. There was a procedural difference in 3B which added two extra conditions: a 'young voice' only condition, and an 'old voice' only condition, each with 260 trials as in Experiment 3A. This was done to assess whether two distinct groups of participants, one hearing a young voice and one hearing an older voice, would show differences to each other. The distribution of participants across conditions was 4 females and 1 male for condition 1 (old voice first), 4 females and 1 male for condition 2 (young voice first), and then 3 females and 2 males for conditions 3 and 4 (old and young voice only).

### 7.4.2 Data Procedure

In line with Experiment 3A, the RT cut-off values were 250 milliseconds for the lower limit, but extended to 4000 milliseconds for the upper limit, to take into account the older age of participants. 92 trials were removed after imposing these limits. This left 5108 trials, or 98% of the available trials, from twenty participants for statistical analysis. This was again log transformed to correct for positive skew. A raw plot of the spread of reaction times, after the cut-offs were applied, is

shown below in Figure 7.5.

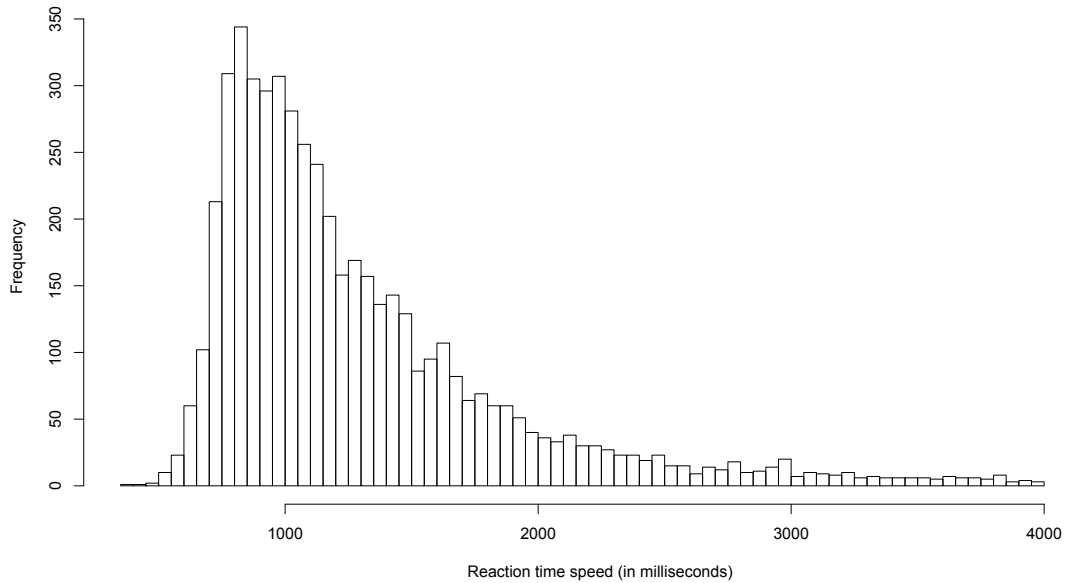


Figure 7.5: Spread of reaction times for older participants

### 7.4.3 Results

The median reaction times for older participants was considerably higher than for the young participants, at 1090 milliseconds for the non-practice trials. The median reaction time in the young voice condition was 1132ms, and for the older voice condition it was 1049 milliseconds. The median reaction time for the practice trials was 1322 milliseconds. As we age, our visual processing and visual search abilities decline along with our reflex speeds (Fozard, 1990), which would account for the difference between the younger and older participants' mean and median reaction times. The mean accuracy score for the older participants in the young voice condition was 91%, and 90% for the older voice condition. Mean accuracy in the practice trials was 93%. As with experiment 3A, significant differences in accuracy between the practice and non-practice trials was checked, and no significant differences were found ( $p=.395$ ).

A linear mixed effects model was fit to the reaction time data. As with Experiment 3A, fully fit complex models were first tested for, including the fixed

and random effects of: order of presentation, stimulus type, block and the added effect of participant sex. These were not significant in any of the iterations. The final model had the dependent measure of reaction time, the fixed effects of voice age and trial, and the random effect of participants. Trial was entered as a non-linear predictor with a 4-knot restricted cubic spline.

Table 7.2: Mixed model output - modelling predicted reaction time by voice age and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	7.347	0.043	169.6	<.001	***
Younger voice	0.067	0.012	5.38	<.001	***
Trials (1st point)	-0.002	0.000	-9.44	<.001	***
Trials (2nd point)	0.000	0.000	-0.13	.898	
Trials (3rd point)	-0.006	0.003	2.44	.014	*

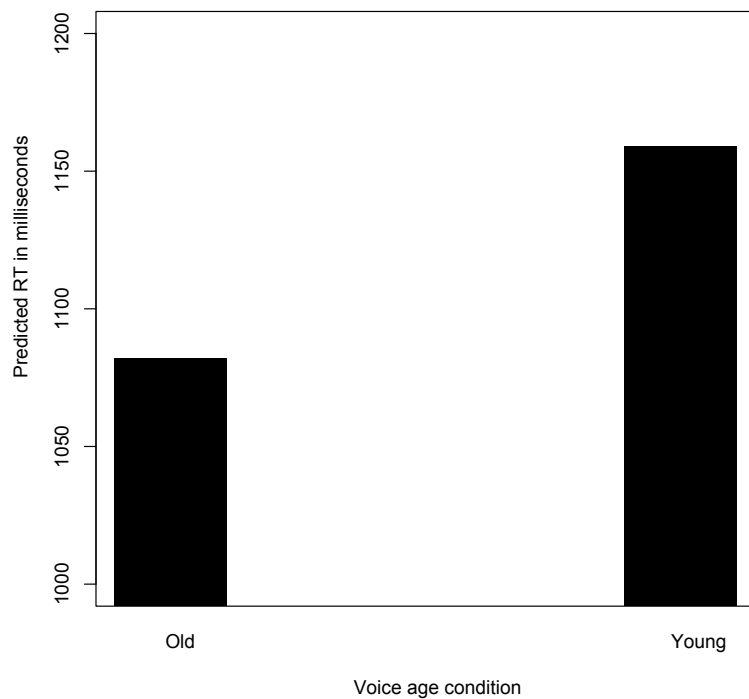


Figure 7.6: Mixed model plot: predicted RT based on speaker age

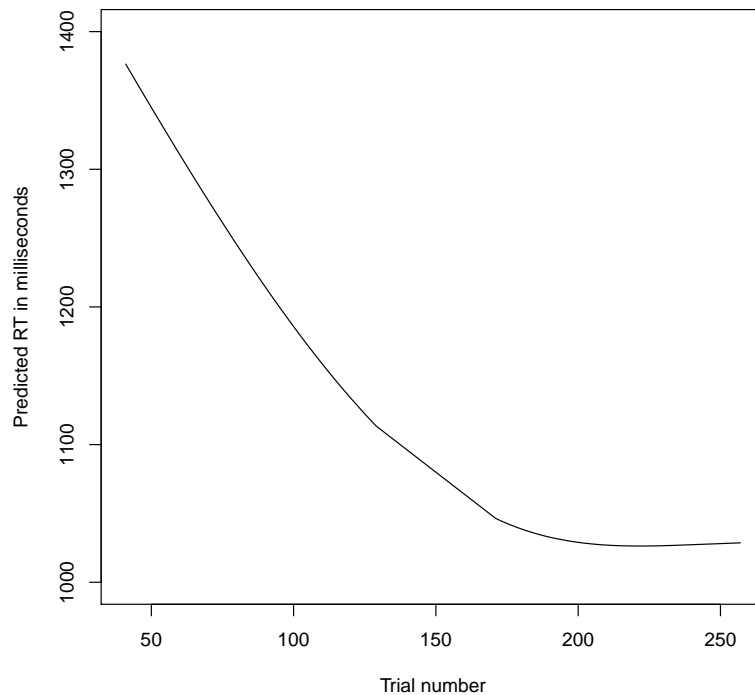


Figure 7.7: Mixed model plot: predicted RT based on trial number

The predictions of the model are shown in Table 7.2. Older participants listening to the young voice were predicted to have significantly slower reaction times than when listening to the older voice. These predictions are shown graphically in Figure 7.6, presented in unstandardised milliseconds. Again, there was also a significant prediction of quicker reaction times as they progressed through the trials (Figure 7.7).

I ran a further analysis to check whether the two groups of older participants, those listening to two voice conditions and those listening to only one voice, were affected by voice age similarly. I ran the same model as the above for the 10 participants who heard both voices, the predictions of which are presented in Table 7.3. Model comparisons indicated that having three knots for the non-linear trial predictor rather than four was optimal.

This model shows that across 260 trials, with half of the participants hearing the older voice first, the age of the voice was a significant predictor and, consistent with the overall analysis presented above, it was the young voice that was associated with slower predicted reaction times. In line with experiment 3A,

Table 7.3: Mixed model output - modelling predicted reaction time by voice age and trial for participants hearing both voices

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	7.386	0.046	159.8	<.001	***
Younger voice	0.071	0.014	5.10	<.001	***
Trials (1st point)	-0.004	0.000	-16.28	<.001	***
Trials (2nd point)	0.002	0.000	9.14	<.001	***

whether this difference existed for the first 130 trials was checked. Like experiment 3A, this returned no significant difference,  $p=.471$ . It was then checked whether the second group of participants, who heard only the younger male or older male voice, showed differences in predicted reaction times. The results for this second group of 10 participants indicated that there was a significant difference for whether the first 130 trials were presented in an older or younger voice, with those in the young only condition predicted to be slower  $p=.002$ .

Looking at the model predictions from both Experiments 3A and 3B, the results indicated that younger and older participants did not react to the voices in the same way: younger participants were, overall, predicted to respond more quickly when hearing the younger voice, and older participants the older voice. This is of course contrary to the primary hypothesis. There were variations in the strength of this effect, and which part(s) of the experiment were more susceptible to it, but the age of participants does appear to lead to different effects.

The data from Experiments 3A and 3B were combined to investigate whether the effect of the voices held across both studies, and to strengthen confidence in these separate results. Trial was entered as a 3-knot non-linear predictor. Table 7.4 shows that the age of voice is a significant factor in participants' reaction times. This is shown graphically in Figure 7.8. A log likelihood test comparing the final model to one without voice age confirmed that voice age was a significantly better model ( $p < .001$ ). Further to the effect of voice, there was a significant difference between the two groups of participants, where young participants were significantly faster than older participants ( $p=.002$ ).



Table 7.4: Mixed model output - modelling predicted reaction time by voice age and trial and the interaction of voice age and participant age

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	7.342	0.034	210.60	<.001	***
Young voice (YV)	0.076	0.011	6.54	<.001	***
Young participants (YP)	-0.366	0.010	-36.37	<.001	***
YP * YV interaction	-0.118	0.015	-7.47	<.001	***
Trials (1st point)	-0.003	0.00	-13.63	<.001	***
Trials (2nd point)	0.002	0.001	9.73	<.001	***

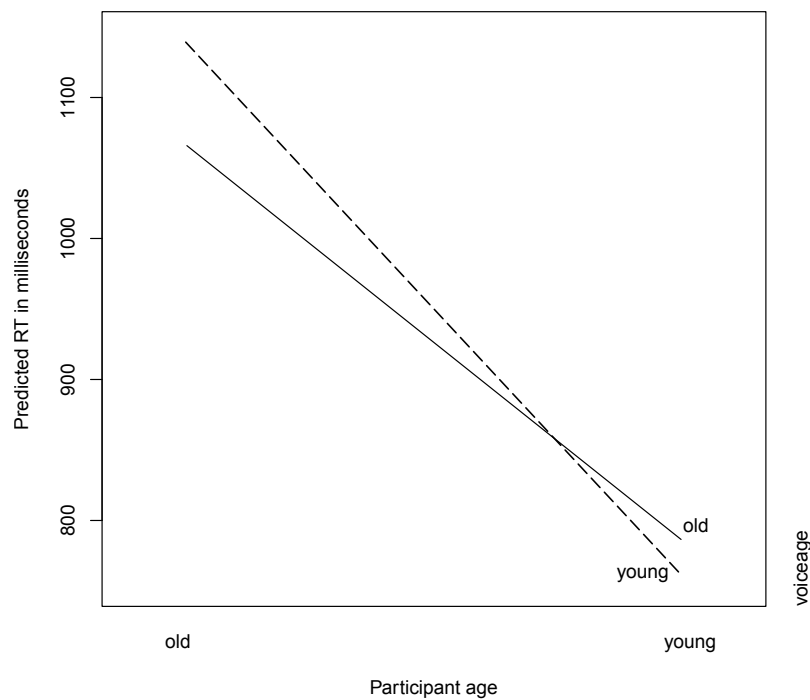


Figure 7.8: Mixed model plot: interaction of predicted reaction times by participant age and voice age

## 7.5 Experiment 3C: Silent Comparison

So far, I have presented evidence to support the prediction that participants would be sensitive to the age characteristics of a voice in a reaction time experiment, and that this would vary between younger and older participants. Contrary to the primary hypothesis, younger participants were slower when listening to an older voice, but older participants were slower when listening to a younger voice. Based on the literature reviewed, it was expected that it was the older voice, indexed to the social concept of *elderly*, that was driving these effects. This was further strengthened by the results presented later in Experiment 3D, which indicate that a processing fluency account is not compatible with these results. The design of experiments 3A and 3B, however, did not allow this to be tested directly. In the present experiment, two new groups of young and older participants completed the reaction time experiment. In this version, there were no voice stimuli. All of the instructions were text based, allowing us to treat the data from this experiment as a type of baseline. The term baseline, however, is one I use carefully and it should not be taken as synonymous with *normal behaviour* and is best thought of as just another condition. If the average reaction time for the young participants in the young voice condition is closer to their reaction times in the silent condition, then it would indicate that the older voice was indeed driving the deviations away from the average.

