Accents on Trial

Accent Discrimination in U.K. and U.S. Courts

Grace Wood

PhD

University of York

Language and Linguistic Science

November 2018

Abstract

Research into urban housing, employment, education, and public perception has found evidence of accent discrimination. However, the role of language and discrimination has been under researched in the legal realm. Trials such as the U.S. *Zimmerman v. State* with witness Rachel Jeantel reveal how damaging accent discrimination can be. In order to research this further, mock trials were put together, collecting "verdicts" from groups of participants in the U.S. who formed mock juries, as well as individual online participants in the U.K. and the U.S. In both countries the national standard accent was compared to a regional accent. This was done by filming a staged cross examination between a prosecutor and a defendant. While the prosecutor's accent remained standard in both guises, participants were given a video that had the defendant testify in either a standard or regional accent. Unlike previous research, these studies were designed to look like psychological studies into jury decision making so that participants were not primed for the linguistic components.

The statistical analysis did not find any significant differences between the two accent conditions. Therefore, while language attitudes affected some of the results, there was no evidence that accent discrimination was present when it came to giving a verdict. Thus the overall findings suggest accent may not always be discriminated against directly; rather it may be the vehicle used to discriminate against protected traits (e.g. ethnicity, gender, religion, etc.). Furthermore, the mock juries differed significantly in their verdicts from the individual online participants. This suggests that the use of individual jurors in any type of trial research significantly inhibits ecological validity. Suggestions for further research are offered to continue learning in what ways language is instrumental in legal contexts.

Contents

List of Tables

List of Figures

Acknowledgements

Paul Foulkes – thank you. Thank you for being such an incredibly supportive and instructive supervisor. Your combined encouragement and constructive criticism from start to finish have been truly invaluable. Thank you for reading everything I ever wrote and teaching me what was expected in academic writing. Thank you for spending hours upon hours over the last few years brainstorming with me regarding various methodologies, hypothesizing over potential outcomes, encouraging me when things did not go according to plan, pushing me when I became complacent, and just in general being excited with me about my research. I couldn't be more grateful to have had you as my primary supervisor.

Thank you to my secondary supervisors, Paul Kerswill and Andrew MacFarlane. Andrew, you were a huge help and support in my first two years, always willing to talk through my methodology with me and teach me how statistics work. PK, thank you for immediately stepping in as Andrew stepped out, and helping me reach the finish line for this PhD over the course of another year and a half. Your experience and knowledge in the field have been so useful as I've written up this dissertation.

A huge thank you to my family and friends. You've been so supportive and encouraging of me as I've gone through this process, and more than understanding in my times of stress and anxiety. Thank you for being the rock I could lean on, a constant source of laughter and love, and a regular reminder of the truly important things in life. I love you all.

Thank you so much to Edie Greene for all your guidance, advice, and the use of your resources. I so appreciated you taking the time just to chat through my research ideas and future plans. Additionally, thank you (and Alan) for allowing me to stay with you while collecting my data. It was such a blessing and honestly the highlight of my entire fieldwork experience.

Thank you to Tarika Daftary-Kapur for giving me a space to work, the resources to collect data, and guiding me in mock jury research and general practices. I'm so grateful you were willing to work with me over an extended period and make my fieldwork possible.

Thank you to Kiara Quiñones for all your help and support. You were an absolute joy to work with. You helped my data collection run smoothly and never failed to lower my stress levels. Thank you.

To my wonderful actors and phenomenal 293 participants across all the pilots and full studies: I literally could not have done this without you. Thank you *so much* for taking time out of your day to be a part of this research. Your involvement was invaluable and your feedback will help me craft stronger studies in the future.

Author's Declaration:

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1. Introduction

On February 26, 2012 in Sanford, Florida a physical altercation took place between Trayvon Martin and George Zimmerman that concluded with Zimmerman shooting and killing Martin. Zimmerman admitted to the shooting but argued that he had been acting in self-defence. The only witness to this altercation was Rachel Jeantel, who had been on the phone nearly the entire time with Martin as the incident occurred. In *Zimmerman v. State* (2013), the prosecution called Jeantel to the witness stand to provide the version of events as Martin had relayed them to her and from what she was able to overhear. As a close friend of his, the prosecution also wanted her to shed light on Martin's emotional state both before and during the altercation (Thompson, 2015).

Figure 1. L-R: Jeantel, Martin, and Zimmerman (Rickford & King, 2016)

Jeantel took the stand for two days, providing an account of the altercation that directly contradicted Zimmerman's. He claimed it was self-defence, citing

Florida's "stand your ground" law (FLA. STAT. § 776.013, 2018), which allows victims to use excessive force when attacked. According to Jeantel, it was Zimmerman who instigated the altercation, following Martin around and antagonizing him before resorting to brutal force (Carlin, 2015). However, rather than hearing a completely different version of events than initially presented, listeners focused on her African American English (AAE) dialect. She was immediately mocked and attacked by the news media in the press, the defence attorney, and occasionally even the prosecutor, in court, as well as across social media for speaking in AAE (Abad-Santo, 2013; Johnson, 2013; Quigley, 2013). She was constantly asked to repeat herself. At times the attorneys would repeat her words back to her, sometimes clearly "fixing" her sentence to be in General American English (GAE) and other times emphasizing the perceived "ungrammatical" aspects of what she had said. An example of one such interaction between Jeantel, the prosecutor, and the court reporter is shown below.

1	*Jeantel*:	He started walkin home. //He told me –
2	*Prosecutor*:	//Ok, let me-// Let me stop you. He said **he started walking home**?
3	*Jeantel*:	Yeah, he's going to leave the area where the mailin area at.
4	*Ct Rep.*:	Ok. Say it again.
5	*Jeantel*:	He's going to leave the mailin area where he was at
6	*Ct Rep*:	**The mailing area**?
7	*Jeantel*:	Yeah. Where you get your mail at.
8	*Prosecutor*:	Alright. And I'll make sure everybody understands what you're saying. Did you say he said **he's gonna leave the *mailing area* where he's at?**
9	*Jeantel*:	Yes.
10	*Prosecutor*:	And what did you take that to mean?
11	*Jeantel*:	That he's leavin the area. He say he's gonna start walking home.

(Sullivan, 2016, p. 6-7)

The bolded speech uttered by the prosecutor and the court reporter in lines 2, 6, and 8 show where they corrected Jeantel's use of the non-standard [n] variant for the standard [ŋ] when pronouncing the –ING variable. The example above is one instance across numerous occurrences where Jeantel's speech is overtly corrected. Simply repeating Jeantel's speech could be interpreted in multiple ways, including a

desire to clarify her specific point to the jury. However, Sullivan notes that corrections such as these regularly occur throughout Jeantel's testimony after she has pronounced the [n] variant, and that the standard [ŋ] variant is emphasised in the corrector's tone (Sullivan, 2016; Sullivan, 2017). Additionally, the pronunciation of the –ING variable is not commonly considered a potential hindrance to understanding; rather it is a variable that carries social meaning (Campbell-Kibler, 2006). Therefore, it would appear that these repetitions of Jeantel's speech do not serve as clarifications, but rather as overt corrections of her speech, which serve no purpose except to highlight her non-standard language use. As Jeantel was the prosecution's own witness, it is doubtful that he intended to discredit her. However, regularly correcting her speech could imply to the jury and any others listening that her speech was uneducated and unintelligible without the help of another to "translate" for them, taking away their side of the communicative burden. If the listener rejects the communicative burden it leaves the speaker, in this case the star witness in the court case, unheard (Lippi-Green, 2012).

In addition to the apparent faults found in her phonetics and syntax, Jeantel was also accused of using racial slurs. In one instance she labelled Caucasians as "crackers," in reference to how Martin had called Zimmerman a "creepy-ass cracker" when speaking about him over the phone to Jeantel on the night of the incident (Thompson, 2015, p. 338). Both "ass" (used here as an intensifier) and "cracker" (a term which can be used to describe anyone who appears ethnically white) are known features of AAE, and the use of them have been defended by linguists as common features of a non-standard dialect (Rickford & King, 2016; Sullivan, 2016). Yet this phrase and other AAE variants were reacted to quite poorly by those present in the courtroom, including the jury (Rickford & King, 2016). In the media, the use of AAE variants such as "cracker" in the courtroom was treated as highly unprofessional and portrayed in some quarters as if Jeantel meant to purposely disrespect the proceedings (Carlin, 2016; Thompson, 2015). Beyond her use of a non-standard, stigmatised dialect, Jeantel was also belittled for her ostensibly low literacy level by the defence lawyers and the media (Carodine, 2015). These attacks on her speech presented her as unintelligent, unintelligible, and difficult to relate to. Characteristics such as these could have caused her testimony to be viewed as unreliable.

Jeantel's language use was overtly corrected, ridiculed, and at times even censored when the court felt she had used racially charged language (Carlin, 2016; Sullivan, 2016; Thompson, 2015). In this way it was made clear that the court and those involved found it unacceptable for Jeantel to express herself in the way most natural and comfortable to her. Rather, it appears that she was expected to conform and speak the standard dialect, something not familiar to her, and received poor treatment as a result of her inability to fulfil those expectations. This treatment is known as accent discrimination.

Accent discrimination, occasionally called *accentism* by some linguists or *linguistic discrimination* (shortened to *linguicism*) by others (Beal, 2006; Howley, 2018), occurs when someone is treated in a biased manner (whether positively or negatively) due to their manner of speech. This can be found in employment, education, housing, and the legal system (Del Valle, 2003; Lippi-Green, 1994; Massey & Lundy, 2001; Wolfram, 1998). Jeantel's experience falls into this category as the overt corrections of her speech in instances such as the –ING variable, where no clarification was needed to improve understanding, is a sign of the biased treatment she received. Though there is evidence found in the lives of Jeantel and others (outlined in chapters 2 and 3) of the presence of accent discrimination, research is still seeking to discover what factor(s) triggers this and how best to protect speakers against it.

In order to prevent discrimination, the underlying cause of it must first be identified. In researching this, psychologists have found that discrimination often occurs where there is an "us v. them" mentality. In this way discrimination is used to bring unity at the cost of directly excluding others. It binds people together over a common "enemy" (Cesario & Navarrete, 2013). For example, without Africans in the United States, it would have been a nation made up of lower class immigrants with nothing more in common than their desire for a new life. Everyone would have been Irish, Polish, Scandinavian, etc., but the addition of Africans gave them the more "impressive" title of white and equal (Milroy, 2000). Similarly, when it comes to language, accent discrimination has also often been found where there is that same "us v. them" ideology (Tamasi & Antieau, 2015). During periods of time when the vast majority of people tended to remain settled in the same areas as their ancestors, neighbouring communities would have been more likely to differ by accent, rather

than race. From then until now, language has been a reliable marker that differentiates between communities, clearly showing who is a part of the "in-crowd" and who is an "outsider" (Collins & Clement, 2012).

It has become common to use linguistic proficiency or variation to reaffirm and justify discrimination. Such actions often rely on stereotypes regarding the speakers when making judgments about them or the belief that non-standard dialects are all grammatically incorrect. However, research has found that many of those stereotypes are overgeneralisations and therefore unreliable (Tamasi & Antieau, 2015). Equally, it is universally agreed by linguists that non-standard dialects are just as grammatically "correct" as the standard dialect (Lippi-Green, 2012; Milroy & Milroy, 2012; Rickford & King, 2016). Despite this evidence, the cases listed in chapters 2 and 3 of this dissertation suggest that numerous people may still be facing discrimination due to their accent (Dixon, Mahoney, & Cocks, 2002; Frumkin, 2007; Wiehl, 2002).

The above issues lay at the heart of this dissertation. While accent discrimination may not be widespread, there is clear evidence in reported experiences and in research to suggest that it is not a random and rare occurrence. However, it is not known whether particular language attitudes alone must be present to cause discriminatory behaviour, or if additional factors must also be involved. This dissertation seeks to uncover further nuances of accent discrimination in order to learn more about what variables may trigger it and in which situations it may be more likely to occur. The specific focus on this research is language within the courtroom and what expectations a jury may have with regards to the standardness of the language used.

1.1. Dissertation Overview

The main question this dissertation strives to answer is *whether the presence of particular language attitudes, as defined and discussed in chapter 2, can lead directly to accent discrimination*. Specifically, the research conducted for this study looks at whether the language attitudes held towards defendants' accents in a courtroom lead a jury to discriminate against them. Four specific research questions (RQs) were identified. These RQs are listed below.

Research Question I: Will a defendant be perceived differently by a jury based on whether he speaks in a standard or regional accent?

Research Question II: If a defendant is perceived differently, will this go as far as affecting the verdict?

Research Question III: Does collecting individual responses lead to reliable results on how a jury would function or does the group condition make a large enough difference to invalidate research results that used only individual participation?

Research Question IV: Are the results of the first two research questions similar enough across western nations, such as the United Kingdom and the United States, as to be valid outside of the country the study was conducted in?

RQs I and II were created in order to help answer the main question, and therefore are the main focus. Moreover, due to the flaws in earlier research outlined in detail in chapter 3, RQs III and IV were added to investigate the prioritised flaws in further depth, and to assess their potential impact. While the United States and the data collected from there are the main discussion points, information about and data from the United Kingdom are also included in this dissertation in order to address RQ IV. These RQs are discussed in detail in chapter 4.

This dissertation is structured as an hour glass. The discussion around accent discrimination begins broadly, looking at the various sectors it can be found in, and gradually narrows down to look specifically at the potential for discrimination against a male defendant due to the accent he spoke in a courtroom. At the end the discussion broadens again, as the impact for this type of research is noted and recommendations for future research are given. A brief outline of each chapter now follows.

Chapter 2 examines language ideologies and attitudes. It begins by considering Standard Language Ideology and how it may have helped legitimise any negative perceptions people hold towards non-standard dialects. From there it moves on to note cases when people have either publicly complained about experiencing accent discrimination or publicly discriminated against language varieties

themselves. The final section in chapter 2 looks at academic studies carried out in housing, employment, and education and the contexts in which they found accent discrimination to be present.

Chapter 3 narrows the discussion into accent discrimination and looks at it with regards to occurrences in the courtroom. It begins by revisiting *Zimmerman v. State*, and the treatment Rachel Jeantel received, before then moving on to other court cases where there is clear evidence of discrimination. Some recommendations are given on ways to set an anti-accent discrimination precedent in the courts based on current U.S. laws. The chapter concludes by examining the known academic studies conducted on accent discrimination in the courts, noting their findings, their strengths, their weaknesses, and any other relevant issues left in the research.

Following these two literature review chapters, chapter 4 notes again the issues left by previous research and how the research undertaken for this dissertation seeks to fill some of them. The pilot studies run for this research are outlined, their results helping to shape the methodologies used in the later full studies described in chapters 5 and 6. The RQs are identified again here and the hypotheses for each of them are laid out. The following three chapters are studies that sought to answer those RQs.