### 7.5.1 Method and Participants

Experiment 3C follows the same set-up for Experiments 3A and 3B with two important differences. First, there were no audio stimuli whatsoever. Secondly, because there were no voices, there was only one condition, and therefore 130 trials rather than 260 trials as in Experiments 3A and 3B. In this instantiation, participants were 10 older New Zealanders (5 male and 5 female, age range 60 to 75 years) to compare with the older participants of Experiment 3B, and 14 younger New Zealanders (9 females and 5 males, age range 18 to 27 years) to compare with those from Experiment 3A. They had no reported speech or cognitive impairments, and none were excluded due to extreme reaction times.

### 7.5.2 Data Procedure

Based on the results of Experiment 3A and 3B, it was clear that participants became quicker and less error prone in the second half of the experiment - please refer to Figures 7.4 and 7.7 to note the increase in speed. To make the results of this silent condition more comparable to the previous analyses, I look at only the first 130 trials for both Experiment 3A and 3B in the analyses that follow. The combined data were analysed separately for the older participants and for the younger participants. For the younger participants, cut-off values of 250 milliseconds for the lower bound and 3000 milliseconds for the upper bound were used. For the older participants, this was 250 milliseconds and 4000 milliseconds respectively, in line with Experiments 3A and 3B.

### 7.5.3 Results for Younger Participants

Figure 7.9 below shows the untransformed spread of reaction times for the younger participants in the silent condition after the cut-offs were applied, and it is much the same as for those younger participants listening to the voices in Experiment 3A.

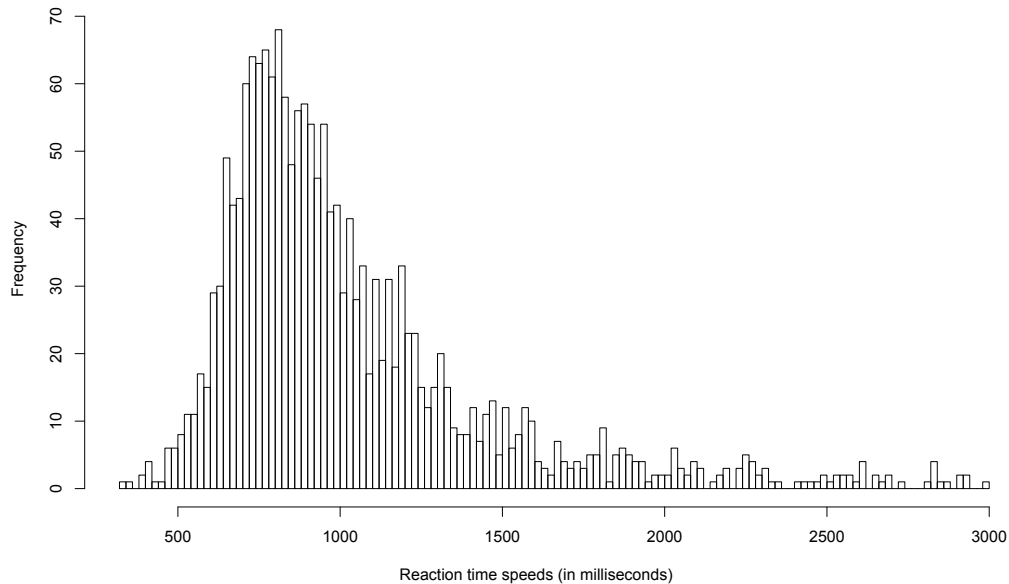


Figure 7.9: Spread of reaction times for younger participants in silent condition

1820 trials were available for analysis. After the exclusion of 22 trials below and

above cut-off values, 1798 trials (99%) remained for analysis. The median reaction time for younger participants in the silent condition was 911 milliseconds, which is slower than the median reaction time from Experiment 3A. Mean accuracy was 86%, which is also lower than the accuracy rates in Experiment 3A. I checked to see whether this difference was significant, using a linear mixed effects model with accuracy, presence of voice, and trial as the fixed effects, and participants and trial as random effects. There was a significant difference in accuracy rates between the young participants in Experiments 3A and 3C ( $p=.045$ ), with those in the silent experiment predicted to be less accurate.

Following the same procedure as the analyses for Experiments 3A and 3B, a linear mixed effects regression model was fit. Tested for were the fixed effects of audio condition (younger voice, older voice, or silence) and trial number. A simple linear predictor for trial was used rather than a non-linear restricted cubic spline component after model comparisons showed that there was no significant improvement in fit using the more complex model ( $p=.562$ ). Participants were entered as a random effect. The dependent measure was reaction time. Participant sex was tested for but was not significant as a main effect or as an interaction with audio condition.

Table 7.5: Mixed model output - modelling predicted reaction time by audio condition and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	6.944	0.036	189.86	<.001	***
Effect of older voice	-0.182	0.055	-3.31	.002	**
Effect of young voice	-0.098	0.056	-1.73	.092	.
Trials	-0.001	0.003	0.74	<.001	***

Table 7.5 shows that the older voice condition was significantly different to the silent condition, with participants in the older voice condition predicted to have faster reaction times than those in the silent condition. The older voice condition also had faster predicted reaction times than the young voice condition, although this difference was not significant ( $p=.179$ ) This is of course contrary to the results presented for Experiment 3A, but not inconsistent with the analysis for the first 130 trials of 3A that showed no difference between the younger and older voice conditions ( $p=.208$ ). The predictions of the model are shown graphically in Figure

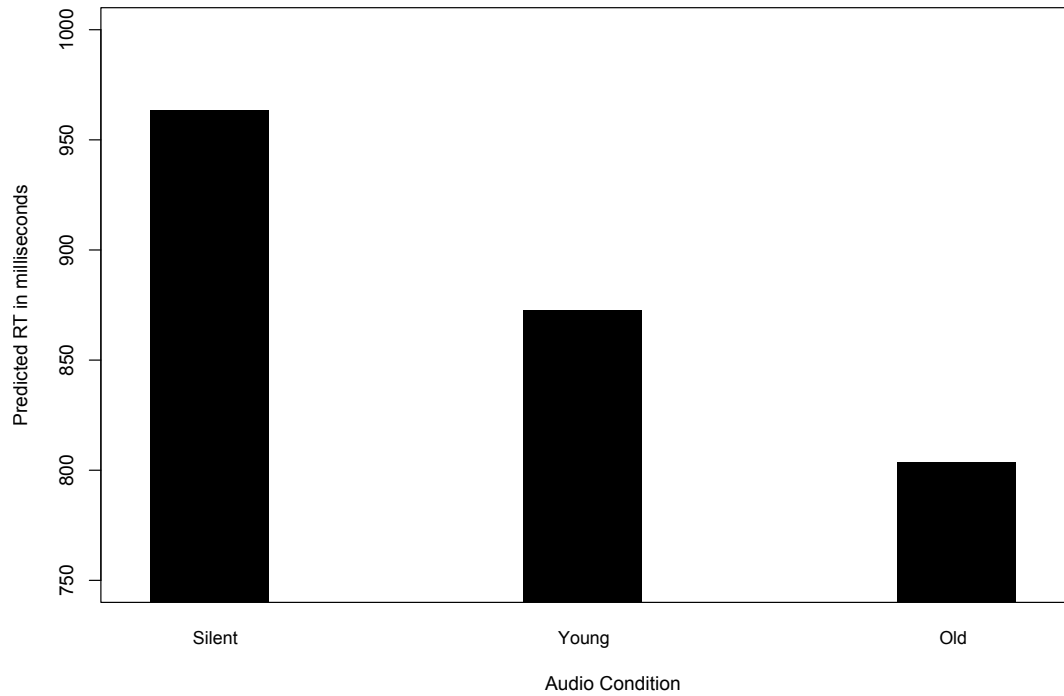


Figure 7.10: Mixed model plot: predicted reaction time based on audio stimulus

7.10. I further investigated whether the presence of a voice overall had an effect on the reaction times, as the coefficients indicated. An identical linear mixed effects model was run, replacing audio stimulus type with presence of audio, again using only the first 130 trials. The predictions from this model can be seen in Table 7.6. Participants who had no voice stimuli were predicted to have significantly slower RTs than those who heard either the older or younger voice ( $p < .003$ ).

Table 7.6: Mixed model output - modelling predicted reaction time by audio presence and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	6.637	7.114	93.31	<.001	***
No audio	1.499	4.693	3.19	.003	**
Trial	-1.020	5.5270	-18.46	<.001	***

The results of study 3C, for the younger participants, show that compared to a silent experiment, participants' predicted reaction times were significantly quicker when a voice was present. In line with Experiments 3A and 3B there was also a significantly increased probability of quicker reaction times as they progressed through the trials. The effect of the age of the voice varied by the number of trials analysed. For the first 130 trials, participants were predicted to be quicker hearing the older voice, but when all 260 trials were analysed, this effect reversed. This

indicates that this effect, whilst significant, is a fragile one that is perhaps more sensitive to changes in voices during a task than to something more stable between voices.

#### 7.5.4 Results for Older Participants

Figure 7.11 below shows the untransformed spread of reaction times for the older participants in the silent condition after the cut-offs were applied, and it is much the same as for those older participants listening to the voices in Experiment 3B.

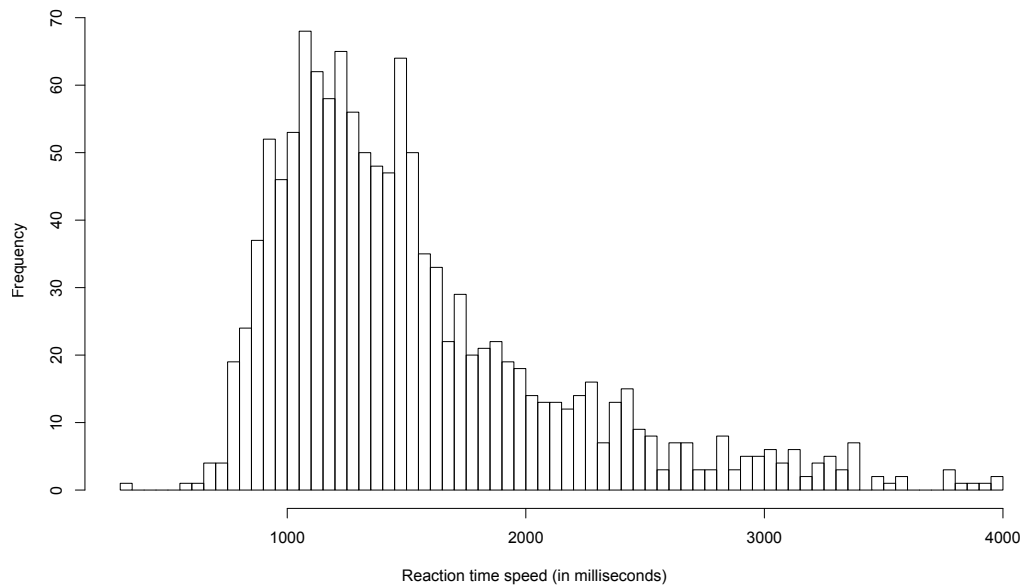


Figure 7.11: Spread of reaction times for older participants in silent condition

1300 trials were available for analysis. After the exclusion of 43 trials below and above cut-off values, 1257 trials (97%) remained for analysis. The median RT for older participants in the silent condition was 1397 milliseconds, which was slower than the mean RT from Experiment 3B. The median reaction in the practice trials was 1338 milliseconds, and mean accuracy was 91%. Following the same procedure as the for the younger participants, a linear mixed effects regression model was fit. The model has the fixed effects of audio condition (younger voice, older voice or no audio) and trial number. A simple linear predictor for trial was used rather than a non-linear restricted cubic spline component after model comparisons showed that there was no significant improvement in fit when using the more complex model ( $p=.375$ ). The silent condition is set as the comparison level here. Participant

and audio condition were entered as random effects. The dependent measure was reaction time.

Table 7.7: Mixed model output - modelling predicted reaction time by voice stimulus and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	7.426	0.060	123.34	<.001	***
Effect of older voice	-0.142	0.083	-1.70	.099	
Effect of younger voice	-0.030	0.083	-0.36	.720	
Trials	-0.002	0.000	-12.29	<.001	***

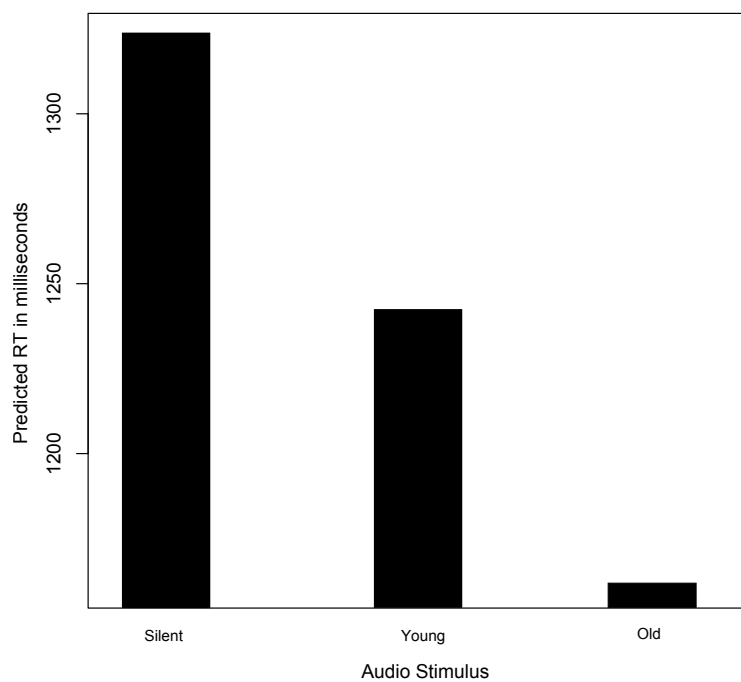


Figure 7.12: Mixed model plot: predicted reaction time by audio stimulus

The results of this analysis, presented in Table 7.7, show that for the older participants both the younger and older voices were associated with faster predicted reaction times. Whilst this did not reach the level of statistical significance, the trend was the same. This is represented graphically in Figure 7.12. The plot of these predictions indicated that the presence of audio overall may be significant. An identical model was run replacing voice condition as a fixed effect with audio presence. The model's predictions can be seen in Table 7.8. The results show that as with the younger participants, the presence of audio for the older participants was associated with faster predicted reaction times.