Chapter 5 describes the United Kingdom version of the study. It begins with a brief literature review on the two accents used, Standard Southern British English (SSBE) and Yorkshire English (YE), before discussing in depth the methodology used. Chapter 5 sought to answer RQ I and II, regarding to what extent a defendant might be perceived negatively by a jury. This was done by creating two short video clips of a cross-examination between a prosecutor and a defendant. Half the online individual participants were randomly given the video with the defendant testifying in SSBE and the other half had the video with the defendant testifying in YE. Participants were then asked to render a verdict of guilty or not guilty, and to answer a series of questions regarding their perception of the defendant. The results are analysed, using specific analysis questions in order to discover if a specific video guise had a significant effect on participant perceptions or verdicts. A discussion is then offered on the results and any methodological weaknesses are critiqued.

Chapter 6 uses a similar methodology to that which was described in chapter 5, but with two main differences: firstly, it was run in the United States, and therefore used GAE and Southern American English as its two accents; secondly, while a similar online version was run with individual participants, it was also run with participants in groups as mock juries in order to test the effect group dynamics may have on a jury. In this way chapter 6 sought to answer RQs I, II, and III. Once again the results are analysed and discussed, including a qualitative analysis of the group deliberations, which were recorded and transcribed. Differences between the individual and mock jury datasets are examined in order to discover whether studies using individual participants as jurors can be considered ecologically valid.

Finally, chapter 7 focuses solely on RQ IV, and whether data and studies from different countries can reliably be compared against one another. Data from the online U.K. participant base described in chapter 5 and data from the online U.S. participant base are cross-compared and analysed. A discussion is offered on these results, and a further note for caution offered regarding the potential for nuances involved in differing cultures' language attitudes.

Chapter 8 concludes the dissertation. It outlines the impact studies like these can have, summarises the main findings of each of the above chapters, once again mentions the weaknesses and gaps left in the research, and offers suggestions for future research. While this dissertation is in no way an exhaustive analysis of accent discrimination, through these eight chapters it seeks to join in the current conversation on the topic and add to the growing body of knowledge.

1.2. Terms Defined

Before this subject is explored further, it is necessary to define and explain the usage of specific terms used throughout this dissertation. Further definitions will be offered in later chapters as well. However, those will generally be relevant to the specific chapter topic whereas the terms defined in this section are fundamental to the overall dissertation.

To begin with, *accent* and *dialect* must be defined, as there are different understandings of these terms. *Accent* refers to the phonetic and phonological patterns belonging to a community of people. This includes pronunciation,

intonation, stress patterns, and pitch (Kerswill, 2007). *Dialect* refers to distinctive semantic and syntactic speech patterns as well as the phonetic and phonological patterns within a speech community (Ghorshi, Vaseghi, & Yan, 2008). Because of the difficulty that comes with deciding when an accent has enough unique characteristics to be considered a dialect, these terms have often been used interchangeably, particularly by the general public (Hughes, Trudgill & Watt, 2012).

For many people, the terms *accent* and *dialect* do not mean anything more than how someone who is *other* and *different* speaks. In fact, many people believe they themselves do not even have an accent (Morley, 1996), something linguists refer to as "the non-accent myth" (Lippi-Green, 2012). The word *accent* itself was originally a Latin translation of the Greek word "prosōidía," (Vaissiere & Boula De Mareuil, 2004) which was used to differentiate between people who spoke "correctly" and those who did not (Oxford English Dictionary, 2018c). This seems to imply that approaching accents as something that is *other* is not a new view and has in fact been the prominent language ideology for millennia.

Yet from a linguist's perspective everyone has an accent, and this accent is an integral part of a person's identity. Because of this, it must be understood that rejecting the way someone speaks can be perceived by the speaker as rejecting them (Nguyen, 1993). It is normal to detect when someone differs from oneself, and noticing this difference does not always lead to discrimination. However, there are a number of situations, some of them outlined in this dissertation, where it is clear speakers were discriminated against, such as with Rachel Jeantel. In the cases where noted differences do cause discrimination, the reasoning may be rooted in any number of stereotypes, such as the speakers being perceived as ignorant, lower class, or more likely to commit a crime (Lippi-Green, 2012; Tamasi & Antieau, 2015). Some studies have observed that occasionally even one non-standard feature can change how the speaker is viewed (Campbell-Kibler, 2006; Wiehl, 2002).

Another term that needs to be properly understood is *standard*. Dialects and accents are often spoken about in terms regarding which are *standard* and which are *non-standard*. This will be discussed further in chapter 2, but when the term *standard* is used throughout this dissertation, it is done meaning a variety that people seek to use uniformly, regardless of which region the speaker is from. Primarily

associated with education and writing, the standard dialect is the *uniform* manner to write in, following the accepted conventions in use of grammar, syntax, and lexicon when preparing an academic paper (Leith, 1997). Although the standard is also generally perceived as being spoken uniformly, linguists have observed that there are still degrees of variation in the accents used (Milroy, 2000). This reveals an inconsistency between the actual patterns of usage (in its variation) and the attitudes towards the usage (perceived as being perfectly uniform) (Mugglestone, 2003). *Non-standard* simply refers to any other variety of language outside of the accepted and more prestigious standard dialect. These non-standard dialects are all still grammatical and fully functioning dialects in their own rights, and therefore in no way *sub*-standard, but they differ to some degree from the standard dialect.

The standard accents that will be the focus in this dissertation are General American English (GAE) from the United States and Standard Southern British English (SSBE) from the United Kingdom. The non-standard accents used are Southern American English and Yorkshire English (YE). While all four of these dialects do have distinctive syntactic and lexical features, only the phonetics features of each were present in the stimuli. Thus, as per the definition, my studies compared *accent* varieties rather than dialectal differences. GAE is more commonly known as Standard American English (SAE). However, this dissertation uses the term GAE (also found in Tamasi & Antieau, 2015), in order to prevent any potential confusion when also discussing Southern American English, as this avoids identical acronyms. GAE also does away with the term *standard* in the title, something more appropriate for a nation that does not have a truly uniform variety of language (discussed in chapter 6).

Finally, the concept of *ecological validity* must be understood. This is used with regards to research and how well an experiment mimicked real-world conditions. This is an important consideration, as the more ecologically valid a study is, the more its results can be trusted to occur in the real world outside of carefully controlled "laboratory" conditions (Bernal, Bonilla, & Bellido, 1995). This term will be used throughout the dissertation as the level of ecological validity in previous studies will be considered, as well as in discussions on how the research conducted for this dissertation strove to reach for greater ecological validity.

Chapter 2. Language Attitudes and Ideologies

"Germans are harsh; just listen to their harsh, guttural consonants. US southerners are laid-back and lazy; just listen to their lazy, drawled vowels. Lower-status speakers are unintelligent; they don't even understand that two negatives make a positive" (Preston, 2002a, p.40-41).

Statements such as those reported by Preston above are common and arise from attitudes that are formed by the ideologies people hold regarding language (Piller, 2015). Language ideologies are social constructions that form the way people think about language. They are used as rationalisations or justifications for people's perceptions regarding how languages are structured and used. These ideologies are tied into how people understand the way language interacts with society and therefore informs their attitudes and responses towards language and the assorted varieties (Cameron, 2003; Woolard, 1992). These ideologies are often constructed in such a way as to follow the political and social concerns of specific groups, and the more politically and socially powerful a group is, the more likely their ideology is to spread further and achieve wider acceptance (Milroy & Milroy, 2012). This can be seen occurring in standardisation efforts, a topic discussed later in this chapter.

People's language ideologies are learned through studying their language attitudes, the outward manifestation of the abstract ideologies held (Cargile et al., 1994). Language attitudes "collectively refer to the perceptions, attitudes, and stereotypes associated with language use, linguistic structure, and, oftentimes, with speakers themselves" (Tamasi & Antieau, 2015, p. 45). These attitudes are subject to variation and change within populations and people. They can be positive (e.g. "Southern [U.S.] speakers sound friendly") or negative (e.g. "Southern [U.S.] speakers sound uneducated"). They can be towards: specific linguistic features (e.g.

11

syntactic use of double negatives); individual people (e.g. complaining about voice pitch); linguistic varieties (e.g. the perception of SSBE speakers sounding intelligent); regional areas (e.g. U.S. Midwest or "Valley girl uptalk"); or any combination of the above factors (Lippi-Green, 2012; Milroy & Milroy, 2012; Tamasi & Antieau, 2015).

These language attitudes are collected and observed through attitudinal studies. The methodology for this was formalised by Giles (1970). In looking at earlier studies on the topic, he noted that they seemed to ignore, or be unaware of, how language attitudes can be related to more than just personality traits, such as how *loyal* a person may be towards a specific dialect. He suggested the need for three more "dimensions," as he calls them (p. 212). These dimensions were *aesthetics*, how pleasant someone found the accent to be, *communicative*, the comfort level a listener had in understanding and interacting with the accent, and *status*, which considers the amount of prestige given to the accent. He believed that he would find that each of these dimensions had an effect on how an accent was perceived and that participants would change their ratings dependent on whether they were answering according to the aesthetics, the ease of communication, or the level of status of an accent. Giles combined this with the newly innovated matched-guise method, in order to show that any changes in perception were in fact due to these various dimensions rather than variables such as voice quality.

Giles' (1970) study was run in the United Kingdom and 177 participants took part. He used a 7-point Likert scale, asking how pleasant/unpleasant the accent was, how comfortable/uncomfortable participants would be when interacting with a speaker of that accent, and how prestigious participants felt the accent was. Giles found that RP was rated highest with regards to prestige when compared against other accents, including other national standard dialects, such as GAE. He noted that the rest of the accents included in the study fell on a type of hierarchal continuum, and that all regional accents had significantly lower levels of prestige than RP. RP also had the highest ratings for aesthetics, and was tied with participants' own accents for highest communicative ratings. Giles' concluded that the standard dialect would always carry the most positive perceptions and considered recommendations that non-standard speakers become bi-dialectal. Giles continued to research language attitudes, and consistently found that the standard dialect received the highest level

of prestige ratings (Creber & Giles, 1983; Giles, 1971; Giles 1972; Giles & Coupland, 1991; Giles & Powesland, 1975; Giles et al., 1992).

Following Giles' methods, over the last four decades numerous studies have been conducted looking at language attitudes through the use of the matched-guise method and asking participants to rate accents on specific characteristics using a 7-point Likert scale (e.g. Bayard et. al., 2001; Creber & Giles, 1983; Garrett, 2002; Giles, et al., 1992; Heaton & Nygaard, 2011; Preston, 2002b; etc.). As numerous characteristic types began to be used, for comparability and simplification purposes they were sorted into either a *status* category trait (e.g. level of perceived intelligence, success, social class, hardworking, education, etc.) or a *solidarity* category trait (e.g. level of perceived trustworthiness, warmth, friendliness, formality, loyalty, cooperativeness, helpfulness, etc.). Across the board these researchers found that the standard dialect tended to receive higher ratings with regards to any *status* characteristics while any non-standard dialect studied was more likely to receive higher *solidarity* ratings with low *status* ratings. This is discussed further in sections 5.2 and 6.2.

Giles' method for studying language attitudes has not only been widely accepted in general perception studies, but has also been utilised when examining perceptions of specific traits in particular circumstances. In looking at hiring decision processes and the potential for discrimination in employment, Cargile (2000) conducted a study looking at how *employable* people were rated to be based on their accent. In considering the legal realm and the testimony of defendants, Seggie (1983), Dixon et al. (2002), and Dixon and Mahoney (2004) all adapted Giles' method to obtain ratings of perceived guilt based on the accent in which the defendant testified. Similarly, Frumkin (2007) ran a study looking at how a witness's testimony would be perceived based on the accent, with regards to characteristic traits such as credibility, accuracy, and deceptiveness. These will all be discussed in greater depth in later sections.

Building upon the studies that researched specific language attitudes, current research is now seeking to discover to what extent these language attitudes may predict how people will behave towards other speakers (Ladegaard, 2000). It has been observed that while language attitudes exist towards all linguistic variables, the

attitudes, and even ideologies themselves, are directly tied to the stereotypes about the *speakers* of those varieties (Cheshire & Milroy, 2013; Preston & Robinson, 2005). In this context, if someone says that the use of double negatives makes a speaker sound stupid, the question that needs to be answered is whether that attitude will then cause the person to act in a discriminatory way towards that speaker.

Although many may recognise these language attitudes as the stereotypes they are, this does not prevent people from acting on them. Psychology research has defined stereotypes as "a type of mental shortcut we rely on to obtain information quickly and effortlessly" (Khan, Benda, & Stagnaro, 2012, p. 3). They are overgeneralizations that allow a person to process data quickly, particularly when someone only has a small amount of information to rely on (Hinton, 2000; Ryan, 2003). Psychologists have found that there are generally some elements of truth in stereotypes, although that truth does not need to be current or representative of the entire population (Ryan, Park & Judd, 1996). It has been observed that people are often more willing to accept the stereotypes than to search for a more nuanced and complex truth (Kristiansen, 2001). In the case of language, studies have found that in some instances, the "mere presence of a regional accent influences stereotype activation and possible discrimination against the speakers" (Rakić et al., 2011, p. 12). Accents can activate stereotypes, and these assumptions and stereotypes open the potential for discrimination to occur (Kristiansen, 2001; Wiehl, 2002).

In more technical terms, language based stereotypes are understood as one type of *indexicality*. This term is used to describe anything that helps create a social identity (language, fashion choice, hobbies, etc.) and allows for micro-categories in the analyses of language variation, along with the traditional macro-categories of age, gender, class, etc. (Eckert, 2008; Johnstone, Andrus & Danielson, 2006). Silverstein (2003) developed three orders of indexicality to analyse linguistic variation and the attitudes associated with them in more depth. The first order of indexicality considers the variable an "indicator," one which carries social meaning but not stylistic variation as speakers are unaware of that variable. The example offered here are languages that have different honorific terms for the *addressee* and the *referent*. Here there is variation in whether a speaker chooses to use an honorific, but the environment dictates whether the exact term used is with regards to the person the speaker is addressing or a third party, nothing more. The second order of

indexicality is known as a "marker," as it is a variable with both social and stylistic variation, and speakers are aware of it. One such "marker" is the –ING variable with the alveolar nasal [n] and the velar nasal [ŋ] as the possible social and/or stylistic variants. Finally, the third order deals with the notion of "stereotypes." These are variables that speakers are not only aware of, but are a topic of social discussion, to the extent that it can result in codification (e.g. Bostonian "Pahk the cah in Havahd Yahd). It is often these second or third orders of indexicality that are elicited through overt language attitude studies when participants are asked to report how attractive, how educated, how intelligent, etc. they find certain language varieties. As a result, it is these (and in particular the third order "stereotype") that will be discussed here and are the main areas of interest.

When it comes to these indexical stereotypes and language attitudes, it may make little difference whether a person's speech contains specific dialectal features or whether those dialectal features accurately and inherently reflect specific characteristics. If people are acting on them as if they are true, then these stereotypes are worth studying. With stereotypes, whether positive or negative, comes the potential for prejudice. However, this prejudice does not automatically cause discrimination. The Oxford English Dictionary defines prejudice as a "preconceived *opinion* not based on reason or actual experience" (2018b; emphasis added) whereas it defines discrimination as an *action*, the *treatment* of someone either more negatively or positively due to prejudice (2018a). The presence of prejudicial opinions does not mean discriminatory treatment has also occurred. The question thus becomes one of whether these linguistic stereotypes are known but recognised to be false, or do people *act* upon these prejudices and treat others differently depending on how they speak?

This dissertation seeks to address the question of whether people act on their linguistic prejudices. The present chapter begins the analysis on a broad scale. Here I will first examine the arguments behind Standard Language Ideology and how it may legitimise negative perceptions of non-standard dialects. This is followed by a discussion on media reports and comments left by readers regarding language and discrimination, looking at individual testimonies and the personal opinions stated on social media in order to address this question. Finally, the chapter will conclude with

an exploration of academic research on accent discrimination across multiple areas such as in housing, employment, and education.

2.1. The Standardisation Ideal

The concept of a standard dialect in Western nations such as the U.K. and the U.S. is attributed by many linguists as the main way people justify accent discrimination (Lippi-Green, 2012; Tamasi & Antieau, 2015; Wiley & Lukes, 1996). Commonly referred to as Standard Language Ideology (SLI), this proposal suggests that there should be one language variety that everyone speaks, at the least within formal or professional settings. Some take a stronger stance, arguing that the standard should be used in every setting, citing a "one nation, one language" rhetoric (Kibbee, 1998; Ricento & Burnaby, 2013; Vogl & Hüning, 2010). SLI reflects a prescriptive view of language and would seek to limit the variety of accents and dialects which differ by region within a nation. Some suggest that by maintaining a standard dialect, communication would be better maintained between regions. Establishing a standard dialect would foster a greater sense of unity and lower the potential for misunderstandings. Within the U.K., this argument dates back to at least the 15th century when communication between villages could be a struggle due to the high level of variation in accents and dialects (Crowley, 1991; Milroy & Milroy, 2012; Watts, 2000). The notion of a prestigious language dates even further back, as French and Latin were perceived as the languages of the educated man in England in the 14th century and earlier, suggesting that language ideologies and attitudes are nothing new (Mugglestone, 2003).

Numerous arguments have been presented regarding SLI, outlining some of the issues inherent with it. To begin with, which dialect should be chosen and promoted as the standard and the one everyone should acquire? Once a dialect is selected, how is the choice of that specific dialect justified? Even after a dialect is chosen, regardless of the justification used, the label "standard dialect" or "Standard English" is not well-defined or understood. Sociolinguists cannot fully agree with one another regarding what Standard English is or what that label means (Kerswill, 2007). Trudgill argues that it is easier to define what Standard English is by defining what it is not, and yet even here he is not in full agreement with other linguists

(Coupland, 2000; Trudgill, 1999). Milroy and Milroy (2012) advise caution when working with this term, as Standard English "is a rather loose and pre-scientific label. What Standard English actually is thought to be depends on acceptance … of a common core of linguistic conventions, and a good deal of fuzziness remains around the edges" (p. 22). They note that standardisation is a process, as the standard dialect is constantly changing with its speakers. Lippi-Green (2012) chose to address this issue by adding an asterisk before a dialect she referred to as "standard," following the precedent set in the academic field of syntax, where an asterisk is placed in front of a sentence that is deemed ungrammatical. She argues that because the standard is an abstract and often fluid concept, to call one specific set of linguistic features the "standard," when other variations may occur and still be considered "standard" by the speakers, is just an untenable as an ungrammatical sentence (Lippi-Green, 2012). Yet despite these issues, dialects are still spoken about in terms of being "standard" or "non-standard."

Because the concept of a standard was never fully defined, the standards within the U.S. and U.K. are an ideal, although SSBE linguistic features are generally regarded as more uniform and better understood than GAE (Fabricius, 2018; Tamasi & Lamont, 2015). It can be argued that there has been some success within individual nations in establishing the use of a single form when writing in formal contexts. Yet due to the constant changing nature of language and its flexibility, no one person speaks the standard dialect just like another. Therefore the goal of everyone using an identical form of speech is unrealistic and unattainable, though the question of whether it was ever actually possible does not seem to have been raised (Lippi-Green, 2012). Instead it remains solely an ideal that fosters misunderstandings regarding the way language functions. Rather than understanding "standard" to mean a perceived *uniform* variety, it has come to mean something more along the lines of *high standards*, in that it is the ideal form of language (Cheshire & Milroy, 2013). This then creates a bias toward an idealised, and often unrealistic, form of language. This same bias has the potential to cause discrimination and may encourage suppression of other forms of language variation (Wiley & Lukes, 1996).

It must be noted that while SLI and standardisation efforts occur in respect of many language and in numerous nations, the exact motivations and driving forces

17

behind the push for SLI varies by area. For instance, France has a long established standard ideology. Considered by some to have one of the most standardised languages, France established its *Académie Française* in 1635 with the specific goal of monitoring language use and deciding what the "correct" way of speaking was (Lodge, 1993). In contrast, while Lithuania has SLI, it is not nearly as strong or as prevalent as France's ideology. In Lithuania it is seen as common sense to use the standard in any formal or official settings; outside of those contexts, however, use of other language varieties is generally acceptable (Nugaraitė, 2017).

Although the U.K. and the U.S. speak the same language, the motivations for and outcomes of SLI also differ between the two nations, leading to different perceptions of the standard variety used in each country by each respective country's citizens. It then follows that their citizens may face discrimination for different reasons (Coupland, 2000). For example there has never been a debate in the U.K. over whether British Black English should be taught in schools, whereas the U.S. has spent a great deal of time arguing over the validity of AAE in the classroom (Milroy & Milroy, 2012; Wolfram, 1998). It is generally agreed that standardisation and discrimination focuses on social class groups in the U.K. while in the U.S. it becomes a racial issue (Coupland, 2000). Therefore, it is important to mention that despite surface similarities, these are two different countries with diverse cultures that may have separate motivations when it comes to language. This will be expanded upon in chapters 5 and 6 when specific dialects are described and the language attitudes held towards them are outlined. In this chapter, SLI is discussed in more general terms as it relates to both the U.K. and the U.S.

In both countries there have been cases where those who did not speak the chosen dialect or language were expected to acquire it. This outcome is clearly seen in some of the more extreme cases where people were forced to learn another language variety. One such example is when the English conquered the Irish and for generations forced them to speak English, justifying it under the same "one nation, one language" rhetoric mentioned earlier (Neill, 1994). However, in most cases the expectation for people to acquire the standard is more subtle than overt government force. Instead it can be seen in the number of "accent reduction" classes, the encouragement teachers give their students to learn the standard, and in some employers' expectations that their employees will use the standard in order to be

hired or promoted (Freeman, 1998; Garrett, 2010; Lippi-Green, 1994; Tamasi & Antieau, 2015). In some cases, to not acquire the standard has been perceived as a sign of ignorance and laziness (Lippi-Green, 2012).

Non-standard dialects have become mistakenly labelled as ungrammatical; in reality they have simply become socially unacceptable (Cargile, 2000). Because the standard dialect is based on the dialect spoken by those with influence and in positions of authority, non-standard dialects have also come to be thought of as the variety spoken by the lower class or only suitable in informal situations (Watts, 2000). As noted in chapter 1, people have often viewed anything under the term *accent* to be *other*. This view is encouraged by SLI, as it is perceived to justify the idea that non-standard dialects and accent are "something less than a language" (Huebner, 1999, p. 1). While the use of a standard dialect is seen as the only "correct" way to communicate (Tamasi & Antieau, 2015), other dialects are understood to be not simply *non-standard*, but rather *sub-standard* (Cheshire & Milroy, 2013), once again showing a misunderstanding of the meaning intended by the term "standard."

This negative attitude held towards non-standard varieties is found not only among those who strive to speak the standard, but also in the speakers of non-standard varieties. Studies have found that some speakers of non-standard dialects may have just as poor an opinion, or occasionally more so, of non-standard varieties than those who speak the standard dialect (Collins & Clement, 2012). This is a trend that has been observed by psychologists in numerous environments, where the out-group favours aspects from the in-group more highly (Jost, Burgess, & Mosso, 2001). With regard to language this could be due to non-standard speakers being more aware of the repercussions and discriminations that come with speaking a variety that is not widely known or accepted. It follows that they would not want their children facing the same issues. Therefore they may encourage an abandonment of the local dialect and a move toward the standard (Frumkin, 2007). Yet this attitude toward their own speech unwittingly strengthens the negative perceptions held toward non-standard dialects and encourages positive discrimination towards speakers of the standard dialect (Collins & Clement, 2012). While people have been complaining about differing varieties of speech for millennia, SLI gives these complaints a legitimacy they previously lacked (Finegan & Rickford, 2004).

SLI has also furthered the non-accent myth; in many cases it is only those who were born into a stigmatized non-standard dialect speaking community who are seen as having an accent. Using the standard is seen as the "norm" while speaking anything else is *other* (Collins & Clement, 2012). When it is a dialect that has positive stereotypes, such as how Americans view U.K. dialects, where the *other* is perceived as something exotic and interesting, positive discrimination is more likely to occur. However, when it comes to stigmatized dialects, whether regional or foreign-accented, the *other* becomes something negative and to be avoided. It is in this context that negative discrimination can be found. An example of this occurs in general communication. In order to effectively communicate, both parties need to be willing to work together and bear the "communicative burden," not simply the speaker (Lippi-Green, 2012, p. 73). The listener must be willing to hear and understand, rather than judge. Yet this is not always the case. In situations where the speaker is deemed to speak with an "accent," generally meaning a stigmatized non-standard variety, the listener may reject bearing their side of the burden and the speaker is blamed and labelled with a "thick accent" (Morley, 1996; Wiley & Lukes, 1996).

Overall, SLI has contributed to over-simplified views on how languages function and may have encouraged the propagation of stereotypes regarding how all other dialects are perceived and then reacted to (Rakić et al., 2011). In some instances the standard dialect has been set up as the route to success, placing other dialects in a lesser light by contrast. Many regional dialects are perceived as having low social class status and as inappropriate for any type of formal context (Milroy & Milroy, 2012). And when people ignore all of this and choose to continue to speak in their regional dialect, they may be ignored or told their accent is simply too "thick" to be understood. People form their ideologies regarding language around the concept of a standard dialect and many may use it as an excuse for any prejudices they hold toward non-standard dialects and the communities who speak them (Collins & Clement, 2012). Stereotypes held towards certain communities may be strengthened by the use of a non-standard variety of language, and eventually these stereotypes become linked to the varieties themselves. It is often these stereotypes that form and inform language attitudes towards particular dialects, and may be the cause of discrimination (Preston, 2002b).

2.2. Discrimination Reported in the Media

While the term "accent discrimination" may not be well known, the concept of it is still widely discussed across social and news media when its consequences are felt. This section will explore various news articles, tweets, and Facebook comments where the topic of language variation is raised, and discuss how the presence of language ideologies regarding standardisation create language attitudes towards specific dialects. This offers a glance into real life situations where accent discrimination was reported, whether by a victim of it or by someone publicly engaging in acts of discrimination themselves. The following examples come from North America and the U.K., looking at discrimination due to regional accents, a minority language, and foreign accented English. This anecdotal evidence is included in order to offer greater ecological validity to any conclusions drawn regarding how likely accent discrimination is, and the contexts it occurs in.

Within the U.S., Pelly (2018) details her experience of being a native speaker of Southern American English and her internal debate over whether to encourage her young son to acquire GAE instead of also using Southern American English. She mentioned how, after giving her first professional presentation, she was disheartened to receive feedback from her supervisors about her accent, with nothing said about the actual content of her talk. Due to that experience and others where it was made clear to her that her accent was viewed as an object of amusement and entertainment at best, a marker of poor education at worst despite having a master's degree, she chose to develop a work persona. She spoke something closer to GAE at work while at home maintaining her Southern accent. However, she had begun to question keeping any aspect of her Southern accent after hearing her 3 year old son being teased by other children for his own already notable Southern accent.

Pelly (2018) was posted to Facebook via the New York Times page. Numerous people commented on it, many in support of Pelly keeping her own accent and encouraging her son's accent, to be proud of where they come from. However, just as many comments were in regards to people's own efforts to "weaken" their non-standard accents in order to gain more community acceptance or be considered for promotions. They felt it was necessary for one reason or another, one person

saying it was "so people would know which side [he] was on;" another person commenting how now that she has "relearned how to speak" she realises "how very stupid [her] family sounds." This can be seen in the top two comments in Figure 2 below.

The south has proven itself backward in many ways. A southern accent implies education in an inferior school, a strong racial bias and uninformed voting tendencies. Of course not all of this is true for all southerners, but most people form initial impressions from very limited evidence since there is little time to do otherwise. Early in my life, I was working in desegregation of schools in the south. I worked hard to rid myself of my Texas accent so people would know which side I was on. I regret that I felt that I had to do that but I have never regretted doing it.

Like · Reply · 9w   240

Yes, you should have carried your office voice home. I grew up in the NC mountains. When I moved to Raleigh, people made fun of my accent. I relearned how to speak and so did my sister that also had moved to Raleigh. Now I go back and I am unnerved at how very stupid my family sounds. I do understand that they would be treated as if they had no education if they left that area.

Like · Reply · 9w   1

I don't think the accent is a problem as long as the grammar is correct and the Southerner doesn't pepper his/her speech with local colloquialisms.

Like · Reply · 9w   73

the only ones thinking a southern drawl reflects a lower intelligence are liberals from New York and surrounding areas. It is best not to associate with those bigots.

Like · Reply · 9w   67

Figure 2. Facebook Comments on Pelly Article (The New York Times, 2018)

Other comments were similar to the third comment in Figure 2, remarking on grammaticality and the use of "colloquialisms," meaning that while it may be fine for the phonetics to be non-standard, every other aspect should reflect Standard English guidelines. The final type of comment noted was in line with the last

22

comment in Figure 2. These people reacted strongly to the idea of being discriminated against due to their speech or regional cultural heritage and responded by negatively stereotyping other regions (The New York Times, 2018). In all four cases third order indexical language attitudes were clearly evident.

Language attitudes are present within the U.K. as well, and there are numerous articles that discuss the potential for negative consequences of those attitudes. Howley (2018) discusses the British reality TV show *Love Island* in its 2018 season. She observes how the contestants who receive the worst criticism appear to be those who speak a dialect that is neither standard nor a non-standard one associated with being "funny" or "friendly." Hayley, one of the contestants, has a Liverpool accent and comments from viewers on twitter have regularly derided her for sounding "annoying," "unintelligent," "fake," and "painful to listen to." The irony in some of these comments is that the grammar and punctuation they used is not always correct, yet they are ridiculing another person for having an "unacceptable" accent. In addition, some of the readers of this article left comments, shown in Figure 3, suggesting they did not disagree with the derision Hayley faced due to her speech. This was in spite of Howley's argument that responses such as these are at best prejudice, and at worst lead to discrimination.

John B Ellis   1 month ago

Scousers aren't 'proper northerners'!

Reply · Flag   -1 likes 👍 👎

John B Ellis   1 month ago

Accent's one thing, coherence is quite another ...