Table 7.8: Mixed model output - modelling predicted reaction time by voice stimulus and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	7.20	0.066	108.86	<.001	***
No audio	0.086	0.025	3.26	.002	**
Trials	-0.002	0.000	-12.29	<.001	***

## 7.6 Experiment 3D: Voice Familiarity

Both Experiments 3A and 3B used voice age to test stereotypic priming. It was possible, however, that processing fluency contributed to the effects. Processing fluency has been described as “the sense of familiarity evoked by presented materials, or the subjective ease with which those materials are perceived” (Bjork, 1999, p 343). Multiple experiments have now been conducted by different researchers working on this, a good overview of which can be found in Schwarz (2004). Reber et al. (1998) showed that easier processing conditions, manipulated by prime-target congruence, could influence liking for the target. Reber and Schwarz (1999) have further demonstrated that participants may infer truthfulness from an easier to process stimulus, such that “Osorno is in Chile” is seen as more truthful when presented in an easier to read font colour and background. People appear to misattribute their feelings of familiarity, or fluency, for feelings of liking, truth, or persuasiveness. Further to this, the less familiar or fluent something is, the more processing effort is required (Jacoby & Dallas, 1981). I reasoned that if voice familiarity played a role in reaction times, then New Zealand participants hearing a New Zealand voice would be faster than when hearing a Scottish voice. Further to this, it would also help clarify the extent to which ingroup/outgroup factors could impact this task.

### 7.6.1 Method and Participants

Experiment 3D follows the same set up for both Experiments 3A and 3B, but with two new voices: one young Scottish male, and one young New Zealand male. The Scottish male spoke Standard Scottish English, and the New Zealand male a standard New Zealand accent. Participants were 31 young to middle aged New Zealanders, composed of 21 females and 10 males. Their mean age was 26



years. They were recruited from local areas, as I was keen to rely not only on the undergraduate population at the University of Canterbury. Further, as age was not the manipulated factor in the voices here, I was able to include a wider range of ages than with Experiments 3A and 3B, from 18 to 55. Occupations were diverse, including prison officer, woodworker, and teacher. I did this in the knowledge that using different groups introduces concerns about strict comparisons of the data, but with the view that the increased ability to generalise from the results to a wider population made this worthwhile.

In line with the young participants from Experiment 3A, the reaction time cut-off values were 250 milliseconds for the lower limit, and 3000 milliseconds for the upper limit. 44 trials were removed after imposing these cut-off values, leaving 99% of available trials for analysis. The median reaction time was 792 milliseconds, much closer to the younger participants in Experiment 3A than the older participants of 3B. The median reaction time during the Scottish voice condition was 784 milliseconds, and 802 milliseconds for the Kiwi voice condition. The mean accuracy score was 93%. Figure 7.13 shows the untransformed spread of reaction times after the cut-offs were imposed, showing a similar distribution and skew to the younger participants in the previous instantiations of the experiment (see Figure 7.2 and Figure 7.9).

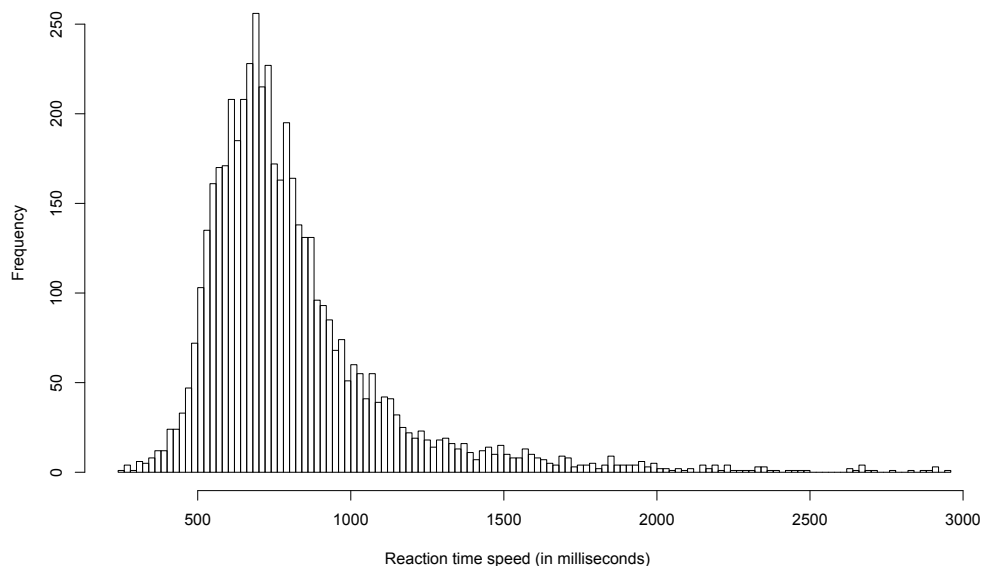


Figure 7.13: Spread of reaction times for younger participants in Experiment 3D

## 7.6.2 Results

Whilst the median reaction times above strongly indicated that any difference between conditions was weak at best, the same data analysis procedure was followed as used for the previous analyses. A linear mixed effects regression model was again fit to the reaction time data. As with Experiments 3A, 3B, and 3C, fully fit complex models were first tested for, including the fixed and random effects of: order of presentation, stimulus type, and block. These were not significant in any of the iterations. The final model has the dependent measure of reaction time, the fixed effect of trial number, the fixed effect of voice, and the random effect of individual participants. Trial was allowed to be non-linear using a 4-knot restricted cubic spline.

Table 7.9: Mixed model output - modelling predicted reaction time by voice condition and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	6.720	0.026	252.6	<.001	***
New Zealand voice	0.004	0.005	0.83	.410	
Trials (1st point)	0.001	0.001	8.57	.002	**
Trials (2nd point)	-0.009	0.005	-15.93	.043	*
Trials (3rd point)	0.029	0.001	17.33	.010	*

The predictions of the model are shown in Table 7.9. The results confirmed that there was no statistical difference in the predicted reaction times for participants in the two voice conditions (where NZ voice is compared against the Scottish voice, which is set as the comparison level). The predictions are shown graphically in Figure 7.14.

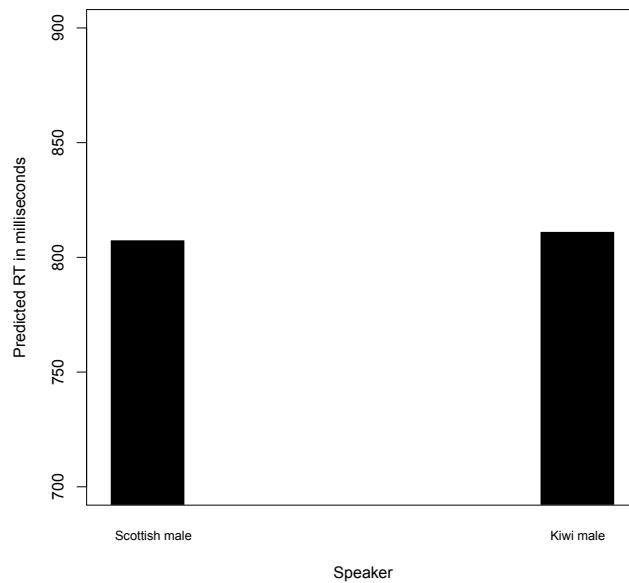


Figure 7.14: Mixed model plot: predicted reaction time by voice condition

### 7.6.3 Presence of Voice Overall

In this final analysis, the silent Experiment 3C data was combined with the data from Experiments 3A and 3D (those with the young participants), to look at whether more generally, the presence of a pre-recorded person speaking has an impact on task performance. To be consistent, only the first 130 trials of both Experiment 3A and 3B were chosen. Audio presence was entered as a fixed effect along with trial number. A simple linear predictor for trial was used rather than a non-linear restricted cubic spline component after model comparisons showed that there was no significant improvement in fit using the more complex model ( $p=.364$ ). Participant was entered as a random effect. The predictions of this model are shown in Table 7.10. The silent condition was entered as the comparison level. Looking at the results we can see that the presence of audio, which includes the young, old, Scottish and Kiwi voices, led to faster predicted reaction times overall. The strength of this speeding effect is a product of the age of the voice and, from what was presented in the results of Experiment 3B, the age of the participants. The model predictions are shown graphically in Figure 7.15. Accuracy did not differ significantly between audio and non-audio conditions in this final analysis ( $p=.249$ ).

Table 7.10: Mixed model output - modelling predicted reaction time by voice condition and trial

	Estimate	S.E.	<i>t</i> value	<i>p</i> value	
(Intercept)	6.722	7.101	94.67	<.001	***
No audio	1.905	4.881	3.90	<.001	***
Trial	-8.556	9.009	-9.50	<.001	***

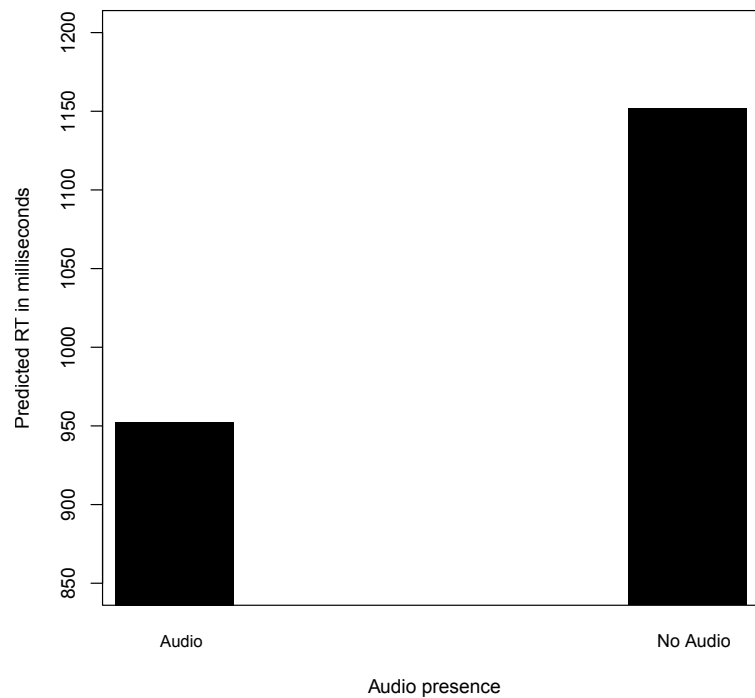


Figure 7.15: Mixed model plot: predicted reaction time based on audio stimulus

## 7.7 Discussion

I hope to have shown in the experiments detailed in this chapter that the presence of a voice, and the age of that voice, can act upon participants' reaction times. Experiment 3A and 3B indicated that younger participants were slower when they heard an older voice, and older participants were slower when they heard a young voice. Experiment 3C indicated that participants' reaction times in a silent experiment were significantly slower than both the older and younger voice conditions, and for both older and younger participants. Experiment 3D provided some preliminary support that voice familiarity, along with simple ingroup/outgroup factors, did not appear to strongly influence these results. A combined analysis of Experiments 3A, 3B and 3C indicated that it is the older voice that drives the effects for both young and older participants. The older voice drives both younger and older participants further away from the silent 'baseline', whilst the magnitude of these effects differs for the two voice conditions. It is important to note, again, that the silent condition is not a valid baseline as such, because the experiments, and the hypotheses, were designed to assess the strength of voice based priming and so any true baseline would necessarily have voice included. It is, however, a useful avenue to explore in future experiments.

In line with other studies on stereotypic priming, such as those discussed in the introduction, we can see an effect of socially marked stimuli on the proceeding behaviour of participants. In their study, Kawakami et al. (2002) found that being primed with *elderly* had the effect of slowing down the reaction time of participants in a lexical decision task. The results presented in Experiment 3A offer some support for this. It is still too early to say whether this was as a result of stereotype activation, and as with Experiment 1, the affect of the voice could mean that it is not priming at all, but something rather more immediate. This needs further exploration, but it is a promising start. Förster et al. (2009), in their chapter on priming, write that it is now generally accepted that exposure to social traits and characteristics, and stereotypes which contain these traits, can lead to behavioural assimilation to these traits, and that this has been shown abundantly. Past studies have used stimuli such as words, images, sounds, music, and so forth, and found that such stimuli produce observable behavioural effects. Sociolinguistic research has shown that aside from the content of speech, both lexical and communicative, speech also carries social meaning and can affect linguistic behaviour.

### 7.7.1 Stereotype Activation

This study was directly influenced by the research on stereotypic priming and automatic behaviour. There were two competing hypotheses. First, and most likely given the prior research in this area, was that the elderly voice would slow participants, regardless of their age group. The weaker hypothesis was that it would do so for the young participants, but that there may be contrast effects for the older participants who attempt to inhibit the negative associations of the elderly stereotype. These results more clearly fit the latter prediction, but this interpretation is not without its own problems. Inhibition would require some deliberation on the part of the elderly participants, yet this task was a cognitively loading one, which should inhibit such effortful processing. This could again point to the effect being affect related, which is more primary than a mediated cognition explanation along the lines of stereotype inhibition.