Reply · Flag   3 likes 👍 👎

Snowdriftsurfer   1 month ago

Sure language evolves, but I've spoken with people who have said," it's
like, like you know what I mean " in all seriousness and it's frustrating
because I don't know what they mean. Save that for talking to your mates
or being on Love Island, not when you're in a business setting. We all
know people have enough stereotypes to fend off without sounding as if
you can't express yourself accent or not.

Reply · Flag   -1 likes 👍 👎

Figure 3. Reader Comments on Howley Article (Howley, 2018)

Other examples of this type of discrimination come from personal
experiences that are directly reported to the media. Writing in the Telegraph, Katie
Edwards, who holds a PhD in Linguistics, describes her experience with being
regularly mocked and ridiculed for her regional accent (Edwards, 2014). Although
she did not specify which non-standard accent she has, she says she has frequently
received recommendations to "tone down" her accent in order to be taken more
seriously; this is despite clearly having a high level of education and working for a
prestigious university. She mentions interviews she conducted, where every woman
she spoke to who did not grow up speaking the standard dialect stated that they had,
to some extent, felt the need to moderate their regional accent. In some cases this
went as far as creating two personas: a work one where they spoke more standardly
and a home one where they spoke in the accent they grew up with, similar to Pelly
(2018). This was due to embarrassing encounters, or even outright bullying, as they
described it. However, many of the men Edwards interviewed said they had
consciously strengthened their regional accents, though one did attribute that as the
reason he was not promoted to a more senior position. The nature of these interviews
was not clear, nor did she describe how robust her data collection was. Therefore,
this may be anecdotal evidence; yet it still offers further evidence of who has felt the

24

need to adjust their speech. The Economist later picked up Edwards' story, and noted how in one of the instances of accent discrimination she describes, it is another female academic discriminating against her, suggesting that while women may receive more discrimination due to their speech, they may be just as complicit in it as men (R.L.G., 2015).

Novelist and broadcaster Dreda Say Mitchell (2017) writes that while she has received many complaints over the years, none have remained as consistent and constant as the complaints about her accent. A native of London, she regularly receives emails and tweets asking why she "drops her 'G's," (i.e. pronouncing the alveolar nasal [n] variant rather than the standard velar nasal [ŋ] for the –ING variable). Others ask why she does not pronounce "the 'th' sound," (instead fronting that to [f] or [v]). These types of non-standard phonetic features rarely inhibit intelligibility. Therefore these complaints most likely stem from people who hold to SLI rather than any genuine desire to understand her broadcasts better. Mitchell notes that she is not alone among broadcasters who receive these types of complaints for not being more "standard" and therefore "correct." One such example is Steph McGovern. A BBC morning show presenter and native of Middlesbrough, she also reported receiving derogatory comments regarding her accent, in one instance being offered £20 to help *fix* her "terrible affliction" (Webber, 2014).

Similarly, the Labour MP and Shadow Education Secretary Angela Rayner, a native of Stockport, has regularly received emails over the course of her tenure as an MP deriding her for not learning to speak "properly," as befits someone of her position. Two such emails can be viewed in Figures 4 and 5.

By improving her English grammar and spoken English, she might have more credibility as Shadow Education Secretary. I cringed a few minutes ago when, on Breakfast Television (BBC) she said, "There is issues…………." oops, plural subject and singular verb….. should have been "There are issues……" Seconds later she was dropping Hs…

I am appalled.

Figure 4. Email 1 to Angel Rayner (Gill, 2016)

Jo Phillips @joglasg

One of the emails Angela Rayner was sent about her accent - can you imagine sending that with your details? Jeeeesus. A teacher sent that? 😤

I am appalled having watched you on question time. If you wanted to represent Northerners as 'thick individuals', then well done you have succeeded. I work in a school and I would not be able to appoint you as you cannot speak Standard English. We have 26 letters in the alphabet and 44 phonemes in the English Language. I suggest you learn them, particularly the letter 'T'. Look at YOUR constituency rather than the labour line. I will happily give you my home address and would welcome a one to one discussion with you. Please contact me directly on .... and I would love to discuss this further with you. I look forward to hearing from you.

An embarrassed constituent.

7:20 am - 25 Jan 2017 From Glasgow, Scotland

Figure 5. Email 2 to Angel Rayner (Phillips, 2017)

Rayner immediately publicly reported receiving this abuse and, unlike the case of Hayley in Love Island where people felt the ridicule deserved, many tweeted in support of Rayner. Figure 5 shows one such case, as the tweet notes anger at the email content attached. Rayner responded to these criticisms by saying that she is proud of the way she speaks as it is representative of both her background and the background of her constituency (BBC News, 2017; Gill, 2016).

Through the above examples it is clear that this discrimination is not limited to one profession or background as the women all have distinctly different jobs from one another. The one common denominator in these stories is that it is all women reporting this discrimination. There is further evidence that the stories listed above are not rare cases when it comes to accent discrimination. Good Morning Britain and the ITV Tonight show, both British shows on news channels, conducted surveys and found that roughly 1 in 4 people across the U.K. feel they have been discriminated

26

against due to their accent. In later interviews it was noted that even children were aware of this phenomenon, and many respondents believed non-standard speakers would need to change their accent in order to be more successful (Itv.com, 2018; Marshall, 2013). However, accent discrimination does not only appear in cases regarding non-standard or regional accents, but also in regards to minority languages and foreign-accented English speakers.

Welsh is a minority language spoken in parts of Wales and attempts are being made to revitalise it, as well as Welsh culture more generally (Coupland & Aldridge, 2009). Freelance news and sportswriter Marcus Stead makes his opinion clear on his twitter account with regards to the attempts to preserve the Welsh language within Wales. Figures 6 and 7 contain only three of many tweets Stead has written deriding the use of Welsh in any context.

Figure 6. Stead Tweet – Welsh Tourism (Stead, 2018b)

27

Marcus Stead @MarcusStead · Aug 5

A lot of noise is coming from the Eisteddfod site near my flat. I can't work out whether the people 'on the mice' are speaking Welsh, or are just very drunk. Probably both. We've got a week of this parochial bore-fest to run. #Eisteddfod2018 #Eisteddfod

💬 254   🔁 10   ♡ 20

Marcus Stead @MarcusStead · Aug 5

The best-known song in the Welsh language is about a saucepan boiling over and a cat scramming a child. I'm not making this up, I promise! Some Welsh nouns: bocsio = boxing, snwcer = snooker. Pys isn't what you're thinking....it's actually peas! #Eisteddfod2018 #Eisteddfod

💬 143   🔁 2   ♡ 11

Figure 7. Stead Tweets – Deriding Welsh (Stead, 2018a)

In Figure 6 Stead blames the use of Welsh as the reason that Wales does not have a booming tourist industry, claiming that it makes visitors feel unwelcome. However, all road signs in Wales are bilingual, with both Welsh and English on them, making it possible for any English speaking visitor to understand them (Beta.Gov.Wales, 2018). Thus the addition of Welsh on the signs cannot be seen as a true hindrance. Figure 7 shows two tweets where Stead ridicules the Welsh language (implying the language sounds like the incomprehensible slurring that comes with being drunk), the Welsh people (in saying they probably are drunk, as if it is a foregone conclusion that the Welsh will always be drunk), and the Welsh culture (by belittling Welsh songs). By all appearances, Stead feels justified in making such derogatory statements and does not seem to view it as any sort of prejudice or discrimination. While not everyone who discriminates against language varieties may be as blatant about it as Stead, it is likely that most are similar to Stead in that they would not view their actions as discriminatory.

Foreign accented English speakers have also reported receiving ridicule and abuse due to their accent. Yunxiang Gao is a professor at a Canadian university and, while a native of China, has spent much of her adult life in North America. She notes that while issues regarding her accent are never brought up directly to her by the students during the semester, it is commented on at the end of every semester in

student feedback. Gao said when she first started teaching, she attended accent reduction classes due to a "recommendation" by someone more senior ranking. She mentions overhearing a student talking about another person who had a "cute" British accent and observes that the British speaker most likely would not have been "recommended" accent reduction classes. This difference in requirements depending on the specific accent would suggest a clear linguistic hierarchy, with some accents being perceived more positively than others (Sathiyanathan & Xing, 2018).

Though accent discrimination may not be systematic, all of these reports across news and social media show evidence of specific attitudes and third order stereotypes being held towards language varieties. Much of it is negative, whether it is a report of being discriminated against or, in the case of Stead, a strong and clearly negative opinion being expressed against a minority language. Additionally, it ought to be noted that in all the cases of accent discrimination being reported, it is a female who has felt its impact. Although this is a small sampling, and may be sexist reporting on the part of the journalists, this is most likely not a coincidence, nor is it surprising. It has long been observed, particularly in employment contexts, that women are more likely to face discrimination and barriers than men (Eagly, 2007; Powell, 2018). This discrimination can come from other women, as seen in Edwards (2014) just as much as it may come from men. Though these stories may be considered anecdotes, as they do not have the weight of statistical analysis to strengthen the conclusions, these news articles, tweets, and Facebook comments all offer recent real life examples of accent discrimination. This would appear to answer the question posed, showing that there are cases when people do act upon the linguistic attitudes they hold, and that this has the clear potential to lead to accent discrimination.

2.3. Research into Accent Discrimination

The previous section looked at real life situations in order to offer greater ecological validity to any conclusions drawn as the likelihood of accent discrimination continues to be evaluated. However, without academic research, it would be difficult to say whether these real life cases are common or simply outliers in society. Thus this section reviews various studies that have equally sought to

answer these questions regarding whether people truly do act upon their language attitudes, and in such a way as to be negatively discriminating against another person.

Purnell, Idsardi, and Baugh (1999) ran studies looking into reactions to GAE, AAE and Chicano English (ChE) and how they compared to one another. As one of the researchers, Baugh, was reportedly fluent in all three varieties, they used him as their voice actor in matched guise experiments. In their first experiment, Baugh called landlords from areas across San Francisco, based on advertisements they found in newspapers. He began every conversation with the same script, but varied the accent used, altering only the phonetics between each guise. The results observed that he was more likely to be offered housing when speaking in GAE than in either AAE or ChE. Any landlords who rejected his inquiry when he spoke in the non-standard accents, but accepted him when speaking GAE, presumably based their decision on auditory information alone. Purnell et al. contend that those landlords held stereotypes regarding those accents and the communities who spoke them, and reacted in accordance with those stereotypes. This, then, is arguably a case of accent discrimination occurring in housing. However, it is unknown if Baugh was able to authentically reproduce all three accents, and do so consistently across every phone call, as the methodology did not mention testing the guises for genuineness. The description of the methodology lacked detail, making it difficult to reproduce this study. For those reasons, although this is a well-known and well-cited study, it ought to be considered with caution and in light of other research.

Around the same period, Massey and Lundy (2001) conducted a similar, yet more robust, study looking into housing in Philadelphia. They tested whether speakers of GAE and AAE would be treated differently by potential landlords. They created a script and broke the AAE portion into two: one guise that used AAE syntax, lexicon, and phonetics and a second guise that used AAE phonetics but Standard English syntax and lexicon. They believed this distinction would lead landlords to perceive speakers with the full AAE dialect as both poor and black while perceiving speakers who only had an AAE accent as middleclass and black. They used both male and female students as their speakers, and in this way tested accent, class, and gender. The students spent a month calling various landlords and letting agents. They used the script created and were instructed to note how many

times they had to call before receiving a call back, if they were told the flat was still available, and whether credit checks or a deposit were mentioned. Overall the results generally showed that males were given better treatment than females, and the more standard the dialect (and therefore the higher the perceived class) the more likely speakers were to receive good treatment. The GAE male speakers received the best treatment, receiving return calls more quickly, more regularly told the housing was still available, and less frequently told a credit check would be necessary. The full dialect female speaker of AAE was treated most poorly in all of those same categories. It would often cost the poor, black female over twice what it would cost the GAE male speaker to get housing, in terms of effort, time, and actual money.

There is evidence of accent discrimination within other sectors as well, such as employment and education. The following examples are real life experiences that were analysed by linguists, and in some cases later encouraged further research into discrimination contexts. Aisha Azmi is one such case. She worked for a school in West Yorkshire, England in the capacity of language support. She was suspended due to her refusal to take off her *niqāb* in the classroom if there was a male present. The school argued that due to the material covering her mouth, the students would struggle to understand what she was saying (it is unclear whether the school had actually received any complaints from students or parents regarding this). Azmi brought this to the local council and asked the case to be investigated for discrimination, as the *niqāb* is tied in to cultural and religious beliefs. However, the council agreed with the school's argument and Azmi lost the discrimination case. There has since been a study conducted investigating whether various types of face coverings, including the *niqāb*, significantly reduce effective communication. Overall they did not find strong effects, suggesting that generally any inability to understand is a result of a decrease in visual/facial information or due to prejudice (Llamas et al., 2008).

Discrimination in employment sectors extends to the U.S. as well. In 1981 Sulochana Mandhare, an Indian immigrant who had lived in the United States for nearly 20 years had her employment terminated and was told it was due to her "heavy accent, speech patterns, and grammar problems" (cited in Lippi-Green, 1994, p. 164). This was despite a master's degree obtained in the United States and an excellent employment record. She filed suit under the U.S. Civil Rights Act and won

the initial civil action, but the decision was overturned in the court of appeals as her evidence of discrimination was not deemed strong enough (Lippi-Green, 1994). Mandhare's experience is not an isolated one. It has been noted that in court cases where the job position requires "communication skills," and an employer claims that the employee cannot be understood because they speak with an accent, the courts regularly side with the employer without further investigation (Del Valle, 2003). The legal issue comes in proving discrimination against her nationality; the linguistic issue deals with comprehensibility. Because the standard dialect is not a fixed point, how is deviation from it to be measured? To what degree must someone differ from the standard before they become truly incomprehensible? With no easy answer to questions such as these, courts are left without a clear precedent and opens up judgments to subjectivity (Kibbee, 1998).

The educational system within the U.S. also contributes to the notion of a single "correct" form, and thus adding to the potential opportunities for accent discrimination (Freeman, 1998). Due to language policies, it is expected by some schoolboards that students will acquire the standard in school, if they are not already learning it at home. They are not only expected to write their reports in GAE, but also to be speaking GAE in their classrooms (Huebner, 1999). This is justified by explaining that because GAE speakers are rated as more employable, teachers are guiding their students toward a better future (Cargile, 2000). However, these policies create a positive feedback loop. Students are taught that GAE is the most acceptable dialect. They adopt this ideology and promote it as employers, choosing to hire or promote only those who speak GAE. This then forces teachers to continue to defend SLI in order that their pupils might be better placed to gain professional jobs (Tamasi & Antieau, 2015). Over the years there has been progress in schools towards acceptance of non-standard dialects (Preston, 1999). However, in many cases children are still expected to adopt the standard due to the perceived prestige inherent in speaking the standard.

A well-known example of this occurred in Oakland, California in 1996. Although over 20 years old now, this case remains relevant when discussing the perceived level of legitimacy in non-standard dialects. The Oakland Unified School District school board passed a resolution recognising "Ebonics" (AAE) as a legitimate language variety. They believed that in the same way non-native speakers

of the English language are given additional assistance in their lessons, non-native speakers of Standard English should receive the same help. Both groups were expected to learn a variety of language not familiar or native to them and therefore if the schools wanted to see them do well academically, both groups should be eligible for the same support. By linguists' standards, this was still a fairly conservative educational reform because the school still expected students to acquire and use the standard dialect; they simply recognised that non-standard speakers were at an academic disadvantage in comparison to standard speakers. However, Oakland's school board decision became a controversial issue across the United States as people disagreed with recognising AAE as any type of recognised language variety. Many misunderstood Oakland's resolution, believing that teachers were going to start instructing in AAE, and that students who had not previously spoken AAE would now be expected to learn it. The general popular opinion was strongly that only Standard English was "correct" and grammatical (Lippi-Green, 2012; Wolfram, 1998).

The previously outlined academic references have all offered examples where language attitudes appear to have, at the worst, caused direct discrimination, or at best been the vehicle by which discrimination against ethnicity was justified. Regardless of which was the guiding motivator, the outcome is sobering as people are left to struggle for housing, employment, or a chance to be well educated. Considered alongside the personal narratives of discrimination discussed in section 2.2., it does not seem to be an overgeneralisation to say that accent discrimination regularly occurs across professional and personal sectors.

However, it must be noted that accent discrimination is not an inevitable and automatic occurrence. Some studies have failed to find any sign of discrimination. Cargile (2000) conducted a matched guise study looking at ratings of employability between GAE and Mandarin Chinese accented English in the United States. He recorded the same script answering the employment question "So how did you hear about this job opening?" using a male speaker and a female speaker for both accents. He then asked participants to rate the speakers they heard according to levels of employment suitability. His results showed no difference in ratings between the two accents, except for the Mandarin Chinese variant being rated as less suitable for positions that specified the need for good communication skills. Cargile theorized

that this result was due to the high economic success rate Chinese immigrants have experienced in the U.S., and thus the stereotypes regarding them may be positive with regards to employment. The one difference in rating could have been due to participants' concern over the comprehensibility of foreign language speakers' English and their potential varying levels of fluency. Additionally, it did not appear that he found any significant differences between the male guise and the female guise. Therefore, while women may experience, or report, more discrimination than men, this study would suggest that neither accent nor gender automatically cause discrimination.

2.4. Conclusion

This chapter examined the ways language ideologies and language attitudes may encourage people to act in a discriminatory manner. One of the main ideologies that may have an effect on this is Standard Language Ideology and the notion that there is one "correct" manner of speaking. The negative language attitudes in which this notion is reflected may be the cause of the discrimination described in news sources, social media, and academic research. It is clear that this discrimination occurs across sectors and potentially more frequently to women than to men. Yet it must be understood that accent discrimination does not always occur when language attitudes are present. Instead, it would appear that there is more nuance to the factors that trigger accent discrimination.

For the very reason that accent discrimination is not an automatic occurrence, it is important to research when it does take place, why, and what can be done to mitigate it. Academic studies into human behaviour, by their very nature, are controlled environments that restrict the amount of exposure participants have to outside distractions. This limits the level of ecological validity experiments can reach. Additionally, the variation between studies with their speaker stimuli and participant listener base can lead to diverse results that must also be taken into consideration when cross-comparing studies. Therefore, while study findings serve as a helpful tool to begin to understand trends and tendencies in society, each result must be considered within the context of its limitations. The studies I conducted, outlined in chapters 5-7, are no exception to this. Yet equally, due to the "messy"

and uncontrollable quality of the real world, it can be difficult to say with any surety that one variable, such as accent, had a larger effect than another variable in many real life situations. Therefore, these two contexts of real life and academic studies must be looked at together in order to gain a broader, and hopefully more accurate, picture. Sections 2.2. and 2.3. sought to do exactly that, and the following chapter continues to look at both real life examples and the findings of studies, but with the specific focus of the law and courtrooms.

Chapter 3. Language Attitudes in the Courtroom

Variation in language is normal and ubiquitous; noticing that variation is also only natural. Discrimination occurs when judgements are made regarding a person based solely on their speech, and then used to justify negative actions towards that person (Nguyen, 1993). If for any reason a listener develops specific attitudes towards the speaker, this is not automatically discrimination (Tamasi & Antieau, 2015). But when those attitudes are acted upon the speaker is left unheard and effectively rendered voiceless. Consider again the case of Rachel Jeantel as outlined in chapter 1; her overall testimony lasted for almost six hours, longer than any other witness testimony given throughout the trial (Rickford & King, 2016), and directly contradicted Zimmerman's own testimony. However, evidence would suggest that her testimony was disregarded because she presented it in a low prestige, non-standard dialect (Carlin, 2016).

One of the jurors later said in an interview that the jury did not understand her speech. Yet they did not raise their lack of understanding as an issue during the trial, nor did the jury ask for a read back of the written transcript of Jeantel's spoken testimony (Thompson, 2015), despite this being an option given to juries under Florida Rule of Criminal Procedure 3.410 (The Florida Bar, 2017). If the jury did not understand her testimony, and no efforts were made to rectify this problem, then it is logical to assume that Jeantel's testimony was not considered when the jury came to a verdict. The same anonymous juror confirmed this, directly stating that Jeantel was not mentioned in their 16 hours of deliberation, and that her testimony did not play a part in the verdict rendered (Rickford & King, 2016). This then presents a serious miscarriage of justice. Regardless of whether Zimmerman was in fact guilty or not, if

the key witness was ignored, and therefore the key evidence for the prosecution was not considered, the verdict cannot be deemed just.

Judging a person based on their dialect can be a form of discrimination when it causes a person to be treated differently. While dialects or accents at their base are terms used to describe the way someone speaks, they are also reflective of how speakers view themselves. They tell a story, giving a glimpse into the identities the speakers hold of themselves (Matsuda, 1991). Therefore, to deny that a variety of speech has value or legitimacy is to reject more than phonetics and the occasional turn of phrase. Rather, to undermine or reject any accent or dialect is to undermine or reject the speaker (Lippi-Green, 1994). Under the current laws in the United States people cannot force, or even ask, another person to change their religion, gender, skin colour or sexual identity (Wadham et al., 2010). Yet it is considered perfectly normal to request a person learn the standard dialect before offering them a job, as discussed in chapter 2 (Cargile, 2000; Nguyen, 1994). The reaction Jeantel, and others like her, faced due to speaking in a low-prestige dialect suggests accent discrimination is thus largely socially acceptable.

Rachel Jeantel's situation provides a demonstration of how people attach varying levels of prestige to dialects, shown in how she was criticized for using AAE in the trial (Thompson, 2015). Her experience was not unique in its occurrence, but rather in how highly publicized it was (Carlin, 2016; Carodine, 2015). While there has not been much research conducted into this topic, there are real-life examples like Jeantel's where it is possible to note the presence of accent discrimination. By comparing experimental research findings alongside transcripts from trials, it is clear treatment similar to that which Jeantel received occurs with some regularity. This chapter will expand on the negative consequences that come with discriminating against someone based on their speech and highlight when accent discrimination has been noted within the courtroom, both in actual trials and in research.

3.1. Discrimination in the Courts

Governments and societies have come far in recent years regarding the recognition of prejudices and racism. Laws have been put into effect to protect people against discrimination. The U.S. Title VII Discrimination Act and U.K. 2010

Equality Act are such laws, covering race, colour, nationality, religion and sex (Wadham et al., 2010). A later addendum was added to the U.S. Title VII, defining discrimination against national origin ". . . as including, but not limited to, the denial of equal employment opportunity because of an individual's, or his or her ancestor's place of origin; or because an individual has the physical, cultural or *linguistic* characteristics of a national origin group" (Office of the Federal Register, 1983, p. 148; emphasis added). This addendum has the potential to protect non-standard speakers against accent discrimination. In fact, a government document specifically notes the potential for accent discrimination, and spends a section discussing it and the duties required of employers when trying to prove they did not discriminate. However, in the process it is made clear that *accent* was misunderstood. The document states, "An accent can reflect whether a person lived in a different country or grew up speaking a language other than English" (U.S. Equal Employment Opportunity Commission, 2016, p. 35). Once again, as discussed in chapter 1, this shows the common misconception in the U.S. that accent is anything that is *other* and non-native, rather than understanding that everyone has an accent. This government definition in no way takes into consideration regional dialects or accents, and how a native of the United States may be discriminated against for speaking anything other than GAE. Therefore, accent discrimination against regional speakers is still technically legal, and court outcomes reflect that.

Due to the potential for lifelong damage that comes with verdicts, it can be argued that the most serious consequences of language attitudes causing accent discrimination is when it occurs in the courtroom (Eades, 2003). This section will examine U.S. court cases and discuss how accent discrimination played a distinct role in the proceedings. There is evidence in numerous trial transcripts and from academic studies (expanded upon in the next section) that some non-standard dialect speakers have been treated differently within courtrooms. Witnesses' testimonies may be considered less reliable when spoken in a non-standard dialect (Frumkin, 2007); as previously mentioned Rachel Jeantel is an example of this. Trial outcomes would suggest that it is not just witnesses who face this potential discrimination but defendants also. This chapter only looks at a few of those trials, namely: *Clifford v. Commonwealth,* 1999; *Clifford v. Chandler*, 2003; *Hyppolite v. State*, 2002; and *Williams v. Frank*, 1991. An additional two cases will be considered with regards to

foreign accented English. However, as they dealt with minors, the court transcripts have been sealed, so only the information as relayed by the defence attorneys to journalists will be detailed. In all these cases the experience of the defendant(s) involved will be the focus.

In U.S. court cases, racial voice identification has been seen as a reliable form of identification, despite it being grounded on stereotypes (Hollien, 2002; Kohler, 2004). This is seen in cases such as *Clifford v. Commonwealth* (1999) and his appeal in *Clifford v. Chandler* (2003) in the U.S. state of Kentucky. In *Clifford v. Commonwealth* (1999) Charles Clifford was convicted of drug trafficking, based on the pivotal testimony of a police officer who gave evidence regarding having heard a "black" voice. This testimony was allowed despite the fact that the officer heard the disputed voice via an audio recording, and the recording itself was not admissible to court, as the judge ruled it inaudible. During the defence's cross examination, the officer did acknowledge that not all blacks sound alike while speaking and some of them may "sound like whites" and vice versa (*Clifford v. Commonwealth*, 1999). Despite this, he maintained that he could tell the disputed voice he heard was black. Clifford had never previously met this officer and he did not testify in court. Therefore, there were no opportunities for either the officer or the jury to hear either his voice or what dialect he had. In addition to this, a white male who was known to be at the scene of the crime had confessed to the crime and testified that Clifford had no part it in. There was also no physical evidence linking Clifford to the crime. Yet due to the testimony of the officer, and the fact that Clifford had been the only African American in the vicinity during the period of the crime, his guilt was assumed (Wiehl, 2002).

Clifford appealed his conviction in *Clifford v. Chandler* (2003), arguing against the admissibility of a testimony based on racial voice identification, along with other subversion of justice claims. However, the court responded that even "circumstantial evidence is substantial enough to establish guilt" (*Clifford v. Chandler*, 2003). Additionally, the court cited the Purnell et al. (1999) experiment in which their participants were able to correctly identify an African American male just over 88% of the time when speaking AAE (Kohler, 2004). The precedent this court ruling set seems to assume that any person speaking AAE is African American and that all African Americans speak AAE, and it ignores studies that have shown

otherwise. It also ignored the fact that the witness never mentioned identifying the "black voice" through AAE, but rather due to the voice quality (Wiehl, 2002), an aspect that Purnell et al. (1999) did not research.

Jacobs-Huey (1997) and Williams (1973) (as cited in Wiehl, 2002) contend that it is not possible to accurately identify a person's background or race based on their dialect alone; rather, people are more likely to react to what they are *expecting* to hear. Jacobs-Huey (1997) recorded three male voices, all with similar levels of education, from a middle-class background, and in their 20s. Two were African American and the third was white, but all three spoke AAE. She then played these recordings to 92 participants. The two African American males' race was correctly identified 92% and 85% of the time. The white male's race was misidentified as African American 92% of the time. It is not surprising that AAE is associated with the African American community. However, Jacobs-Huey's results suggest that people expect only African Americans to speak AAE, thus misidentifying the white male.

Williams (1973) (as cited in Wiehl, 2002) recorded a single GAE child and then played that one recording to participants with a video of either a white, African American, or Hispanic child. Participants were then asked to comment on the level of proficiency of the speech they heard. In this study participants rated the white child as both fluent and standard, the African American child as fluent but non-standard, and the Hispanic child was judged lowest overall (as cited in Wiehl, 2002). Yet all the participants listened to the same voice. These results suggest that what they heard was filtered through stereotypes they held towards each community. Jacobs-Huey and Williams' studies argue that dialects are strongly associated with the community who speaks them—African Americans speak AAE, and are perceived as speaking less "properly," while white are believed to be more likely to speak GAE than non-white. The results of both studies, which were already published and available at the time of Clifford's initial trial and his appeal, furthers the argument that Clifford's conviction, regardless of whether he was in fact guilty, was based on racial voice identification and therefore should have been overturned.

Thomas (2002) conducted a review of sociolinguistic perception studies. In contrast to Jacobs-Huey (1997) and Williams (1973), most of the studies Thomas

examined found that their participants could accurately identify the speaker's ethnicity above the level of chance; although it was noted that no research could agree on a specific linguistic marker that caused this. It was suggested that participants were most likely relying on stereotypes to answer correctly, such as naming the community group/ethnicity who was most associated with the specific dialect they heard. For example, if it was an AAE speaker, then they must belong to the African American community, which could only make them African American themselves. While this may be true in most cases, as shown in Jacobs-Huey's (1997) (as cited in Wiehl, 2002) research, other ethnicities may speak AAE and thus be falsely identified due to a reliance on stereotypes.

Further research supports the conclusion that a person's background and ethnicity cannot always be accurately identified based on their speech alone; even trained experts can struggle (Cheshire, Fox, Kerswill, & Torgersen, 2013; Fraser, 2011). There are a number of confounding factors. For instance, if a person has moved around or spent a large amount of time with speakers of another dialect, that person may converge in their language, taking on speech markers of other communities beyond the one they grew up in. This can make it difficult to place exactly where a person is from. Additionally, in cases where a person sets out to commit a crime, there have been known instances when they purposely changed their voice quality and accent. While experts could note certain phonetic features that clearly were not natural to the person (and thus the observation that the speaker had disguised their voice), it can be a near impossible job to then say which features *are* natural to the speaker and what their native dialect is (Watt, 2018). With this in mind, when it comes to criminal cases caution must be exercised when accepting a description of any voice, particular by a non-linguist.

Discrimination within courts can also occur due to misunderstanding general linguistics and poor (or outright wrong) interpretations of linguistic research. This can be seen in *Clifford v. Chandler* (2003) where the court misinterpreted and misused the Purnell et al (1999) results. A further example of misinformation regarding linguistic principles is found in the court *Hyppolite v. State* (2002). Larry Hyppolite, a Haitian-born American, was convicted of selling cocaine in 1996 in the state of Indiana. A police officer visually identified him as the seller. However, there was also an audio recording of the sale interaction between the officer and the seller,

and the speaker's voice did not match Hyppolite's. The voice in the audio recording was regularly mentioned in the court transcript as having "no accent," yet the defendant spoke English with a Haitian accent. Despite this, the prosecution convinced the jury to convict the defendant by arguing that the defendant had purposely changed his accent in order to disguise his identity. The prosecution was able to do this in three main ways, utilising the defence attorney's, the judge's, and the jury's lack of linguistics knowledge in order to make unsubstantiated, and blatantly false, linguistic claims.