Wheeler and Petty (2001) conducted a meta-analysis of the literature on stereotype activation and its behavioural effects, and wrote that the ideomotor theory predicts strong self-stereotype effects would occur because there are two pathways: ideomotor automaticity, and also stereotype threat (see Steele & Aronson, 1995). I did not find that elderly participants behaved more slowly to the elderly voice (ideomotor account), which in turn also suggests that they did not activate a negative self-stereotype or relevant behavioural script. That this did not occur does not, however, discount a stereotype priming effect and indeed these results are consistent with those found by Hess et al. (2004).

Stereotypes are multifaceted, situation dependent, and culturally dependent. Levy and Langer (1994) and Levy (1996) have documented how the elderly can overcome negative self-stereotypes by being exposed to positive age-related primes, which we might presume a voice acting as proxy for a psychology experimenter might be. For the younger participants, who do not have a self-stereotype of ageing, the prime may have behaved quite differently. Moskowitz and Skurnik (1999) have noted that “Exemplar primes are narrow and exclusive and, as discussed by Heider(1944), can serve as a context or ground in which behaviour is viewed. For these reasons, exemplars are likely to be used as standards of

comparison and to produce contrast through a perceptual process” (Moskowitz & Skurnik, 1999, p 916). If the older voice was seen as atypical of the elderly, then this is worth considering. The results are consistent too with Harwood’s (1996) finding that negative age stereotypes are weaker for middle aged and older people than for younger people. Allen and Sherman (2011) noted that when one’s ego is threatened, one may be motivated to inhibit the negative traits of the self-stereotype or to foreground the negative associations of out-group stereotypes. For the elderly, such a negative stereotype exists and within that stereotype there are associations about speed and brainpower. Both elderly and younger people share this stereotype, but the elderly can be motivated to respond to this stereotype differently by this process of ego-protection. For the Scottish and Kiwi voices, there are no particularly negative stereotypes, and certainly not with the type of slowed performance and physicality of the elderly versus young stereotype.

The slower reaction times that the young participants showed when listening to the older voice compared to the younger voice in Experiment 3A need not be because of negative stereotyping. It may have brought forth concepts of mental and physical slowness - or, on the other hand, caution, authority, and wisdom. When listening to the younger voice, it could have been agility and vigour, or carefree and whimsical. The duality of stereotype valence is important to consider (see Brewer & Lui, 1984; Levy, 1996). It is important also to consider that the older voice speeded participants compared to a silent version of the experiment, not that it does so more generally. In for example Bargh et al’s 1996 study, they did not compare their participants in the *old* word puzzle condition to those in a no puzzle condition, but rather those in a neutral puzzle condition. There was no baseline per se. The same holds for this experiment. The older voice condition is best compared against the young voice condition. From there, the presence of audio can then be compared to the the absence of audio.

Chasteen, Schwarz, and Park (2002) conducted a lexical decision experiment designed explicitly to test how older and younger participants responded to positively and negatively valenced age stereotypes. They found that both younger and older adults shared the same stereotypes of younger and older people, but that there was very weak evidence of a young stereotype having been activated. Their findings would, then, predict that in the current paradigm, both groups

would primarily activate the elderly stereotype, which would be consistent with the results presented showing that the older voice condition differed most from the silent conditions. It would be with some difficulty to uncover what precisely is activated, such as wise or cautious or calming or authoritarian. Investigating this might make use of implicit association experiments, where participants would hear the voice and respond to a series of stereotype related labels and images. A problem arises here, however, that if they do this before the task, the priming effects would almost certainly drop because attention has been brought to bear on the voices. If it was done after the task, the recent exposure to the voices could influence the results of this IAT (see Dasgupta & Greenwald, 2001).

To answer what social information in particular was activated by the voice -if indeed it was social information and not a more immediate emotion - is a difficult question that will require further experiments. Human voice is, more so than a picture of zimmer frame, or the word *Alzheimers* flashed on screen, a particularly complex stimuli. Moreover, it is one that can be categorized in multiple ways. Because of this categorization ambiguity, a competition for dominance ensues, the winner of which is better specified at an individual rather than group level. As Macrae and Bodenhausen wrote, "Category salience, chronic accessibility and temporary goal states are all factors that are likely to confer an activational advantage to particular categories in such a competition" (Macrae & Bodenhausen, 2001, p 250). Without knowing the category salience of Asians, or how chronically accessible Asians are, or whether people have a temporary goal state relating to Asians, the particulars of how they use that voice are hidden. Further to categorizing, there are perceptual challenges with voices too. Whilst a photograph of a zimmer frame could certainly produce different affective reactions in people, people would see it as a zimmer frame all the same. This is not so with voices. What sounds like an older white man to one person might sound like a younger black man to another, which is particularly problematic for speaker identification in legal situations (Baugh, 2003; Nolan, 2005). I was careful to ensure that the voices were perceived as members of the category they belonged, but this was a directed task, with an absence of background noise and social context. Even so, ratings were not always 100% accurate and presumably influenced by participants' prior knowledge of the probability of a person being from one category rather than another. That is to say that, if you have a feeling that there are only 20 gay men



for every 200 heterosexual men, your rating of whether a voice sounds gay should be influenced by this knowledge - the voice stimuli is just part of the rating. If you imagine how voices are heard in a day-to-day context, what could influence their perception could be as disparate as their hairstyle to the traffic noise. What we hear is often literally in the ears of the listener. In day to day life, we may activate one, all, of none of a particular stereotype. Whether people are accessing multiple stored exemplars of the category, that is to say every time they have heard a similar voice in the past, and all of the social and contextual information associated with those encounters, or whether they access a more abstract schema of elderly/youth, is an open question, but one worthy of future endeavour.

### 7.7.2 Social Facilitation

The results presented above have shown that the presence of a voice improved task performance (in the sense that being faster is an improvement, given that accuracy didn't suffer). This happened regardless of the age or accent of the voice speaking over the experiment. If we consider that the voices were not merely ambient, but actively directing task behaviour and giving messages of encouragement, they were akin to having someone present. Research on social facilitation has shown that the physical or implied presence of others can improve task performance. In one earlier meta-study, Bond and Titus (1983) concluded that the presence of others can have the effect of improving task performance, particularly when that task is one where the dominant tendency is to give the correct answer, which is referred to as a simple task in the literature (Zajonc, 1965; Huguet, Galvaing, Monteil, & Dumas, 1999). Huguet et al. (1999) replicated the earlier findings on social facilitation effects on simple tasks, but found that social presence also facilitated stroop task performance. In their study, they noted that the present 'other' should be seen as a 'slightly superior coactor'. "How people experience their social world, or what they come to believe about it, can play a significant role in determining primitive operations of cognition" (Huguet et al., 1999, p 1024). Similar findings have been replicated more recently by Sharma, Booth, Brown, and Huguet (2010). These effects are not, however, limited to physically present humans. Similar social facilitation effects have also been reported with robot presence (Riether, Hegel, Wrede, & Horstmann, 2012) and even implied social presence (Risko & Kingstone, 2011). It appears that the presence of voices led to social facilitation effects in the

reaction time experiments described above. The audio voices in the experiments played the role of experimenter, and as such, they would almost certainly qualify as ‘slightly superior coactors’ in the estimation of many of the participants. Further, the task was nearer the simpler end of the scale with accuracy scores across the experiments varying a little around ninety percent.

### 7.7.3 Processing Fluency and Familiarity

Preliminary evidence was provided indicating that voice familiarity plays little part in these results. This does not discount perceptual fluency. It seems more likely that the voices simply did not differ in fluency to any magnitude. This indicates that familiarity, and fluency, must be treated as separate entities, as the New Zealand voice would be much more familiar to these participants than the Scottish one. When people have repeated exposure to the same stimulus, they can often report liking that stimulus more (see for example Reber & Schwarz, 1999). Following from that, the less familiar or ‘fluent’ something is, the more processing effort is required which in turn leads to perceived dislike (see for example Jacoby & Dallas, 1981). This is true both of recently encountered items, and also historic knowledge of things, such as *robin* being more ‘fluent’ as a bird than *penguin* (see Oppenheimer & Frank, 2008). The results of Experiment 3 showed that New Zealand participants exposed to a native and non-native accent had nearly identical reaction times in both conditions. I am also unaware of any work that shows that processing fluency effects continue after the presentation of the stimuli. To test this properly, it should first be established that certain voices do indeed cause different effects using established paradigms, such as reading aloud an exercise regime and then asking participants to say how easy it would be to undertake these exercises, and whether they would do it themselves (Song & Schwarz, 2008). After it had been established that two voices showed distinct task differences, these voices could then be used for the reaction time paradigm as described above.

## 7.8 Experiment 3 Conclusion

This experiment demonstrated that two different voices, one younger and one older, impacted the reaction times of younger and older participants in significantly different ways. Younger participants were faster in a simple cognitive task if they had recently listened to a younger voice. Older participants were faster if they had

recently listened to an older voice. Secondly, it was demonstrated that having two accents that vary not by age, but by familiarity, did not appear to result in similar effects, offering some initial evidence against a *simple* perceptual fluency account. These effects are a result of voice presence and an interaction between the age of the voice, and the age of the participants. Participant groups in Experiment 3A and 3B appear to have been influenced more by the older voice, possibly by activating elderly congruent traits. A challenge going forward would be to investigate whether it is indeed a social priming effect, or something more like Zajonc's primary affect, the 'feeling before thinking' (Zajonc, 1980), that is responsible for the different voice effects. If younger participants simply like younger voices, and vice-versa for the elderly participants, then it is not social priming as is traditionally conceived.

A further conclusion is that much like a physically present other, or a robot, or even an implied social presence, having a voice during a task such as this appears to significantly facilitate task performance. All of the participants, whether listening to a younger, older, Scottish or New Zealand voice were faster than those participants who heard no voice during this task.

## Part III

# Discussion and Conclusion

## Chapter 8

# General Discussion

There were two overarching goals for this thesis. The first of these was to show that voice is a relevant environmental object that can affect behaviour in non-linguistic ways. In the introduction I stated my belief that actions can be traced to causes. In the experiments presented in this thesis, I hope to have identified voices as a candidate for causal stimuli. Each experiment had its own smaller discussion section detailing the specific effects of the voices and possible causes for the priming behaviour. This larger discussion section draws together the component experiments, and different strands of research, to try to outline commonalities, unanswered questions, and possible future directions for research of this kind.

It might have been more surprising had voice not produced any observable behavioural effects. It is a stimuli rich with social meaning, linking both speaker and listener to their social world. The effects, however, occurred under specific conditions - they were not global. As was described for each of the individual experiments, there are examples of the voices affecting participants in different conditions differently, and examples of voices having little or no effect. It is important to remember that the effects described in this thesis are directly task related. There is no suggestion that they would occur when there is no correspondence between the stimulus and the environment. Bargh et al. (1996) noted that their participants did not go and “buy condos in Florida” after being presented with the elderly stereotype. There is similarly no presumption that my participants booked flights to Beijing or went home and watched *Braveheart* after hearing the Asian or Scottish voices respectively. The results were not always neat, but participants in different voice conditions, or in silence, differ frequently enough

for me to be confident that people can use social information from voices to guide task behaviour. The second goal was to begin to posit what it is about certain voices that cause these effects: what social-psychological constructs are they connected to.

One commonality which seems to act as a criterion for these priming effects is *situational relevance*. This is one of the key propositions Schwarz et al. (2003) offered as a foundation of social cognition research. They proposed that accessible information guides behaviour by influencing how the actors interpret their situation. This is to say that, at the encoding stage, particular stimuli can colour the situation so that with different primes, the same situation may be interpreted differently. In Experiment 1, where different people heard the same statements in different voices, the situation should be interpreted differently. An Asian male asking questions about ethnicity, or a gay man asking questions about same sex marriage, could easily be construed as task relevant. This possibility is highlighted by a comment made by one of the participants on the older male's voice. They guessed that he was homosexual '...purely because it's logical to assume a survey about perception of homosexuality (among other things) would be prepared by someone homosexual'. Offensive as it is to suggest that only homosexuals could be interested in matters of sexual equality, it belies, I think, a more widespread belief that minority topics are the preserve of minority groups. How individual participants in my experiment construed relevance is something that would require further research. In Experiment 2, where an Asian voice asks people to judge an excerpt of Cantonese, the relevance is again clear. The voice is relevant not only to interpreting the stimulus, but to the situation as a whole. The construal of a situation by accessible constructs is central to the work of Sedikides and Skowronski (1991), and is succinctly described by Fazio, Sanbonmatsu, Powell, and Kardes (1986) who wrote that "behaviour in any given situation is largely a function of the individual's definition of the event that is occurring" (Fazio et al., 1986, p 237). In Experiment 3, the relevant criterion was not the relatedness of the voice to the task, but the rich stereotype and associative network that accompanies an elderly sounding voice.

The results presented here do not support a direct perception-behaviour link, which makes a cognitively mediated account the more likely option, with the caveat that the affect and valence of the voices might have been partly or solely responsible

for some of the effects. Schwarz has written that “We need to know if X is used in constructing a mental representation of the object of judgement, or of the standard against which the object is evaluated. The former use results in assimilation effects, but the latter results in contrast effects” (Schwarz et al., 2003, p 53). I would argue that for participants with a low stereotype relevance, they judge the object in congruence with the accessible construct. For those with personal relevance, they judge the object against that construct. This is a view further supported by Rothman and Schwarz (1998), who have shown that personal relevance to a topic promotes participants to rely on historic instances of behaviours to form current judgements, and that low personal relevance promotes heuristic judgements, such as subjective availability. In Experiment 1A, a matching of voice condition and social theme was expected, such that questions about immigrants would be most affected by hearing an Asian voice, or that questions about gay men most influenced by the gay voice. This did not happen. We cannot however discount relevance as a factor here. New Zealand participants are likely to have had more discussions about gay rights, or women’s rights, or the elderly, with members of their own community - that is, younger New Zealand males and females. If they have already-stored knowledge of these discussions, and which groups are likely to have certain opinions, then it would be the male and female voices of younger New Zealanders that would be relevant for them, as if they were answering to their peers. Perhaps the distance of their own social category to the Asian, gay, and older male voices was more important than category matching. Again, this is something further exploration would help shed light on.