First the prosecution noted that the defendant had been a part of the army, and there had the title of "linguist" as his job role. By his own admission, the defendant was an interpreter for the army. However, there is no evidence that the defendant had any formal linguistics training. The prosecution defined linguistics as being primarily concerned with sound change, and therefore argued that the defendant knew exactly how to change his speech in order to perfectly mimic another accent. While historical linguistics and sociolinguistics, both branches of linguistics, may be primarily concerned with sound change, it is wrong to suggest that linguistics as a whole focuses on that one area.

From there the prosecution moved on to speak about what sociolinguists define as "style." This refers to when people change their level of formality in speech depending on the context and the amount of attention they are giving their speech (Bell, 1984; Eckert, 2000; Giles & Powesland, 1975). Through a series of questioning, the prosecution succeeding in getting the defendant to admit that he would not swear in front of a priest but would swear in front of his friends. The prosecution used this to argue that this showed the defendant regularly changed his speech to suit his needs.

Finally the prosecution brought up the fact that the defendant could speak five languages. From this, the prosecution suggested that the defendant only spoke with a Haitian accent as an act, and could adopt other accents since he had been able to learn other languages (Rodman, 2002). This ignores linguistic research that discusses the "critical age" theory and how after a certain age, which the defendant was, it was near impossible to learn a foreign language without the speaker's native language

42

interfering and being evident as a clear "non-native" accent (Birdsong, 1999; *Hyppolite v. State*, 2002).

In all of this, it is unclear whether the prosecution deliberately misrepresented linguistic principles for their own purposes or if their own lack of linguistic knowledge contributed to the gross amount of misinformation. The defence did not know enough about linguistics to object to the inclusion of such "evidence." This left the jury misinformed, believing that despite the voices not matching, the audio recording evidence could still be Hyppolite's. Due to this and other evidence brought, the jury found Hyppolite guilty and the judge sentenced him to 20 years in prison.

Hyppolite appealed this conviction in 2002 and as a part of that appeal had his own voice and the audio recording forensically compared by Robert Rodman. Rodman's report stated that the recorded voice did not match Hyppolite's. However, as the court did not fully understand how this analysis had taken place and its legitimacy, it was not admitted in to the official appeal and Rodman was not allowed to appear as an expert witness. Hyppolite's appeal was eventually denied on the grounds of his not having provided enough evidence to support his arguments (*Hyppolite v. State*, 2002; Rodman, 2002). Hyppolite was convicted partially based on unsubstantiated and outright incorrect claims, and his appeal was denied in part due to ongoing linguistic ignorance. While it is not clear whether the prosecution was aware they were giving false information, this is still discrimination. They found arguments that fit the narrative they wanted to tell, and did not do their due diligence in discovering what linguistic research would actually say or asking a linguist to testify as an expert witness. Rodman summarises this issue, noting,

> "If a person were convicted of a crime because a jury was convinced of a geocentric solar system, or of the phlogiston theory of combustion, or of the four humours, there would be outrage, and the conviction would (one prays) be overturned. Yet the chain of linguistic non sequiturs that allowed [Hyppolite] to be convicted is no less absurd, no less naïve in today's society, than these once highly respected ancient theories" (Rodman, 2002, p. 101).

As seen in *Hyppolite v. State* (2002), despite the protection promised by Title VII and its later addendum, foreign nationals and L2 speakers may still be

discriminated against in courts. However, this could be due to prejudices as well as ignorance. Journalist Weiss (2008) discusses a court case in the state of Pennsylvania where four Latino men between the ages of 17-22 were accused of criminal conspiracy to commit robbery. The judge sentenced them to either 24 months in prison or to learn English. They were given a year to pass an English proficiency test and if they failed it they would be expected to serve out their 24 months in jail. This was due to the four men requiring translators during their trial, something the judge appeared to take exception to as he was reported to having asked them, "Do you think we are going to supply you with a translator for the rest of your life?" The judge later defended his decision, saying he was giving the four men a chance to improve their lives and gain full-time employment.

Another journalist Barry (2005) wrote an article discussing ongoing discrimination in the state of Tennessee. County Judge Barry Tatum had become known for threatening to separate immigrant parents from their children unless they learned to speak English with a specific degree of fluency. One 18 year old woman from Mexico was reported for not immunizing her toddler. At her hearing Tatum's response was to tell her to learn English and use contraceptives. In another instance a Mexican mother was reported for neglecting her 11 year old daughter. After considering her case, Tatum set another court date for six months later, saying she had that long to acquire at least a fourth grade level of English fluency. The court order was reported to say, "The court specially informs the mother that if she does not make the effort to learn English, she is running the risk of losing any connection — legally, morally and physically — with her daughter forever." During those six months the 11 year old would be taken away and put in foster care. The court order implied this would act as a helpful incentive for the mother to learn English. At the time of the article being written, Tatum was known to have given these types of orders to immigrant parents in at least five cases, something that was viewed by the local community as a helpful encouragement to the immigrant population to fully assimilate (Barry, 2005).

Other non-standard dialects not associated with specific ethnicities or nationalities can be affected as well, although this can be even harder to prove in court. While race is protected under the U.S. Title VII Discrimination Act, regional speech and dialects are not. One example of this is in the court case *Williams v.*

44

Frank (1991), seen in how the court responded to Williams' claim of accent discrimination. Williams was fired from his job as a window clerk for the US Postal Service in New York City due to multiple counts of misconduct. Williams brought his employer to court, stating he was actually let go due to discrimination with regards to the fact he spoke Southern American English. It is incredibly difficult to win a discrimination court case, as there is a heavy burden of proof on the plaintiff and judges are more cautious before ruling in favour of the plaintiff (Del Valle, 2003). In this instance, the court rejected the evidence Williams brought regarding the way his supervisors made fun of his dialect, asserting that his supervisors only spoke about his speech, not his race or any other characteristic protected under Title VII. The court declared that as "southerness is not a protected trait" there were no grounds for a Title VII discrimination suit (*Williams v. Frank*, 1991). While there were other significant grounds for Williams' dismissal from his job, *Williams v. Frank* has become known for this statement, and is cited by other case rulings when dealing with disputes regarding regional heritage rather than national (*Langadinos v. Appalachian School of Law*, 2005; *Storey v. Burns International Security Services*, 2004; *Thomas v. East Orange Bd. Of Educ.*, 2014).

Williams v. Frank set a legal precedent, saying accent discrimination is not unlawful, or even true discrimination, unless it can be proved that the discrimination is linked to another characteristic such as race, religion, gender, etc. Yet even in cases where a difference in race or nationality is involved, such as with Clifford and Hyppolite or the trials described by Weiss (2008) and Barry (2005), language can still be used as a medium to justify discrimination and outright racism. Thus, even if research finds that language attitudes generally do not go far enough to be the cause of overt discrimination, because it can still be a vehicle used for discrimination, efforts should be made to add *language* and *speech* (regardless of their national or regional origin) to the protected traits list. One potential avenue for accomplishing this, within the United States, is through the constitution. The first amendment gives the right to freedom of speech. This is commonly interpreted as having the right to say *what* one desires, but it could also be understood as giving the right to speak *how* one desires (Miner, 2013). Therefore, this amendment has the potential to protect US citizens from discrimination based on the accent they have, whether regional or foreign, thereby making accent discrimination unconstitutional and prosecutable in a

court of law. So far as I am aware, however, the first amendment has never been used by lawyers in this way. Due to that, this possibility first needs to be brought to the attention of lawyers and used in such a way in order to set an anti-accent discrimination precedent.

The trials outlined here are a small sampling of the trials that have been marked as having some aspect of accent discrimination. Not only may people be discriminated against due to their speech, but they may also be disadvantaged from the start of the trial by not being able to understand the highly formal and technical language used. This could be just as true of native speakers of English as it is for non-native speakers (Kibbee, 2016). If a person does not know the "proper" courtroom procedure, such as was the case with Jeantel, there is the potential to be viewed in contempt of court. Equally if a person cannot understand what is being said around them, how can they be expected to offer a strong defence or a clear testimony? Defendants and witnesses should not only be given their full first amendment rights to testify in the language or dialect they speak most fluently, but the courts also need to recognise the same defendants and witnesses deserve to have all proceedings "translated" for them. This could be in a different language or by having someone explain the procedures and expectations in simplified terms. If a person's ability to testify is hampered, then it can be argued that the full evidence available in the case was not provided, and leaves the potential open for a miscarriage in justice.

For further case studies, an extensive list of trials can be found in Lippi-Green (2012, pp. 175-178). The trials there only name the ones where it is clear accent discrimination occurred. It is highly likely that the number of cases where this has occurred is much higher, as it can be difficult to reliably identify accent discrimination as the one variable out of many that played a large enough role as to affect the verdict. With additional variables such as defendant or witness gender, pre-deliberation biases on the part of the jurors, the type of jury instructions given by the judge, pre-trial publicity, whether the jury is to offer a dichotomous verdict or given a multiple response option, the size of the jury and whether any of them have previously served on a jury before, it is no surprise that accent discrimination can be difficult to discern (Dhami & Belton, 2016; MacCoun, 1989; Steblay et al., 1999; Werner et al., 1985).

The court cases listed above all came from the United States. Despite speaking with legal experts and searching through online trial databases as well as academic articles, I was unable to find cases in the United Kingdom, or elsewhere, where I could clearly point out the presence of discrimination due to linguistic features. This does not mean they do not exist, simply that I was unable to find them owing to the nature of trials – with the vast number of variables involved in a court case, it can be difficult to measure which ones had a stronger or weaker effect. Additionally, lawyers have no obligation to justify to the public their various approaches to trials, so it is not always obvious if they are using linguistic stereotypes when arguing their case. Equally, if lawyers are using them, it may be unclear whether they are doing so consciously or subconsciously based on trial transcripts alone (Levinson, 2018). Finally, juries and judges generally do not explain their decision making processes (in fact they are often either strongly encouraged not to or directly forbidden to do it), which again makes it hard to know which evidential variables mattered and which did not (Ferguson, 2002; Gov.UK, n.d.; United States District Courts, n.d.). Due to these unknowns, in many cases it would be unverifiable to claim that language had a clear effect on the trial and its outcome. It would be surprising, however, if this was an issue only the U.S. struggled with. There is every possibility that bringing this issue to light in the U.S. will cause it to be noticed and dealt with elsewhere. Therefore, while other locations are not specifically mentioned, the examples here may be no less relevant for them.

3.2. Research into Accent Discrimination within the Courtroom

From being denied housing to being convicted mainly on the basis of stereotypes, there are clear and severe repercussions to accent discrimination. Yet in order to establish accent discrimination as a true form of discrimination in a court of law, and encourage dialects to be considered as a protected trait, research needs to show two things: firstly that dialects and accents are an integral part of a person's identity and therefore ought to be protected; and secondly that people do discriminate and treat others differently based on how they speak, similarly to discrimination due to race, ethnicity, gender, etc. Research has answered—and continues to answer—the first part of the question. Numerous sociolinguistic studies have shown how a person will often reflect their identity through the way they

choose, whether consciously or subconsciously, to use language (e.g. Hall, 2013; Omoniyi & White, 2006). However, there have been significantly fewer studies researching potential cases of accent discrimination in real life environments, with arguably the best known one being the Purnell et al. (1999) study, discussed in chapter 2. What follows is a critical analysis of studies, spanning the last four decades across multiple continents, which have specifically researched accent discrimination in a courtroom setting.

It ought to be noted that comparing studies must be done with caution. The use of different methodologies across locations, cultures, and time periods, which will affect the language ideologies of the participants, could lead to unreliable conclusions if overgeneralised are applied outside of the context in which each study was conducted. The studies discussed in this section have all helped to form the basis of the research conducted for this thesis (discussed in the following chapters). However, while their results can inform future research, their contexts and limitations should also be taken into consideration.

Seggie (1983) conducted what appears to be the first study looking at the effects of dialects in a legal context. Building on research that showed there was a preference for specific dialects in some occupations (e.g. that non-standard speakers are perceived as being more suitable for low prestige jobs than high prestige ones), Seggie questioned whether dialect and the type of crime would reveal a similar bias based on the occupation of the accused. He hypothesized that standard dialects would be more likely to be associated with "white collar" crimes whereas non-standard would be likely to be associated with violent crimes. For his study conditions he chose to simulate three types of crime (violent crime against a person, violent crime against property, and a theft involving embezzlement but no violence) and three dialects (English Received Pronunciation, Broad Australian, and Malaysian Chinese English). Seggie had one male speaker who recorded all three guises. He ran this study in Australia, and chose those dialects as he assumed his participants would be familiar with and even have regular contact with speakers of these dialects.

75 Australian university student participants were randomly given a piece of paper containing a summary description of one of the three crime types and were

told the police had three men in custody for that crime. After reading the summary of the supposed crime, they were asked to listen to an audio clip from each of the three men (what was said on the audio clip was not described). Participants were then to rate, on a Likert scale of 1-10, the likelihood of each speaker having committed the crime condition they were given. After attributing levels of the likelihood of guilt, participants were asked to listen to the voices again and recommend a punishment they deemed appropriate to the crime, and also to identify the country of origin of the speakers. These were open questions allowing the participants to write in any answer. Analysis of the data found that the English RP voice was significantly more likely to be accused of the crime of embezzlement than the other two crimes, whereas the Australian voice was significantly more likely to be found guilty of the violent crimes than embezzlement. The Malaysian Chinese guise did not differ significantly between any of the crime types. Seggie suggests that this is evidence that general stereotypes regarding that dialect were not yet well formed, either negatively or positively, due to Malaysians only recently immigrating to Australia in large numbers; therefore it appears the responses stayed fairly neutral. There were also no significant differences between the recommended punishment and the dialect.

This study by Seggie (1983) concludes that speakers of more prestigious dialects, such as RP, are perceived as more likely to commit non-violent crimes. However, this study cannot say if dialect could have made a difference in whether a defendant was convicted, as this was not tested. Instead this study offers insight into the stereotypes held regarding the lifestyle of different dialect speakers (e.g. RP speakers must hold higher positions in business in order to have the opportunity to even commit embezzlement, much less seem likely of it). It does not suggest that speakers of the dialects were treated differently in the legal aspects, as the recommended punishments were similar enough to not be significant. Additionally, if one dialect was discriminated against to the point where it would have affected the outcome of a case, one would have expected to see it rated more poorly against the other dialects, regardless of the crime type condition.