## 8.1 Speech Variation as a Tool

Introduced in Section 4.5 was the idea that people might vary their speech (tone of voice, pace, intonation, and breathiness for examples) not only to make themselves better understood, or more liked, or because of some automatic perceptual-production loop, but because they have learned to associate certain voices with certain outcomes. It was then posited in Chapter 4 that, if this were true, current models of language variation and change would need updated, as they presently make no reference to automatic priming as a driver of change. Experiment 1 indicated that voices can certainly impact how people respond to socially targeted questions. The variation between the voice conditions must come from participants’

past experiences with related voices, and the groups from which those speakers belong. Whether this is a product of a spreading-activation of social information, or affect based, is as yet unanswered. As I have argued, the results do not seem to be related to more automatic mechanisms such as perceptual fluency (Reber et al., 1998), or the perception-behaviour express way (Dijksterhuis & Bargh, 2001). In Experiment 2, it was similarly argued that the listeners' history with certain voices is used, such that the voice anchors the search for an answer around a particular value or concept. The Asian male voice encouraged participants to rate the music as Asian not because of something universal and internal, but because each of those participants would have had some history with Asian speakers, and have associated knowledge about them. The same is true of the voices in Experiment 3. The result of this means that behavioural outcomes may feature in individual language change. This has been tested for perceptually in this thesis. The next step must be from the speech production angle. If it can be shown that people will vary their speech based on a history of certain outcomes being associated with certain speech styles, this would inform sociolinguistic and sociophonetic research, which to my knowledge, has overlooked this possibility.

## 8.2 Social Behaviour

Fiske and Linville (1980), writing about the difficulty of finely specifying models of social behaviour in the same way as cognitive models, stated that "Social processes involve affect, self-reference, unavoidable complexity, considerable ambiguity, and stimuli that are themselves causal agents" (Fiske & Linville, 1980, p 549). The results presented here are no exception to this complexity, and the stimuli used are indeed products of causal agents. With the added complexity of sociophonetic variation, the results pose many more questions than is currently possible to answer. I wish also to make clear that Ulric Neisser's call for better ecological validity was not ignored. A criticism often levelled at laboratory based experiments, particularly those that rely on what are often termed 'artificial' paradigms, is that they do not reflect, let alone predict, what might happen in the real world. Nonsense words, subliminally flashed stimuli, categorization tasks and memory tasks have all come in for criticism in this vein.

In designing the experiments that form this thesis, there was thus a tension



between controlled (yet undeniably limited) prediction, and usable results. In the first two experiments, there is certainly some real-world validity to speak of. Listening to meaningful questions in isolation, and answering them, is a behaviour that we can find in many real life situations. I have used call centres as an example before, and for offering an example of ecological relevance, I can do no better than the call centre. A helpline provides a good scenario to consider this. An employee sits in their chair, hour after hour, and has to answer a wide and often bewildering array of questions from the customer. If one can show, experimentally, that certain questions asked in a female voice have the tendency to be agreed with, either because listeners emotionally prefer the voice, or associate certain female characteristics with it, then this has direct consequences for the business. Another example is a person at home doing their ironing whilst the television booms in the background. An advert for a new brand of shampoo is playing, in which the disembodied voice asks “Are you ready for your hair to feel like a million dollars?”. I hope to have shown that certain voices may be more suited to this task than others. Finally, when you are contacted by a polling company representative, who asks “Do you agree with same sex marriage”, then I strongly argue that, over a number of instances, differences in the answers will be stratified by who asks the question.

The reaction time experiments are a different type of task altogether, one which people might be tempted to label artificial, or lacking in ecological realism. This is a misunderstanding of what it means to be ecologically valid. To answer this, we can defer to Neisser’s own words in his original call for enhanced ecological validity. He wrote:

A psychology that cannot interpret ordinary experience is ignoring almost the whole range of its natural subject matter. It may hope to emerge from the laboratory some day with a new array of important ideas, but that outcome is unlikely unless it is already working with principles whose applicability to natural situations can be foreseen (Neisser, 1976, p 4).

We may interpret this to mean two things. First, ordinary experience should guide our investigations. Second, whether it take place in the laboratory or not, the variables should be, as far as possible, real-world in nature. Over 30 years later, Mortensen and Cialdini (2010) make the same call to social psychologists as

Neisser had to cognitive psychologists. They argued for naturalistic observation, and the oft repeated but seldom listened to idea that not finding an effect where one might expect is a theoretically useful avenue for exploration. Again, to use their words, “Although many interesting and important social psychological phenomena have been found by focusing on where the light is best (i.e., illuminated by past theory, literature, and highly controlled laboratory research), this is not always where we should be looking” (Mortensen & Cialdini, 2010, p 53). There are two considerations here. First, that laboratory based effects serve a utility. Second, that the balance should be evened by real world observation. If we return to thinking about the reaction time experiments, I believe that they satisfy at least half of what Neisser (1976) and Mortensen and Cialdini (2010) have argued for. It is a mistake, not to mention somewhat naive, to label a computerised, monotonous laboratory task as ‘artificial’. A research assistant involved in transcription, or a factory worker whose sole responsibility it was to press a lever at certain times would be inclined to disagree. The whole premise of scientific management, which has remained in many industries, was to monotonize tasks. To label similar laboratory tasks artificial, or to claim they have no real world validity, is to ignore the wide number of real life tasks that are equally monotonous - like ironing.

I turn from the paradigm itself to the stimuli. Real voices were used in this thesis, unaltered, saying real, meaningful task related sentences to people involved in a task. The only manipulation to these voices was in asking the speakers to follow, as much as possible, the pace and style of the first person to record, which I accept puts limitations on claims to naturalness but was necessary as a control for perceptual fluency amongst other considerations. The criticisms of laboratory priming have typically been directed at nonsense words (or other nonsense stimuli), and for this reason such techniques were not used in this thesis.

### 8.3 Limitations

Clearly one of the main unanswered questions here is whether or not voices are acting as primes, in the traditional sense of the word and activating associated information, or whether they have a more immediate, affective impact on listeners. The results presented in this thesis show examples of category assimilation, such

as younger participants slowing in response to elderly voices, but also of contrast, and at other points of something quite unpredicted. The working hypothesis that listeners would in some way assimilate their behaviour to the social category of the voice was not reliably documented. It is likely, given the literature on affect and valance (see e.g., Sinclair, Moore, Mark, Soldat, & Lavis, 2010; Winkielman, Knutson, Paulus, & Trujillo, 2007; Schwarz et al., 2003; Lerner, 2000), that certain voices in these experiments were more pleasing to listen to, or produced an uneasy feeling in listeners, and that it is these states that influenced task performance and not social information as predicted through a spreading activation account. However, it must be considered that affect and cognition cannot be neatly separated, as Zajonc (1980) was careful to note. One could hear an elderly voice and think about his or her grandmother (the spreading activation account), and then feel quite sad because grandma is a nasty old lady (an affect/valance account), which then leads to he or she performing badly on a task (which would then be called automaticity). Whether we call this a direct, or mediated effect, is a complicated consideration.//

This alternative hypothesis based on affect does not neatly fit with the ambiguous prime results in Experiment 2, but as noted in the previous section, there is also no requirement that each set of results is explained by the same mechanism. A crucial next step for this research is to more narrowly specify a hypothesis, and through thorough pre-testing of the voices along a number of dimensions including affect, credibility, dominance and persuasion and such like, investigate whether a spreading-activation or affective mechanism is the more likely candidate, or whether they are inextricably linked in some cases.

This thesis, and much of the literature that was reviewed and that it is hoped in some way contributes back into, is concerned in one way or another with human behaviour. Indeed the reasons for doing this PhD, and the impetus behind the present experiments, stem mostly from a curiosity about what feeds into human behaviour. However, the participants that were used, like those participants of very much of behavioural science research, were predominately college aged students. To borrow from Henrich et al. (2010), they were WEIRD: Western, educated, industrialised, rich and democratic. This is hardly indicative of the global human population at large. The authors note that of six top-tier journals, 96% the subjects in their papers came from Western, industrialized

countries. Further, in one of the top journals for social psychology (*The Journal of Personality and Social Psychology*), 80% of subjects from non-US countries were psychology undergraduates. This would be less of a problem if we thought that human behaviours were largely similar across place and culture, but we know this to be untrue.

In his book “*A Geography of Thought*”, Nisbett (2003) lays out a convincing case for taking cultural variation seriously. The book focusses mostly on the differences in East Asian versus Western thoughts and behaviours, such as a study by Han, Leichtman, and Wang (1998) that investigated how children report on their daily events. The authors found that children report about themselves more than they do about others. American children, however, referred to themselves three times more than comparable Chinese children, and referred twice as much to internal states. These cultural variations extend to smell and taste perception (Han et al., 1998) and also to economic proposals such as those of the ‘ultimatum game’, in which many researchers have found that people would rather forego a small amount of money than be treated unfairly (Oosterbeek, Sloof, & van de Kuilen, 2004). These findings, and others like them, remind us that many behaviours are temporally and geographically specific, not traits of mankind as might previously have been suggested, such as by Nowak, Page and Sigmund (2000).

The experiments in this thesis predominantly utilized undergraduate students in a Western, industrialized and democratic society. Where possible, non-typical populations were also used, such as the elderly participants in Experiment 3B, and people from diverse occupations and ages in Experiment 3D. The fact remains, however, that due to the nature of these experiments, it should be clear that these results are culturally, demographically and temporally locked. For this reason, we should look forward to future replications with more atypical populations.

It would be remiss for a thesis on priming not to address the controversy surrounding priming research. This is a controversy that spilled over outside of psychology’s walls, with articles about failed replication attempts in *Nature* (Yong, 2012a) along with multiple blog and news posts that a cursory search of Google will make clear. Much of this stems from the Doyen et al. (2012) paper that failed to replicate the Bargh et al. (1996) age priming study, and latterly

the Shanks et al. (2013) paper that found much weaker evidence for intelligence priming than Dijksterhuis and van Knippenberg (1998) presented. Indeed, an email was circulated to psychologists by Daniel Kahneman that, while sympathetic to the premise of priming, reprimanded researchers for not being more rigorous about replication and can be seen on Ed Yong's post on the matter in *Nature* (Yong, 2012b). Cesario, in a recently published article, has addressed many of the current concerns with priming research (Cesario, 2014). Relevant to the present work is Cesario's argument that the nature of priming, and its relative infancy, make applying the same rules and standards that other theories and replications are subjected to a difficult task. He writes that psychologists "should expect priming effects to vary by a wide range of moderating individual difference and experimental context variables, and absent well-developed theories for specifying such variables, the conclusions of replication failures will be ambiguous" (Cesario, 2014, p 41).

The results in this thesis certainly sound like those Cesario speaks of. They varied widely in response to individual differences and experimental variables, and, it is clear from the reported failures to replicate, along with Wheeler et al's (2013) review paper, that this is not uncommon among priming research. The effects were often (although not always) small, and we should be hesitant to predict the same results again even using the same voices and participant demographics. My reasons for this are much the same as those Cesario argues for, but are perhaps best summed up by Neisser, whose final chapter in *Cognition and Reality* (1976) lays out a convincing argument on the limits of prediction for human behaviour. He wrote that "The prediction and control of behaviour in the real world requires detailed knowledge of that world to a degree that we do not usually have, and that in any case falls outside the realm of psychological expertise" (Neisser, 1976, p 184). There are numerous examples in science of things not yet available for accurate prediction.

To illustrate the difficulty involved in predicting priming effects, however, I'll use the analogy of a runaway train to first talk about predicting non-human behaviour. In order to predict what will happen to this train, one would need to know the length of the train, its weight, the power of its engine, the speed it was travelling at before the driver lost control, the direction and speed of the prevailing wind, rain conditions, bends in the track, the rigidity of the track, and

many more variables. The information will only ever be *nearly* perfect, and as Stiglitz remarked about the banking collapse of the past few years, *almost perfect* information is not the same as *perfect* information (Stiglitz, 2010, p 243). Now imagine that rather than a runaway train with fixed characteristics on a fixed track and a reasonably fixed environment, you have humans with agency (or some degree thereof) in a changeable environment interacting with other similar changeable humans. This is what priming researchers work with. Joseph Cesario asks psychologists to “stop drawing sweeping conclusions about the ubiquity of priming effects before such conclusions are warranted” (Cesario, 2014, p 45). In his article he rightly points out that based on what we know about humans, and how the mind operates, it is unreasonable to suggest a universal method of priming that works for all people in all environments. The results from this thesis certainly support not drawing sweeping conclusions. The results were not neat, and the effect sizes not great.

A concern with social psychology and priming, particularly in light of the recent controversy, is the lack of attention paid to whether or not direct replications under different circumstances can really be thought of as testing the same theory as the original research (Stroebe & Strack, 2014). The authors write that “Repeating a specific operationalization of a theoretical construct at a different point in time and/or with a different population of participants might not reflect the same theoretical construct that the same procedure operationalized in the original study” (Stroebe & Strack, 2014, p 61). In this thesis, young citizens from a liberal country were used to test ideas about whether voices can influence their answers to social questions. Although the *macro theory* is voices will have effects, there is an assumed *underlying theory* that informs the direction I expect the effects to follow. If this study were to be repeated in the United Arab Emirates, the macro theory would remain the same but this underlying theory would, of course, change. Similarly for the reaction time study, there is an underlying theory about how the young and elderly view each other. This would change from place to place and person to person. This was expressed more clearly in Section 7.2 in the discussion of the work of Brewer and Lui (1984) and Levy (1996), on how stereotypes are multifaceted, non-fixed entities.