The second study researching potential discrimination in courtrooms, by Sobral Fernández and Prieto Ederra (1994), was published over a decade later and run in the Basque region of Spain. The researchers set up the experiment to look into

whether eyewitnesses were perceived with varying levels of favourability depending on which regional Spanish dialect they spoke. Although conducted in the Basque region, the Basque language was not used, only the Spanish dialect for that area, as well as other regional Spanish dialects. They used three actors as the eyewitnesses and trained them in not only the phonetics of the regional dialects across Spain, but also in the production of pauses and intonation in order to achieve native or near native proficiency in each dialect employed. Their participants, 200 university students, took the role of individual mock jurors, judging the reliability of the testimony they heard. The results suggested that overall participants were more likely to trust the testimony spoken in the dialect closest to their own. They appeared to feel more solidarity towards the speaker if he or she spoke like someone from their home region. This research proposes that within the Basque region familiarity, compared to prestige, plays the larger role in juror perceptions of testimonies. This is not a surprising result within the context of Spain and the Basque, as the Basque region is known for its (in some cases extreme) nationalism and struggle for independence from Spain (Olivieri, 2015; Sullivan, 1988). Therefore it follows that local participants would be more likely to choose solidarity and familiarity characteristics over prestige and status. Additionally, this is in agreement with earlier research into language attitudes which found that some people have "accent loyalty," where they are more likely to ascribe positive characteristics to their own accent than to others (Giles, 1971).

The studies by Seggie (1983) and Sobral Fernández and Prieto Ederra (1994) are notable in pioneering methods of looking at perceptions of dialects in courtrooms. However, there are criticisms that can be made regarding aspects of the design or interpretation. Both studies had participants come in and contribute individually, rather than in groups. This reduces the ecological validity of results as jury verdicts always involve group dynamics, not individuals, and research suggests that the group aspect does have an effect. This is seen in how the minority position within a group will generally acquiesce to the majority position, making the majority opinion the eventual verdict (Bornstein & Greene, 2011). Another relevant group dynamic involves individuals becoming more confident in a decision after being a part of a group. This may have the outcome of encouraging stronger reactions, known as *group polarization*, as people feel more justified or generally "right" when

they know others agree with them (Myers & Lamm, 1976). This is often found in cases where people have separated themselves into groupings, whether that is along lines of race, religion, political party, etc. (Iyengar & Westwood, 2014). It has also been observed in online groups (Yardi & Boyd, 2010). These are a few examples of how group dynamics can have an effect, and that effect can occur while the group is together as well as continue to inform their decision making even after the group has dispersed (Forsyth, 2017). Therefore, it is important to consider the use of groups when conducting research into trials and juries, something Seggie and Sobral Fernández and Prieto Ederra appear to have disregarded.

Additionally, both Seggie (1983) and Sobral Fernández and Prieto Ederra (1994) only played audio recordings to their participants. This has the potential to "clean up" the study environment too much, as it leaves mock jurors with very little to take into consideration beyond the speech. This bring speech into focus in an artificially prominent, and therefore unrealistic, way. In the vast majority of court cases, jurors would be able to both see and hear the person testifying, and judge their reliability and truthfulness over long periods of interaction, not just via recording. Sight is taken away only in the extremely rare cases when an innominate jury is used – this entails placing a screen in front of the jury to protect their identity from anyone in the courtroom, leaving them only the ability to hear but not see anything. Furthermore, research into the use of innominate juries has found they are actually 15% more likely to convict than regular juries (Mangat, 2018). This would suggest that giving only audio most likely significantly lessens the ecological validity. Stereotypes and discrimination may be triggered that otherwise would not have if the participants had more variables to base their judgments on, as they would in a general trial. Not only that, but it is possible that mock jurors would find it easier to discriminate against a voice they have no association to, when "the other" is in sharp focus, whereas seeing the speaker as well as hearing them could humanize that person and make the jury less likely to act on their stereotypes.

Another flaw in both Seggie (1983) and Sobral Fernández and Prieto Ederra (1994) is that they used only university students as participants, which are not an accurate representation of the entire demographics that make up communities. These studies are also old enough that their results could be outdated due to changing language ideologies around standardisation, as recent years have brought an increase

in positive perceptions of non-standard dialects (Coupland & Bishop, 2007). Due to this, these studies ought to be considered in light of more recent research. Finally there is no evidence that the researchers' interest in language was hidden, and therefore the participants may have been primed for the linguistic component, lessening the ecological validity of these studies' results.

In order to address the question of when accent discrimination may occur in legal contexts, Dixon, Mahoney and Cocks (2002) chose to replicate Seggie's (1983) study, conducting their study in England. For their study, they compared RP with a "Brummie" accent (i.e. the accent from the Birmingham area). They used a male actor to portray the suspect in order to create a matched guise study. He was able to reliably produce both accents, as confirmed by naïve raters. While their main independent variables were accent, they also tested whether the ethnicity (black or white) and the crime type (blue collar: armed robbery or white collar: cheque fraud) would affect perceptions of guilt, replicating Seggie's (1983) study. Participants heard physical descriptions of the suspect in the audio, as well as descriptions of the crime committed. Dixon et al. based the stimuli on the transcript of a police interrogation of a suspect from 1995, creating a two minute audio recording that they asked participants to listen to and afterwards provide a guilty rating from 1 (innocent) to 7 (guilty). 119 white undergraduate students at University College Worcester acted as participants, 24 of whom were male.

Dixon et al.'s results showed that the "Brummie" guise was rated more negatively with regards to perceived levels of "superiority" and "attractiveness," and that the black person speaking Brummie who committed a blue collar crime was perceived worst of all the conditions, at a statistically significant level. This is consistent with other research which has found standard dialects are generally preferred over non-standard (Coupland & Bishop, 2007; Giles & Coupland, 1991), but is in contrast to the study results of Sobral Fernández and Prieto Ederra. This could be due to the difference in cultures between the U.K. and Spain, particularly due to the Basque's previously mentioned separatist values (Olivieri, 2015). Additionally, while the regional accent was perceived as more likely to be guilty overall, and therefore these results broadly agree with Seggie's (1983) results, there was no evidence that the type of crime had an effect. This could be due to a change in society in the two decades that passed between Seggie and Dixon et al., due to a

52

difference in culture between Australia and England, random fluctuation, or combination of any of the above.

Dixon and Mahoney (2004) replicated their 2002 study, adding in the effect of weak evidence and strong evidence while removing the race variable, for a total of six conditions. They did this in order to test the *liberation hypothesis*, which says that extra-legal factors (such as language) will generally only have a significant effect when the legal evidence is weak. They predicted that accent would have a stronger effect on the results when the legal evidence was weak, in that participants would perceive the suspect as "guiltier" (the researchers' term) when hearing a non-standard accent with weak legal evidence. Their audio stimuli were the same as the initial study, and they added a booklet for participants to read about the crime scenario that contained either strong (the suspect had a criminal background or the stolen items were found in his house) or weak (no criminal background or no physical evidence found connecting to the suspect) evidentiary information. Once again Dixon and Mahoney used university students as participants, with 171 women and 28 men, and their participants were randomly given one of the six conditions. They found similar results to their first study, in that the regional-accented speaker was rated more negatively than the standard speaker, although they were unable to say whether this effect was directly or indirectly due to accent. However, accent did not have an effect on participant responses with regard to the crime type or the strength of evidence. This contrasts with the results of Seggie's 1983 study and refutes the liberation hypothesis.

This interaction between strength of evidence, type of crime, and accent could be studied in more contexts and with differing methodologies. For instance, the use of video stimuli rather than only audio recordings could prove useful, as, similarly to the studies that came before them, both the Dixon et. al. (2002) and the Dixon and Mahoney (2004) studies used audio recordings with no video component. As previously mentioned, this puts the focus on speech alone in a way that would generally not occur in a real trial. While there were no significant findings in the two studies beyond the standard accent being preferred over the non-standard, it is likely that there are additional variables at play and it would be helpful to know when the linguistic stereotypes will be triggered and by what. In light of that, it would be of interest to test the effect of the liberation hypothesis further in order to better

understand when accent discrimination may become relevant in court cases. Additionally, while they did exclude any participants from Birmingham, it does not appear that they considered the large discrepancy regarding the gender of their participants in their statistical analysis. In both their 2002 study and the follow up study in 2004 there was a much larger number of female participants than male participants. Not only that, but both of their studies used only university students. It is possible the results were affected by both of these issues as their participant bases were not an accurate reflection of a general populace. Finally, as with the previous studies, there is no evidence that they masked their interest regarding the linguistic variable and therefore may have accidentally primed their participants.

Frumkin (2007) ran the first study on this subject in the United States at the University of Maryland, examining the effect of foreign accented English in courts, rather than regional dialects. Following a pilot study, she decided to use three nationalities, German, Mexican, and Lebanese, as the pilot study results indicated the level of prestige would decrease with each one. She recruited three female actors who had all grown up bilingual in one of the three targeted languages and GAE. They were given 12 months to practice both guises, and at the end of that time the authenticity of each guise was confirmed by other linguists. Additionally each actress was reported to resemble the stereotypical "look" of someone from the national group she was representing. However, no pictures or explanations were provided regarding in what way the look was "stereotypical." With the use of a script, the actresses recorded two videos each in a moot courtroom: one video testifying in the foreign accented English of the country they were representing and a second video testifying in GAE. They gave the same testimony, acting as a witness in a courtroom, with the only difference in the script coming when they state their nationality. Altogether there were six video variables: 1. German appearance, GAE; 2. German appearance, German accented English; 3. Mexican appearance, GAE; 4. Mexican appearance, Mexican accented English; 5. Lebanese appearance, GAE; and 6. Lebanese appearance, Lebanese accented English. Frumkin called the GAE guises "accent-free conditions."

Responses from 174 participants were gathered for the study, 63% of whom were female and only 11% of the participants had ever travelled outside of the United States. The participants, in groups ranging from 1 to 16, were shown one of

the six videos, following which they were each asked to complete a survey regarding their perceptions of the testimony. This included questions regarding the various degrees (on a Likert scale of 1-10) to which the participants felt the witness was credible, accurate, deceptive, and prestigious. Additionally participants were asked whether they believed the defendant to be guilty and if so what an appropriate punishment would be; this served as another way to test the degree to which participants believed the eyewitness testimony.

It was found that in the accent-free conditions there were no significant differences between the three ethnic appearances. However, all the accent-free conditions were more positively rated than the accented conditions. Each actress was perceived more negatively in all four witness characteristic questions when she testified in the foreign accent guise than when she testified in GAE. As for the foreign accented conditions, contrary to expectations, there was no significant differences between German accented English and Mexican accented English. Both were rated more favourably than the Lebanese accented English guise. Irrespective of the speech guise, the actress who stated she was German was deemed more credible than the Mexican actress, and equally the Mexican actress was perceived as more credible than the Lebanese actress.

While Frumkin's study did not find as many significant differences between the accented varieties of English, there was still a clear result in that the non-accented English guises were preferred, regardless of the ethnicity of the speaker, over any of the accented English guises. This further supports the theory that the standard dialect is preferred and given more prestige over any non-standard dialect or accent. Additionally, as the study was run in the United States, Frumkin argued that participants' likely familiarity with Mexican accented varieties combined with the higher prestige attached to German (Kibbee, 1998), due to being a Western European language, may have led them to give similar ratings to those guises, while Lebanese accented guise was rated more negatively. This explanation would align with Sobral Fernández and Prieto Ederra's findings (1994), in that jurors will give a certain amount of preference to varieties that are familiar (such as Mexican) when contrasting with less familiar varieties (such as Lebanese). However, Frumkin does not appear to have collected any information regarding which ethnic groups her participants had had the most contact with, and therefore claiming that Mexican and

German would be most familiar is potentially relying on stereotypes and an overgeneralisation. Moreover, social psychology research has found that there is truth in the phrase "familiarity breeds contempt," in that the more people learn about someone or something, the more likely they are to find fault with them or it (Norton, Frost & Ariely, 2007). Therefore, even if Frumkin's participants were more familiar with German and Mexican accents, this does not automatically explain why they rated those guises higher than the Lebanese guise.

Despite the fact that juries come to decisions as a group comprising 6-12 jurors, of all the studies conducted only Frumkin (2007) attempted to have group participation in her study. Yet her groups ranged from numbers of one person (thus in fact completely removing the group aspect) to 16 people (a group that is both larger than any jury and, with so many people, would be difficult to monitor them all). This large a spread gives little consistency, making it questionable how much the groups could be compared to one another. She also did not specify how many groups she had in total. Additionally, she, like Seggie (1983), did not use regional dialects but instead used L2 English, speech associated with different nationalities or ethnicities. This opens the possibility that their participants rated the speaker based on stereotypes regarding the nationality or ethnicity rather than the linguistic features. However, Frumkin's finding that all the non-accented varieties were rated more positively than any of the accented varieties would indicate that language and not ethnicity was the deciding factor in this case. Once again though, as with the previous studies, there is nothing to suggest that Frumkin did not accidentally prime her participants for the linguistic variable.

Further research into this topic has also been conducted on a smaller scale by university students. Gray (2010) studied how listener age may effect perceptions of guilt depending on the accent. Her two male "suspects" adopted a Scottish, Newcastle, or RP accent guise, though she was particularly interested in the Newcastle and RP responses, with the Scottish guise added in as "filler." These were semi-scripted, which allowed her voice actors to speak as naturally as possible. The situation was set as a mock police interview, with a police officer questioning the suspect in regards to the supposed crime. Similar to Seggie (1983), Dixon et al. (2002), and Dixon and Mahoney (2004), she also added crime type as a variable, though it was not explained how she differentiated between a blue collar crime and a

white collar crime. Gray had a total of 80 participants from across the United Kingdom, as she conducted her study online, recruited through email and social media. However, she did not specify how she ran her study online, whether it was through an online survey software or if she sent each participant the necessary material. Participants were randomly assigned one of the two crime conditions (giving 40 participants per condition with an equal split of older and younger participants). Participants listened to each of the three recordings, and were asked a series of questions, the most important one being a 6-point Likert scale regarding how likely they thought a jury would find each speaker guise guilty of the crime.

While no significant effects were found regarding accent or crime type, Gray found that older listeners tended to give weaker "guilty" responses than younger listeners. She attributed this to older listeners being more likely to note the lack of case information and therefore finding the suspect less guilty than younger listeners who may have taken minimal evidence as a clear sign of guilt. Due to a number of missing details regarding her methodology, it is difficult to say why Gray did not find any other significant effects. It could be that this is evidence that accent discrimination is not an automatic occurrence, and requires additional variables to make a person be viewed more negatively, such as race, religion, etc.

Maerowitz (2014) looked at British South African English, GAE, and AAE for an undergraduate honours dissertation for the University of Arizona in the United States. He did not give a definition of British South African English, a term he appears to have coined himself, but does note that in his pilot study participants referred to this guise simply as "British." He took his stimuli from the American TV show "Judge Judy," using a domestic argument case and a clip of just over a minute of the male defendant telling his side of the story in AAE. Rather than using the matched guise method, Maerowitz opted to have authentic dialects. He transcribed the audio from the TV defendant and then gave it to two speakers, one speaker of British South African English and another of GAE, instructing them in how to match the intonation, pauses, and other suprasegmental features of the defendant. In this way Maerowitz created his stimuli of three audio clips with an identical testimony. He used an online questionnaire and recruited 60 participants from across the United States using Amazon Mechanical Turk, gaining an even split of male and female participants, and having 20 participants per condition. After listening to one of the

three audio clips, participants answered questions relating to perceptions of honesty, credibility, guilt, etc. on a 5 point Likert scale. They also gave a recommendation on punishment.