## 8.4 Implications and Future Directions

The implications of my work are directly influenced by the limitations outlined above. I would argue that the first point is to develop a better understanding of the environment and the stimuli, before making ever more complex theories of the mental mechanisms that may or may not respond to and process them. In this respect, I am in agreement with Neisser (1976) and De Houwer (2011), who have both argued for such an approach. De Houwer makes a case for a functional approach, where behavioural effects are defined in terms of the “causal impact of the environment on behaviour” (De Houwer, 2011, p 204). This requires thinking about the effects in a different manner to most of the research presented above, including those of the theories that were offered as possible explanations for the effects. It would say that the effect of a given voice on a given behaviour is another given behaviour. That, for example, hearing a Chinese voice whilst listening to a certain type of music leads people to perceive the music as more Asian. This behaviour would change in line with any contextual and environmental change, as well as any change in stimuli, as Neisser (1976) sets out. This certainly makes global predictions futile, but as argued above, global predictions should not be the aim. Any real world predictions must involve a constant monitoring of the contextual and environmental landscape so that those predictions can be updated, or abandoned, as new information becomes available. If we have in front of us a photograph of a concrete slab, we may call it a slab. If we see the same slab with someone lying on it wrapped in a blanket, we may call it a bed. We are inferring the construct not only from the stimulus, but from the behaviour of people with the stimulus. So too a certain voice is one thing in one context, but possibly something entirely different in another - a product of the listener, the context, and the environment.

To deal now directly with the effects, I would argue that these experiments offer initial evidence that when we hear people speaking, such as on the telephone, over a radio, on a sat-nav system and so forth, there is a possibility that the speech can have subtle and automatic effects on their attitudes and beliefs at that time. Further, we might conclude that these effects would go further than just attitudes and beliefs, and into the realm of choice and decision making. The type of voices advertisers choose to use, or the voices we hear when we call our telephone banking services, are just a few examples of environments where the voice may do more

than simply present a product or perform a transaction. In industries where voice is a commodity, much effort is spent on recording voices that are likeable, or task appropriate. In future, there should be a consideration of voice outcomes, both theoretically and practically.

There are instances of research that seek to draw the connections between voice and social behaviour more generally. (MacFarlane & Stuart-Smith, 2012) showed that certain accents in Glasgow were associated with certain brands. Particularly, middle class accents were more strongly identified with middle class brands, the reverse being true for working class speech and brands. People clearly associated voices with certain personalities and ideals, and it is no great surprise that different voices would prime different levels of agreement. Advertisers, again choosing their context (their audience, time of day and so forth), may find that the voices they use to advertise their products have effects over and above the advert being liked, or the voice being recognised. If people associate certain voices with certain ideals or ways of life or behaviours, then I see no obstacle to the voices having behavioural effects.

As mentioned above, future work should also address this speech-to-behaviour link from the opposite direction, from behaviour-to-speech. Based on both the present results and the results of other works that have looked at socially motivated language change, a natural prediction follows that individual speech could be manipulated by behavioural outcomes. If a person comes to associate speaking in a breathy tone with more sales in their job, or a higher pitch with being interrupted more in the boardroom, then I would predict that they increase their breathiness or decrease their pitch over time, largely automatically. There is little if any empirical work looking at how individuals develop their speech styles. Work on this area is almost exclusively at the group level, perhaps understandably so. Yet it feels true that our voices develop as we pass puberty and into adulthood, and that they take on certain characteristics as we progress through life - characteristics not accounted for by standard models of language change. We might look at an individual's history of voice-based reinforcement, both implicit and explicit, and begin to explore how this may shape voice variation.

Finally, it is important to discover how far this priming can extend. The types of



behaviour it can affect, the strength of the vocal cues needed, and also the length of time these effects might last. Priming has been shown typically to have short-lived effects, and indeed it must do, for if the environment is to change, so too should the priming effects. Yet what else is our day but a succession of temporary events that we perceive in a flowing state. It is no great criticism of priming that it produces effects that are short lived. Certain types of priming may only last for a minute, but what happens in that short window could have much longer lasting effects. Notwithstanding the limitations of the present experiments as I have described, and the difficulty of working with human voices as a tool, this should prove a fruitful area of research with some surprising results.

## References

- Allen, T. J., & Sherman, J. (2011). Ego threat and intergroup bias: A test of motivated-activation versus self-regulatory accounts. *Psychological Science, 22*, 331–3.
- Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison Wesley.
- Allport, G. (1985). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 1–46). Reading, MA: Addison-Wesley.
- Allport, G., & Cantril, H. (1934). Judging personality from voice. *Journal of Social Psychology, 5*, 37–55.
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*, 715–727.
- Arndt, J., Greenberg, J., & Cook, A. (2002). Mortality salience and the spreading activation of worldview-relevant constructs: Exploring the cognitive architecture of terror management. *Journal of Experimental Psychology: General, 131*, 307–324.
- Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humour, 16*, 243–260.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.
- Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000, January). Mixed-effects models in psychophysiology. *Psychophysiology, 37*(1), 13–20.
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*, 603–618.
- Banaji, M. R., Lemm, K. M., & Carpenter, S. J. (2001). The social unconscious. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 134–158). Malden, MA: Blackwell.
- Bargh, J. A. (1994). The four horsemen of Automaticity: Awareness, intention, efficiency, and control in social cognition. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A. (2005). Bypassing the will: Toward demystifying the nonconscious control of social behavior. In R. R. Hassin, J. S. Uleman, & J. Bargh (Eds.), *The new unconscious* (pp. 37–58). New York: Oxford University Press.

- Bargh, J. A. (2008). Free Will Is Un-natural. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? psychology and free will* (pp. 128–154). New York: Oxford University Press.
- Bargh, J. A., Chaiken, S., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*, 893 – 912.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Political Science Review*, *54*, 462–479.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, *71*, 230–44.
- Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925–945.
- Bargh, J. A., Raymond, P., Pryor, J. B., & Strack, F. (1995). Attractiveness of the underling: an Automatic power to sex association and its consequences for sexual harassment and aggression. *Journal of Personality and Social Psychology*, *68*, 768–81.
- Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*, 593–605.
- Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, *15*, 1–4.
- Bartlett, F. (1937). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bates, D., & Maechler., M. (2008). *lme4: Linear mixed effects models using S4 classes. R package version 3.0.1. 2008*. Retrieved from <http://CRAN.R-project.org>.
- Baugh, J. (2003). Linguistic profiling. In S. Makoni, G. Smitherman, A. F. Ball, & A. K. Speakers (Eds.), *Black linguistics: Language, society and politics in africa and the americas* (pp. 155–163). London: Routledge.
- Bazerman, M. (1983). Negotiator judgement. *American Behavioral Scientist*, *27*, 211–228.
- Beattie, G. (2003). *Visible thought: The new psychology of body language*. East Sussex: Routledge.
- Bell, A. (1984). Language style as audience design. *Language in Society*, *13*,

145–204.

- Bell, A. (1992). Hit and miss: Referee design in the dialects of New Zealand television advertisements. *Language and Communication*, *12*, 327–340.
- Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge: Cambridge University Press.
- Bem, D. J. (1972). Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality*, *40*, 17–26.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*, 354–364.
- Bjork, R. (1999). Bridging the gap between basic and applied research on the cognitive regulation of performance. In D. Gopher & A. Koriat (Eds.), *Attention and performance xvii - cognitive regulation of performance: Theory and application* (pp. 3–14). Cambridge, MA: MIT Press.
- Bodenhausen, G. V., & Lichtenstein, M. (1987, May). Social stereotypes and information-processing strategies: the impact of task complexity. *Journal of personality and social psychology*, *52*(5), 871–80. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3585699>
- Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. Wyer (Ed.), *Stereotype activation and inhibition: Advances in social cognition* (pp. 1–52). Hillsdale, NJ: Erlbaum.
- Bodenhausen, G. V., Schwarz, N., Bless, H., & Wanke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened or generalized appraisals? *Journal of Experimental Social Psychology*, *31*, 48–63.
- Bohner, G., & Chaiken, S. (1994). The role of mood and message ambiguity in the interplay of heuristic and systematic processing. *European Journal of Social Psychology*, *24*, 207–221.
- Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*, 265–92.
- Bornstein, R. F., & Dagostino, P. R. (1992). Stimulus recognition and the Mere Exposure effect. *Journal of Personality and Social Psychology*, *63*, 545–552.
- Braun, A. (1996). Age estimation by different listener groups. *Forensic Linguistics*, *3*, 65–73.
- Brewer, M. B., & Lui, L. (1984). Categorization of the elderly by the elderly: Effects of perceiver's category membership. *Personality and Social Psychology*

- Bulletin*, 10, 585–595.
- Broverman, I., Vogel, S. R., Clarkson, F., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues*, 28, 59–79.
- Brunel, F. F., & Nelson, M. (2003). Message order effects and gender differences in advertising persuasion. *Journal of Advertising Research*, 43(3), 330–340.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7, 585–614.
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological review*, 106, 676–713.
- Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, 16, 330–50.
- Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. Unpublished doctoral dissertation, Stanford University.
- Carli, L. L. (1989). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology*, 56, 565–576.
- Carlston, D. (2010). Models of implicit and explicit memory representation. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 38–61). New York: Guilford Publications.
- Carlston, D. E., & Skowronski, J. J. (1994). Savings in the Relearning of Trait Information as Evidence for Spontaneous Inference Generation. *Journal of Personality and Social Psychology*, 66(5), 840–856.
- Carroll, M., Byrne, B., & Kirsner, K. (1985). Autobiographical memory and perceptual learning: A developmental study using picture recognition, naming latency, and perceptual identification. *Memory & Cognition*, 13, 273–9.
- Cartwright, D. (1949). Some principles of mass persuasion: Selected findings of research on the sale of United States war bonds. *Human Relations*, 2, 253–267.
- Carver, C. S., Ganellen, R. J., Froming, W. J., & Chambers, W. (1983). Modeling: An analysis in terms of category accessibility. *Social Cognition*, 16, 225–252.
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9, 40–48.
- Cesario, J., Plaks, J. E., & Higgins, E. T. (2006). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social*

- Psychology*, 90, 893–910.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
- Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79, 115–153.
- Chasteen, A. L., Schwarz, N., & Park, D. C. (2002). The activation of aging stereotypes in younger and older adults. *Journal of Gerontology*, 57, 540–7.
- Chebat, J.-C., Gelinat-Chebat, C., Hombourger, S., & Woodside, A. G. (2003). Testing consumers' motivation and linguistic ability as moderators of advertising readability. *Psychology and Marketing*, 20, 599–624.
- Chen, S. X., Benet-Martínez, V., & Ng, J. C. K. (2013). Does language affect personality perception? A functional approach to testing the Whorfian hypothesis. *Journal of Personality*, 1–14.
- Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, 36, 1514–28.
- Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Correll, J., Judd, C. M., Park, B., & Wittenbrink, B. (2010). Measuring prejudice, stereotypes and discrimination. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *Prejudice, stereotyping and discrimination* (pp. 45–62). London: Sage.
- Costa, P., & McCrae, R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.
- Coupland, N. (1980). Style-shifting in a Cardiff work-setting. *Language in Society*, 9, 1–12.
- Coupland, N. (2001). Language, situation, and the relational self: Theorizing dialect-style in sociolinguistics. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 185–210). Cambridge: Cambridge University Press.
- Coupland, N. (2010). The authentic speaker and the speech community. In C. Llamas & D. Watt (Eds.), *Language and identities* (pp. 99–112). Edinburgh: Edinburgh University Press.

- Crawford, M. T., McCarthy, R. J., Kjaerstad, H. L., & Skowronski, J. J. (2013, March). Inferences are for doing: the impact of approach and avoidance states on the generation of spontaneous trait inferences. *Personality & social psychology bulletin*, *39*(3), 267–78. doi: 10.1177/0146167212473158
- Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007, May). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472. doi: 10.1016/j.jesp.2006.04.003
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–14.
- De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*, 202–209.
- Deaux, K. (1985). Sex and gender. *Annual Review of Psychology*, *36*, 49–81.
- Decoster, J., & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review*, *8*, 1–27.
- DeMarree, K. G., & Loersch, C. (2009). Who am I and who are you? Priming and the influence of self versus other focused attention. *Journal of Experimental Social Psychology*, *45*, 440–443.
- Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology*, *36*, 531–544.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (33rd ed., Vol. 33, pp. 1–40).
- Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D., Koomen, W., Knippenberg, A. V., & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology*, *75*, 862–871.
- Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology*, *74*, 865–77.

- Dinsmoor, J. (1999). Foreword. In B. Thyer (Ed.), *The philosophical legacy of behaviourism* (pp. 1–8). Netherlands: Kluwer.
- Dockrell, J., & Messer, D. (1999). *Children's language and communication difficulties*. London: Continuum International Publishing.
- Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Volume 2 ed., pp. 1084–1119). Hoboken, New Jersey: John Wiley and Sons.
- Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS one*, *7*, 1–7.
- Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *The American Psychologist*, *64*, 644–58.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*, 543–558.
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly*, *15*, 203–216.
- Eckert, P. (2000). *Linguistic variation as social practice*. Oxford: Blackwell.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 453–476.
- Egan, O. (1980). Intonation and meaning. *Journal of Psycholinguistic Research*, *9*, 23–39.
- Epley, N., & Gilovich, T. (2010). Anchoring unbound. *Journal of Consumer Psychology*, *20*, 20–24.
- Ervin-Tripp, S. (1972). Alternation and co-occurrence. In J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 218–250). Rinehart and Winston.
- Fagot, B. I., & Leinbach, M. D. (1989). The young child's gender schema: Environmental input, internal organization. *Child Development*, *60*, 663.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*, 115–141.
- Fazio, R. H., & Olson, M. a. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.



- Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 216–246). New York: Guilford Publications.
- Fiske, S. T. (2005). Social cognition and the normality of prejudice: 50 years after Allport. In J. F. Dovidio, P. Glick, & L. Rudman (Eds.), *On the nature of prejudice* (pp. 36–53). Malden, MA: Blackwell.
- Fiske, S. T., & Linville, P. W. (1980). What does the schema concept buy us? *Personality and Social Psychology Bulletin*, *6*, 543–557.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (23rd ed., pp. 1–74). San Diego, CA: Academic Press.
- Fiske, S. T., & Taylor, S. (1991). *Social cognition* (2nd ed.). New York: McGraw Hill.
- Fitzsimons, G. M., Chartrand, T. L., & Fitzsimons, G. J. (2008). Automatic effects of brand exposure on motivated behavior: How apple makes you ‘think different’. *Journal of Consumer Research*, *35*, 21–35.
- Forehand, M. R., Deshpandé, R., & Reed, Americus, I. (2002). Identity salience and the influence of differential activation of the social self-schema on advertising response. *Journal of Applied Psychology*, *87*, 1086–1099.
- Forgas, J. P. (1992). On mood and peculiar people: Affect and person typicality in impression formation. *Journal of Personality and Social Psychology*, *62*, 863–875.
- Förster, J., Liberman, N., & Friedman, R. S. (2009). What do we prime? On distinguishing between semantic priming, procedural priming, and goal priming. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action. social cognition and social neuroscience* (pp. 173–192). New York: Oxford University Press.
- Fozard, J. L. (1990). Vision and hearing in aging. In J. Birren & K. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 150–159). New York: Academic Press.
- Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs.

- Experimental Psychology*, 49, 171–180.
- Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology*, 89, 672–85.
- Gawronski, B., & Strack, F. (2012). Cognitive consistency as a basic principle of social information processing. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 1–18). New York: Guilford Publications.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–82.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87–105.
- Giles, H., & Billings, A. (2004). Assessing language attitudes: Speaker evaluation studies. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 187–209). Oxford: Blackwell Publishing Ltd.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge: Cambridge University Press.
- Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars*. Mahwah, NJ: Lawrence Erlbaum.
- Gilovich, T., & Griffin, D. (2010). Judgement and decision making. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 542–588). New Jersey: Wiley and Sons.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–79.
- Goldstein, E. (2008). *Cognitive psychology: Connecting mind, research and everyday experience*. Belmont, CA: Thomson Wadsworth.
- Gough, H. G. (1956). *California Psychological Inventory*. Consulting Psychologists Press.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual

- differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–80.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.
- Han, J. J., Leichtman, M. D., & Wang, Q. (1998). Autobiographical memory in Korean, Chinese, and American children. *Developmental Psychology*, *34*, 701–13.
- Harwood, J., Giles, H., Ota, H., Pierson, H. D., Gallois, C., Hung Ng, S., . . . Somera, L. (1996). College students' trait ratings of three age groups around the Pacific Rim. *Journal of Cross-Cultural Gerontology*, *11*, 307–317.
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *4*, 865–892.
- Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh : Exemplar priming in speech perception. *The Linguistic Review*, *23*, 351–379.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*, 61–135.
- Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, *51*, 1106–1115.
- Hess, T. M., Hinson, J. T., & Statham, J. A. (2004). Explicit and implicit stereotype activation effects on memory: Do age and awareness moderate the impact of priming? *Psychology and Aging*, *19*, 495–505.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Psychology*, *13*.
- Hilsenrat, M., & Reiner, M. (2011). The impact of subliminal haptic perception on the preference discrimination of roughness and compliance. *Brain Research Bulletin*, *85*, 267–70.
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*, 237–71.
- Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.
- Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the 'coordinates of truth'. *Perspectives on Psychological Science*, *6*, 253–271.
- Hogg, M. A. (2010). Influence and leadership. In S. T. Fiske, D. T. Gilbert, &

- G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 1166–1206). New York: Wiley.
- Hovland, C. (1959). Reconciling conflicting results derived from experimental and survey studies on attitude change. *American Psychologist* (January), 8–17.
- Hovland, C., Janis, I., & Kelley, H. (1953). *Communication and persuasion*. New Haven: Yale University Press.
- Hovland, C., & Pritzker, H. A. (1957). Extent of opinion change as a function of amount of change advocated. *Journal of Abnormal Psychology* (March), 257–261.
- Hovland, C., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly, Winter*, 635–650.
- Huang, J. Y., Song, H., & Bargh, J. A. (2011). Smooth trajectories travel farther into the future: Perceptual fluency effects on prediction of trend continuation. *Journal of Experimental Social Psychology, 47*, 506–508.
- Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology, 77*, 1011–25.
- Hummert, M. L. (1990). Multiple stereotypes of elderly and young adults: A comparison of structure and evaluations. *Psychology and Aging, 5*, 182–93.
- Hummert, M. L., Shaner, J. L., & Garstka, T. A. (1995). Cognitive process affecting communication in older adults: The case for stereotypes, attitudes, and beliefs about communication. In J. Nussbaum & J. Coupland (Eds.), *The handbook of communication and aging research* (pp. 105–132). Mahwah, NJ: Erlbaum.
- Hundhammer, T., & Mussweiler, T. (2012). How sex puts you in gendered shoes: Sexuality-priming leads to gender-based self-perception and behavior. *Journal of Personality and Social Psychology, 103*, 176–93.
- Hymes, D. (1974). *Foundations of sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Iidaka, T., Anderson, N. D., Kapur, S., Cabeza, R., & Craik, F. I. (2000). The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. *Journal of Cognitive Neuroscience, 12*, 267–80.
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*, 995–1006.

- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology. General*, *110*, 306–40.
- Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*, 283–93.
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*, 1338–1339.
- Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions and insurance decisions. *Journal of Risk and Uncertainty*, *7*, 35–41.
- Johnson, M. K., & Hasher, L. (1987). Human learning and memory. *Annual Review of Psychology*, *38*, 631–86.
- Johnston, L. (2002). Behavioral mimicry and stigmatization. *Social Cognition*, *20*, 18–35.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrarm Straus and Giroux.
- Kahneman, D., & Frederick, S. (2002). Heuristics of intuitive judgment: Extensions and applications. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgement: Extensions and applications* (pp. 1–30). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.
- Kawakami, K., Young, H., & Dovidio, J. F. (2002). Automatic stereotyping: Category, trait, and behavioral activations. *Personality and Social Psychology Bulletin*, *28*, 3–15.
- Keller-Cohen, D., & Gordon, C. (2003). On trial: Metaphor in telling the life story. *Narrative Inquiry*, *13*, 1–40.
- Kenton, S. (1989). No Title. *International Journal of Business Communication*, *26*, 143–157.
- Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, *16*(8), 617–25.

- Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity*, 37–41.
- Krieglmeyer, R., & Sherman, J. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, 103, 205–24.
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, 129, 522–544.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14, 334–9.
- Langer, E. (1989). *Mindfulness*. Reading, MA: Addison Wesley Longman.
- Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.
- Lerner, J. S. (2000). Beyond valence : Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, 14(4), 473–493.
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15, 374–379.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, 71, 1092–107.
- Levy, B., Ashman, O., & Dror, I. (2000). To be or not to be: The effects of aging stereotypes on the will to live. *Omega*, 40, 409–20.
- Levy, B., & Langer, E. (1994). Aging free from negative stereotypes: Successful memory in China and among the American deaf. *Journal of Personality and Social Psychology*, 66, 989–97.
- Lim, L. (1997). Intonation patterns characterising three ethnic varieties of English in Singapore: Observations and implications. In *Intonation: Theory, models and applications. proceedings of an esca workshop*. Athens: ESCA & University of Athens.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.
- Linville, P. W., Fischer, G. W., Fischhoff, B., & Pryor, J. B. (1993). AIDS risk perceptions and decision biases. In G. D. Reeder (Ed.), *The social psychology*

- of *hiv infection* (pp. 5–38). Hillsdale, NJ: Lawrence Erlbaum.
- Lips, H. (1988). *Sex and gender: An introduction*. Mountain View, CA: Mayfield.
- Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.
- Lombardi, W. J., Higgins, E. T., & Bargh, J. A. (1987). The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin, 13*, 411–429.
- Lowrey, T. (1998). The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology, 7*, 187–206.
- Luna, D., & Peracchio, L. A. (2001). Moderators of language effects in advertising to bilinguals: A psycholinguistic approach. *Journal of Consumer Research, 28*, 284–295.
- MacFarlane, A. E., & Stuart-Smith, J. (2012). ‘One of them sounds sort of Glasgow Uni-ish’: Social judgements and fine phonetic variation in Glasgow. *Lingua, 122*, 764–778.
- Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92*, 239–255.
- Macrae, C. N., & Johnston, L. (1998). Help, I need somebody: Automatic action and inaction. *Social Cognition, 16*, 400 – 417.
- Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 37–47.
- Markus, H., & Zajonc, R. (1985). The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 137–230). New York: Random House.
- McCrae, R., Yik, M. S., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. *Journal of Personality and Social Psychology, 74*, 1041–55.
- McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental*

- Psychology. Learning, memory, and Cognition*, 18, 1155–72.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McVee, M. B., Dunsmore, K., & Gavelek, J. R. (2005). Schema theory revisited. *Review of Educational Research*, 75, 531–566.
- Mendoza-Denton, N. (2008). *Homegirls: language and cultural practice among latina youth gangs*. Malden, MA: Wiley.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Milroy, L., & Gordon, M. (2002). *Sociolinguistics: Method and interpretation* (2nd ed.). Oxford: Wiley Blackwell.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W. (1997). Was the cognitive revolution just a detour on the road to behaviorism? On the need to reconcile situational control and personal control. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition, vol 10*. (pp. 181–186). Mahwah; NJ: Lawrence Erlbaum.
- Mortensen, C. R., & Cialdini, R. B. (2010). Full-cycle social psychology for theory and application. *Social and Personality Psychology Compass*, 4, 53–63.
- Moskowitz, G. B., & Skurnik, I. W. (1999). Contrast effects as determined by the type of prime: Trait versus exemplar primes initiate processing strategies that differ in how accessible constructs are used. *Journal of Personality and Social Psychology*, 76, 911–927.
- Murphy, S. T., & Zajonc, R. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64, 723–39.
- Musolff, A. (2006). Metaphor scenarios in public discourse. *Metaphor and Symbol*, 21, 23–38.
- Mussweiler, T. (2006). Doing is for thinking! Stereotype activation by stereotypic movements. *Psychological Science*, 17, 17–21.
- Mussweiler, T., & Damisch, L. (2008). Going back to Donald: How comparisons shape judgmental priming effects. *Journal of Personality and Social Psychology*, 95, 1295–315.
- Mussweiler, T., & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology*,



- 78, 1038–1052.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology*, *106*, 226–254.
- Neisser, U. (1976). *Cognition and reality*. San Francisco: WH Freeman and Company.
- Neumann, E., & Deschepper, B. G. (1992). An inhibition-based fan effect: Evidence for an active suppression mechanism in selective attention. *Canadian Journal of Psychology*, *46*, 1–40.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*, 62–85.
- Nisbett, R. (2003). *The geography of thought*. London: Nicholas Brealey.
- Nolan, F. (2005). Forensic speaker identification and the phonetic description of voice quality. In W. J. Hardcastle & J. Mackenzie (Eds.), *A figure of speech* (pp. 385–411). Mahwah, NJ: Lawrence Erlbaum.
- Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, *7*, 171–188.
- Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, *106*, 1178–94.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–93.
- Piliavin, J. A., & Martin, R. R. (1978). The effects of the sex composition of groups on style of social interaction. *Sex Roles*, *4*, 281–296.
- Podesva, R. J., Roberts, S. J., & Campbell-Kibler, K. (2006). Sharing resources and indexing meanings in the production of gay. In D. Cameron & D. Kulick (Eds.), *The language and sexuality reader* (pp. 141–150). London: Rout.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–32.

- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*, 338–42.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on perceptual judgements. *Psychological Science*, *9*, 45–48.
- Riether, N., Hegel, F., Wrede, B., & Horstmann, G. (2012). Social facilitation with social robots? *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12*, 41.
- Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: Implied social presence, eye tracking and attention. *Attention, Perception and Psychophysics*, *73*, 291–6.
- Robinson, J., & McArthur, L. Z. (1982). Impact of salient vocal qualities on causal attribution for a speaker's behavior. *Journal of Personality and Social Psychology*, *43*, 236–247.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.
- Rosenkrantz, P. S., Vogel, S., Bee, H., & Broverman, I. (1968). Sex-role stereotypes and self concepts in college students. *Journal of Consulting and Clinical Psychology*, *32*, 287–295.
- Rothman, A. J., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin*, *24*, 1053–1064.
- Russell, C. A. (2003). Investigating the effectiveness of product placements in television shows: The role of modality and plot connection congruence on brand memory and attitude. *Journal of Consumer Research*, *29*, 306–319.
- Russet, B., Starr, H., & Kinsella, D. (2010). *World politics: The menu of choice*. Belmont, CA: Wadsworth.
- Sapir, E. (1933). Language. In *Encyclopaedia of the social sciences* (pp. 155–169). New York: Macmillan.
- Sartre, J.-P. (1965). *Essays in existentialism* (1993rd ed.). New York: Citadel.
- Schilling-Estes, N. (1998). Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society*, *27*, 53–83.
- Schilling-Estes, N. (2006). Dialect variation. In R. Fasold & Connor-Linton J. (Eds.), *An introduction to language and linguistics* (pp. 311–341). Cambridge: Cambridge University Press.
- Schilling-Estes, N. (2008). Introduction. In J. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change*.

Oxford: Blackwell Publishing Ltd.

- Schön, D. (1979). Generative metaphor: A perspective on problem-setting in social policy. In *Metaphor and thought* (pp. 254–283). Cambridge: Cambridge University Press.
- Schötz, S. (2006). *Perception, analysis and synthesis of speaker age*. Unpublished doctoral dissertation, Lund University.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*, 332–348.
- Schwarz, N. (2011). Feelings-as-information theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology: Collection: Volumes 1 & 2* (pp. 289–308). Thousand Oaks, CA: Sage.
- Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. In M. Zanna (Ed.), *Advances in experimental social psychology* (24th ed., Vol. 24, pp. 161–199). New York: Academic Press.
- Schwarz, N., Bless, H., Wanke, M., & Winkielman, P. (2003). Accessibility revisited. In G. Bodenhausen & A. J. Lambert (Eds.), *Foundations of social cognition* (pp. 51–77). Mahwah, NJ: Lawrence Erlbaum.
- Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 436–457). Malden, MA: Blackwell Publishing Ltd.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*(3), 513–523.
- Schwarz, N., Strack, F., Kommer, D., & Wagner, D. (1987). Soccer, rooms, and the quality of your life: Mood effects on judgements of satisfaction with life in general and with specific domains. *European Journal of Social Psychology, 17*, 69–79.
- Sedikides, C., & Skowronski, J. T. (1991). The law of cognitive structure activation. *Psychological Inquiry, 2*, 169–184.
- Serpell, L., & Green, J. (2006). Parental decision-making in childhood vaccination. *Vaccine, 24*, 4041–6.
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*, 207–22.

- Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., ... Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PloS one*, *8*, 1–10.
- Sharma, D., Booth, R., Brown, R., & Huguet, P. (2010). Exploring the temporal dynamics of social facilitation in the Stroop task. *Psychonomic Bulletin & Review*, *17*, 52–8.
- Shepard, C., Giles, H., & Le Poire, B. (2001). Communication accommodation theory. In P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33–56). Wiley.
- Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*, 13–39.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Sinclair, R. C., Moore, S. E., Mark, M. M., Soldat, A. S., & Lavis, C. a. (2010). Incidental moods, source likeability, and persuasion: Liking motivates message elaboration in happy people. *Cognition & Emotion*, *24*, 940–961.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. Middlesex, England: Penguin Books.
- Smith, E. R., & Queller, S. (2001). Mental representations. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 111–133). Malden, MA: Blackwell Publishing Ltd.
- Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological science*, *19*, 986–8.
- Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Duinn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152.
- Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672.
- Stangor, C., & Schaller, M. (1996). Stereotypes as individual and collective representations. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 3–37). New York: Guilford Publications.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences*,

23, 645–726.

- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.
- Stiglitz, G. (2010). *Freefall: America, free markets, and the sinking of the world economy*. New York: W. W. Norton & Company.
- Strack, F., & Mussweiler, T. (2003). Heuristic strategies for estimation. In G. Bodenhausen & A. J. Lambert (Eds.), *Foundations of social cognition* (pp. 79–95). Mahwah, NJ: Lawrence Erlbaum.
- Strack, F., & Schwarz, N. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23*, 53–62.
- Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology, 38*, 556–568.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*, 59–71.
- Stuart-Smith, J., Timmins, C., & Tweedie, F. (2007). ‘Talkin’ Jockney’? Variation and change in Glaswegian accent 1. *Journal of Sociolinguistics, 11*, 221–260.
- Tagliamonte, S. (2011). *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley Blackwell.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*, 189–208.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & S. Madigan (Eds.), *Organization of memory* (pp. 381–402). New York: Academic Press.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency. *Cognitive Psychology, 232*, 207–232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–31.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59*.
- Valentino, N. A. (1999). Crime news and the piming of racial attitudes during evaluations of the president. *Public Opinion Quarterly, 63*, 293.

- Valentino, N. A., Hutchings, V. L., & White, I. K. (2002). Cues that matter: How political ads prime racial attitudes during campaigns. *American Political Science Review*, *96*, 75–90.
- Wheeler, S. C., & Berger, J. (2013). When the same prime leads to different effects. *Journal of Consumer Research*, *34*, 357–368.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127*, 797–826.
- Wigboldus, D., Sherman, J., Franzese, H., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*, 292–309.
- Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J. L. (2007). Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology*, *11*(2), 179–192.
- Winkielman, P., Zajonc, R., & Schwarz, N. (1997). Subliminal affective priming resists attributional interventions. *Cognition & Emotion*, *11*, 433–465.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology*, *37*, 244–252.
- Woodland, J., & Voyer, D. (2011). Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, *26*, 227–239.
- Wundt, W. (1907). *Outlines of psychology*. Leipzig: Wilhelm Englemann.
- Wyer, R., & Carlston, D. (1979). *Social Cognition, Inference, and Attribution*. Hillsdale, NJ: Erlbaum.
- Wyer, R. S. (2007). Principles of mental representation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 285–307). New York: Guilford Publications.
- Yabar, Y., Johnston, L., Miles, L., & Peace, V. (2006). Implicit behavioral mimicry: Investigating the impact of group membership. *Journal of Nonverbal Behavior*, *30*, 97–113.
- Yong, E. (2012a). Bad Copy. *Nature*, *485*, 298–300.
- Yong, E. (2012b). *Nobel laureate challenges psychologists to clean up their act*. doi:10.1038/nature.2012.11535
- Zajonc, R. (1965). Social facilitation. *Science*, *149*, 269–274.
- Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality*

*and Social Psychology, 9, 1–27.*

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35, 151–175.*

# Appendices



# Appendix A

## Statement List for Experiments 1A, 1B and 1C

Below are the statements that were designed to test the original hypothesis of social support varying as a function of voice condition. The original coding from Experiment 1A is shown in brackets, where *supportive* is in reference to support for the social category.

### Sexual orientation themed questions

- Same sex adoption should be treated equally to traditional adoption (*yes* answer = supportive)
- Sexual preference is biologically determined (*yes* answer = supportive)
- I would be comfortable if a close friend came out as gay (*yes* answer = supportive)
- It is unfair that gay men and women can refer to themselves using words such as queer and dyke, when straight people using these words would be considered bigoted (*no* answer = supportive)
- I am comfortable with the idea of same sex marriage (*yes* answer = supportive)

### Gender themed questions

- Women still face the glass ceiling in organisations, i.e. there is a limit to how far they can progress in an organization (*yes* answer = supportive)
- I am comfortable with seeing breastfeeding in public (*yes* answer = supportive)

- Men are better drivers than women (*no* answer = supportive)
- The 14 weeks maternity leave currently offered by the New Zealand government is enough (*no* answer = supportive)
- Men require the same amount of paternity leave as women require maternity leave (*no* answer = supportive)

#### **Age themed questions**

- I would give up my seat on a bus to an elderly, but physically capable person (*yes* answer = supportive)
- It is reasonable to expect people to work until they are 70 (*no* answer = supportive)
- The elderly, i.e. post-retirement, are wiser (*yes* answer = supportive)
- Children have an obligation to care for their parents when they are no longer able to care for themselves (*yes* answer = supportive)
- I am comfortable with growing old (*yes* answer = supportive)
- I prefer to spend time in company of people younger than myself (*no* answer = supportive)

#### **Ethnicity themed questions**

- Immigrants have equal access to work and employment opportunities in New Zealand (*no* answer = supportive)
- I would consider an immigrant to be local if they had lived in my country for a reasonable length of time (*yes* answer = supportive)
- I am comfortable interacting with non-native speakers of English (*yes* answer = supportive)
- The following excerpt of speech is pleasant to listen to (Cantonese excerpt) (*yes* answer = supportive)
- The culture of a country suffers when immigration is high (*no* answer = supportive)

## Appendix B

# Question list for Experiment

## 2

### Ethnicity themed questions

- Please guess the country of origin of the following excerpt of music (*The music was Mandarin-spoken rap music played in reverse to obscure obvious language identification*)
- The following excerpt of speech pleasant to listen to (*The excerpt was a piece of dialogue between male and female character in a Hong Kong movie*)
- Name a country that comes to mind when you think about human rights violations (*There was no accompanying stimulus for this question*)

### Sex themed questions

- Please estimate your confidence, from 0% to 100%, whether the following individual is male (*The image was a composite of a Caucasian male and female face*)
- Please estimate your confidence, from 0% to 100%, whether the following voice is female (*The audio clip was an excerpt of the American actress Kathleen Turner, who has a distinctly deep voice*)
- Please guess the percentage of the crowd in this photograph that is male (*The photograph was an image of the Wimbledon tennis crowd*)

### Sexual orientation themed questions

- Please estimate your confidence, from 0% to 100%, whether the following voice is from a gay male (*The speaker was a young Caucasian gay male reciting a tongue twister*)
- Please estimate your confidence, from 0% to 100%, whether the following photograph is of a heterosexual man (*The photograph was of a young gay male in a formal, corporate-style photograph*)
- Please estimate the percentage of a poll of 1000 men that listed Lady Gaga among their favourite artists (*There was no accompanying stimulus for this question*)

#### **Age themed questions**

- What is the best age to be? (*There was no accompanying stimulus for this question*)
- How old do you think the person in this photograph is? (*The photograph was a 63 year old Caucasian male*)
- How many times per year does an average person visit their doctor? (*There was no accompanying stimulus for this question*)

## Appendix C

# Information and Debrief Sheets

### C.1 Information and Debrief for Experiments 1 and 2

## INFORMATION

**This was be presented in the email invitation to the volunteers. A further consent button and reminder were presented on the first page of the questionnaire, indicating whether participants still wish to continue.**

You have been invited to participate in a questionnaire, assessing your views on a range of socially sensitive topics where you will hear a series of questions being read out (this will require you to have audio access on your computer) and options for you to answer with. It is being conducted under the New Zealand Institute of Language, Brain and Behaviour. Your participation is strictly anonymous. Please be as honest as possible in the questions, and don't try to 'second guess' the questions; we want your real views/guesses, not the ones you think you *ought to* report.

You have the right to withdraw from the project at any time, including withdrawal of any information provided.

The results of the project may be published, but you may be assured of the complete anonymity of data gathered in this investigation.

The project is being carried out as part of a larger PhD study by Andrew Euan MacFarlane and under the supervision of Professor Jennifer Hay, who can be contacted at [jen.hay@canterbury.ac.nz](mailto:jen.hay@canterbury.ac.nz). She will be pleased to discuss any concerns you may have about participation in the project.

**Your agreement to participate in the study, and permission for your data to be used in subsequent analyses, is implied by completion and submission of the questionnaire.**

The project has been reviewed *and approved* by the University of Canterbury Human Ethics Committee.

# Debriefing Information

This was presented in the penultimate page of the online questionnaire

The primary focus of this study is to establish whether people's answers to socially sensitive questions depend in part on the voice asking the question. Different versions of the questionnaire were sent to volunteers; they contained the same questions, but were read out in different voices. You were not told of the importance of the voice because it could have invalidated any differences we find across groups; priming (an earlier stimulus e.g. voice influencing a response to a later stimulus e.g. question) typically requires that the participant is oblivious to the prime. If you wish to know the results of the study, you may email the address given below.

If you are happy to allow your results to be used to help with our study, please press 'submit' now. If you no longer wish to participate, you may close the window.

I would be more than happy to answer any questions you have about this study, and indeed hear your own thoughts on its nature.

With many thanks,

Andrew Euan MacFarlane

[Andrew.macfarlane@pg.canterbury.ac.nz](mailto:Andrew.macfarlane@pg.canterbury.ac.nz)

## C.2 Information and Debrief for Experiment 3





Te Kāhui Roro Reo • New Zealand Institute of  
**Language Brain & Behaviour**



## INFORMATION

**This was presented to participants on paper before the start of the reaction time experiment. If they were satisfied, they signed a consent form and continued.**

You have been invited to participate in a reaction time study investigating response times to visual stimuli. The experiment should last no longer than 15 minutes, and involves looking at images flashed on a computer screen, and responding to those images by pressing specific keys. It is being conducted under the New Zealand Institute of Language, Brain and Behaviour. Your participation is confidential.

You have the right to withdraw from the project at any time, including withdrawal of any information provided.

The results of the project may be published, but you may be assured of the complete anonymity of data gathered in this investigation.

The project is being carried out as part of a larger PhD study by Andrew Euan MacFarlane and under the supervision of Professor Jennifer Hay, who can be contacted at [jen.hay@canterbury.ac.nz](mailto:jen.hay@canterbury.ac.nz). She will be pleased to discuss any concerns you may have about participation in the project.

**Your agreement to participate in the study, and permission for your data to be used in subsequent analyses, is implied by completion and submission of the questionnaire.**

The project has been reviewed *and approved* by the University of Canterbury Human Ethics Committee.

## **Questionnaire**

Please state your age:

Please state your sex:

Please state your country of birth:

Please state your subject course:



## Debriefing Information

**This was presented to participants on paper after completion of the reaction time experiment.**

The primary focus of this study is to establish whether people's response times may be influenced by the age of the speaker who recorded the instructions. Some participants heard the young voice first, and some the older voice; I am interested in whether or not the response times between these two conditions is significantly different.

You were not told of the importance of the voice because it could have invalidated any differences we find; priming (an earlier stimulus e.g. voice influencing a response to a later stimulus e.g. question) typically requires that the participant is oblivious to the prime. Once the results are available I'm more than happy to inform you of the outcome of my experiment, and indeed your own individual results.

If you are happy to be included in this experiment, and allow your results to be used in both publication of this research and future work, you may proceed to read and sign the consent form. If you wish to withdraw from the study now, please indicate to the experimenter and your results will be discounted and no further action will be required on your part.

I would be more than happy to answer any questions you have about this study.

With many thanks,

Andrew Euan MacFarlane  
[Andrew.macfarlane@pg.canterbury.ac.nz](mailto:Andrew.macfarlane@pg.canterbury.ac.nz)