Maerowitz found that overall on perception questions, British South African English ranked most highly, followed by AAE, with GAE being least favoured. This result is a surprising finding, in that the standard appears to be perceived most poorly, and thus is directly against SLI. Despite this, when it came to a question on recommended punishment, British South African English and GAE had equally high numbers for "no punishment," whereas AAE only had three participants out of 20 recommend "no punishment" and a few who were "unsure" and wanted more information, something that was not deemed necessary with the other dialects. Maerowitz suggested that the initially surprising perception rankings of the dialects could have been due to participants seeking to compensate for their responses to the punishment question, where AAE was clearly treated much more poorly.

There are a few notable issues with Maerowitz's (2014) and Gray's (2010) studies. Because AAE is associated with a different ethnicity than GAE, it can be argued that Maerowitz's results could be due to race or ethnicity rather than language. In order to study ethnolects, it must first be established that the ratings are due to the dialect and not the ethnicity, as Frumkin (2007) was able to determine through her methodology. Another issue is the lack of the matched guise method. Maerowitz (2014) chose not to use it at all, while Gray (2010) only partially utilised it, as she had two voice actors for the three accents. This leaves the possibility open that participants reacted to other linguistic cues such as a difference in vocal pitch between speakers. Both also chose to use only audio as their stimuli. As discussed with regards to previous studies, limiting participants' senses to just sound may severely inhibit how ecologically valid the study is. Neither mentioned taking steps to mask their interest in the linguistic variables, and therefore may have accidentally primed their participants, lessening the ecological validity of their results. Finally, as these are both non-published, student dissertations, their methodologies and findings ought to be treated with more caution as they may not be directly comparable to the prior published research listed. However, their efforts show further interest in the subfield of accent discrimination in legal settings and still add to the ongoing discussion on this topic.

The studies discussed in this section have all contributed to further understanding accent discrimination within a simulated courtroom context. However, the majority of the studies mentioned used only university students as their participants. This is a widespread and convenient methodological decision, but it leads to a subject sample that is not an accurate representation of the make-up of any jury, thus potentially threatening the validity of the results. As many studies rely on university (and in particular undergraduate) student participants, it has long been a concern that this could be skewing results. Bornstein (1999) and Bornstein et al. (2017) tested the effect of using only university students in comparison to using a more demographically diverse participant base, specifically in regards to mock juries. Both studies report no significant differences between student only mock juries and more demographically diverse mock juries. Due to this, the use of university students in these studies may not be the issue it once would have seemed, and may not remove any significant ecological validity in that respect.

Beyond the use of university students as participants, a few things ought to be taken into consideration. Only Frumkin (2007) attempted to collect data from groups in order to gain a better understanding of how a jury, rather than individual jurors, would react to dialectal differences. As previously noted, however, her group sizes were inconsistent, making it difficult to directly compare them. Frumkin was also the only researcher who used female, rather than male, guises in her study. While the consistent use of male guises in the other studies make them more comparable, it does show a significant gap in research on this topic regarding the effect of women versus men in the courtroom.

Another issue in many of the studies conducted is found in the dialects the researchers chose to study. Often foreign accents or ethnolects were chosen to compare against the national standard dialect. This leaves open the possibility that judgments were made based on stereotypes associated with nationality or ethnicity rather than the linguistic qualities of a person's speech. Additionally, regardless of the variety chosen, not everyone employed the matched guise method, again leaving open the chance that judgments were made on something other than the dialectal features. While people may very well make decisions based off of national or ethnic stereotypes, or on voice quality in addition to dialectal features, because these studies were testing the effect of dialect, those other variables ought to have been controlled

for. With the exception of Sobral Fernández and Prieto Ederra (1994) and Gray (2010), those who did use matched guise only altered the phonetics features of the guises, and therefore might have relied to some degree on Standard English grammar in their scripts. This would suggest that they depended on the standard dialect to an extent in both standard and non-standard guises and is therefore misleading to call the guises used dialects rather than accents.

Except for the student dissertations, all of the published studies used only students as participants. While this may not have skewed their results according to Bornstein (1999) and Bornstein et al. (2017), none of the researchers acknowledged in their results that students represent a single demographic, not the entire populace. Additionally, many of them had a much higher proportion of female participants than male participants, showing another way their participant demographics could skew the results. Another issue is that many of the studies used only an audio stimulus rather than one that presented visual information as well. Since juries are generally able to not only hear but also see the speaker testifying, the lack of video lessens the ecological validity in those studies. Beyond that, each stimulus was only a small representation of the amount of information juries would receive in real trials, therefore lessening the ecological validity. Finally, in all of the articles above, no researcher mentioned actively attempting to hide their interest in linguistic issues from their participants. This could have primed participants, possibly making them more aware of their own language attitudes, and thus potentially skewing the results.

3.3. Conclusion

George Zimmerman was acquitted of second-degree murder in the fatal shooting of Trayvon Martin (*Zimmerman v. State*, 2013). It would be erroneous to suggest that this verdict was returned solely due to the dismissal of Jeantel's testimony. There were many variables in the trial that led to Zimmerman's acquittal. It is furthermore clear that Jeantel was also viewed as an unreliable witness due to more variables than her speech, such as her weight, her apparently low literacy level, and her reluctance to testify (Carlin, 2016; Thompson, 2015). However, she was the prosecution's key witness, and the juror who spoke to the media specifically mentioned her dialect as a factor in their decision to disregard her testimony

(Thompson, 2015). Therefore, there is evidence that her non-standard dialect did have a crucial impact on the outcome of the trial.

Accent discrimination studies have the potential to impact fields beyond linguistics and affect lives. In U.S. court cases, it rarely occurs to either the judge or the jury that racial voice identification could be founded in prejudices (Kohler, 2004). Similarly, they are frequently unaware that their impressions of a defendant or witness are at least partially based on his/her accent due to the stereotypes attached to that variety (Frumkin, 2007). Yet it is not known whether the court cases described in this dissertation are typical and part of a regular and systematic discrimination against non-standard speakers or if they are instead outliers. Therefore, research on attitudes regarding non-standard dialects is important, and it is necessary to learn when, and how regularly, those attitudes cause discrimination.

Chapter 4. Research Questions

Through the court cases outlined in section 3.1, I have argued that accent discrimination, while not an automatic occurrence in every situation, has the potential to affect trial outcomes. In conjunction with this, the studies discussed in section 3.2 all sought to determine when linguistic stereotypes are triggered and the extent to which they can affect jury perceptions and verdicts. Through the use of various methodologies across three continents, it can be argued that they have begun to establish that accent discrimination is a possible variable within courtrooms. However, further research is needed to fill the issues their studies left and continue gaining a better perspective on this phenomenon before any recommendations can be made regarding policy changes in order to mitigate the effects of accent discrimination in courtrooms. The research conducted for this dissertation seeks to address some of those flaws and add to the growing body of knowledge on this topic.

I chose to conduct experiments in both the United Kingdom and the United States. Because prior research does span such a large geographical spread, I wanted to establish how valid it was to use the results of a study from one country and apply it in another country. There are inherent differences in the culture and the justice systems involved which could lead to incompatibility, making any comparison faulty and unreliable. As the U.K. and the U.S. have had a long history of contact and cooperation, they are similar enough to compare their results with some degree of reliability. By running comparable experiments in both countries and analysing whether the data shows the same trends, it may be possible to say whether studies like Dixon, Mahoney, and Cocks (2002) and Frumkin (2007) can be reviewed jointly or must be kept in the context of their respective countries. Broadly speaking, it

could offer evidence towards how generalizable studies are and to what extent researchers need to consider specific cultural issues and differences.

The accents used in both the U.S. and U.K. versions of this study have been well described, both phonetically and in terms of documented evaluations, in published research, listed in chapters 5 and 6. With this information, I have independent sources on which to base my predictions. It also helped me choose U.K. varieties that would in some way be comparable to those used in the U.S. version of this study, consequently allowing me to test the results between the two versions. In the U.K., I used Yorkshire English (YE) to compare against Standard Southern British English (SSBE). Southern American English was chosen to compare against GAE in the U.S. version. I chose regional accents that held contrasting stereotypes from the national standard accent. SSBE and GAE are both the standard accents of their respective countries and thus can be compared with regard to "standardness" and the high levels of overt prestige attached to that concept (Coupland & Bishop, 2007; Heaton & Nygaard, 2011). YE and Southern American English are comparable as they both have lower levels of overt prestige but are stereotyped as having higher levels of characteristics relating to solidarity and covert prestige – although it must be noted that Southern American English may have slightly more negative perceptions than YE (Tamasi & Antieau, 2015; Wales, 2006). Both non-standard dialects are associated with a lower class than that of the standard dialects. Finally, both regional accents are well recognised varieties within their respective countries (Allbritten, 2011; Hughes, Trudgill, & Watt, 2012), an important consideration as there is then a higher likelihood that participants will be familiar enough with the variety to also hold specific language attitudes towards that variety.

In addition to considering the specific stereotypes associated with the varieties I chose, I also sought to minimise potential confounding variables. Within the U.S., research findings suggest that race is one of the strongest extra-legal influencers in a courtroom (Esqueda, Espinoza & Culhane, 2008). As I was seeking to discover what effects *language* has in a courtroom, I did not want to use dialects that were associated with differing ethnic groups, as any significant result could be attributed to the variable of ethnicity just as much as it could be correlated to language. Due to this, I chose accents that were not associated with any specific minority group in order to avoid confounding variables such as race, nationality, etc.

(Hurwitz & Peffley, 1997). All four accents are associated with those who are ethnically white, the majority ethnic group of both countries, and the non-standard varieties are both associated with a specific region of their respective countries, rather than a specific community group. Therefore, while these varieties do still index a specific ethnicity (white), it is the same ethnicity between all four, and no minority group is signalled out. This made comparisons between the non-standard varieties and the standard varieties more viable, as GAE and Southern American English, as well as SSBE and Yorkshire, would not be likely to be perceived differently due to race or ethnicity. Through all this, I sought to determine whether ratings were based on linguistic stereotypes due to regional (and class) differences, rather than national or minority ethnic biases.

I explicitly labelled both the U.S. and U.K. studies as psychological research, telling participants that I was looking into jury decision making. I did not include any references to linguistics or even the department I was a part of, instead only mentioning the universities I was working with in all the recruitment advertisements, participant information and consent sheets, and any other forms. This allowed me to mask my interest in the accent variable and not accidentally prime participants for the linguistic differences, thereby achieving greater ecological validity.

When collecting data from the U.K., due to time and financial constraints, I chose to collect individual responses only online through Qualtrics Survey Software. I was still able to directly compare these against the U.S. individual responses even though I did not have groups from the U.K. to compare against the U.S. ones. Due to the U.S. being my main interest, the U.K. study was set up in part as a large scale pilot for the U.S. version. In that way I was better able to test every aspect of the study, from putting it together to how well the questions gathered the desired information from participants, to discovering any issues in the Qualtrics software itself that would require time to fix. After gathering all the U.K. data, it was decided that there was enough information for it to be a full study rather than just a large pilot, and would allow direct comparison between U.K. and U.S. language attitudes.

As for the U.S. version of this study, I set up the study on Qualtrics in order to gain responses spanning the whole country. However, as this would only give me individual responses, I also travelled to multiple locations across the U.S. to run the

mock trials with focus groups. As the stimulus and questions were exactly the same, the only differences being whether participants deliberated in groups or individually online, this would allow me to test whether deliberating as a group had a significant effect on the results. If not, then previous studies that only used individual participant responses could be regarded as more ecologically valid and generalizable to group contexts such as jury deliberations. It would also make future data collection less strenuous. However, if there are statistically significant effects, it could cast doubt on the reliability of those same aforementioned studies' results.

4.1. Pilot Studies

In order to better inform my decision making processes as I developed my studies, I ran two pilot studies. This allowed me to test various methodologies and guided me in knowing which issues to focus on covering in this research. The first pilot study tested the effect of gender. Because all except one of the previous studies used male guises in their stimulus, I sought to learn whether it would be worthwhile to add the gender of the defendant as a variable in my larger study. The second pilot study tested the effect of audio versus video stimuli as many of the previous studies had only used audio. In both pilot studies I told the participants it was a psychological study, a small methodological detail I later used again in my full studies, in order to avoid priming them for the linguistic variable.

4.1.1. Gender Effect Pilot Study

The first pilot looking at gender had four conditions: male GAE speaker, male Southern American English speaker, female GAE speaker, and female Southern American English speaker. I had two voice actors, both in their 20s from the Midwest of America who were able to produce both a GAE and Southern American English guise. Their guises were marked as authentic by 28 naïve American raters who listened to the audio. The term "naïve rater" refers to people who have no knowledge of any other part of the study and therefore are able to approach it from a neutral and objective perspective

This pilot study was set up as a psychological study focusing on memory. Participants were told the study was investigating whether they would be more likely

65

to remember a list of words if they read the list or heard the list read aloud to them. I created two word lists, seen in Tables 1 and 2, that contained a number of the phonetic differences between GAE and Southern American English. Following are the word lists and justifications for their inclusion:

Word List #1			
White	Dating	Sat	Heat
Side	Shirking	Led	Tray
Nine	Caring	Chin	Row
Bike	Fielding	Mole	Noon

Table 1. Gender Effect Pilot Study - Word List #1

Word List #2			
Fight	Bating	Cat	Sheet
Tide	Working	Said	Pay
Mine	Faring	Thin	Go
Like	Yielding	Roll	Moon

Table 2. Gender Effect Pilot Study - Word List #2

- **Columns 1:** the vowels are diphthongs [ai] in GAE but lengthened monophthongs [a:] in Southern American English.
- **Columns 2:** GAE pronounces the –ING variable as the velar nasal [ŋ] while Southern American English pronounces the –ING variable as the alveolar nasal [n].
- **Columns 3:** the vowels are monophthongs in GAE but diphthongs in Southern American English.
- **Columns 4:** each word uses a vowel not yet pronounced in the previous columns in GAE. This was done due to the vowel shift currently occurring in Southern American English. It has not yet been found to be consistent or regular (Allbritten, 2011) and I left it open to my voice actors how they wanted to pronounce it. They pronounced *sheet*, *go*, and *moon* the same in both GAE and Southern American English, but they said *pay* with a diphthong in GAE and with a monophthong in Southern American English.

Participants read Word List #1, but heard Word List #2 in either GAE or Southern American English as a part of the pseudo memory task.

In addition to hearing a word list, participants also listened to the instructions in an audio clip. The speaker introduced him or herself as the researcher, thanked the participants for taking the time to be a part of the study, and then explained how the memory task would work (the script is included in Appendix I). In this way I was able to have the participant listen to the audio stimulus twice, once when given the instructions and a second time when they listened to the word list read aloud. I chose to have the speaker claim to be the researcher in order to attach the notion of a real person behind the voice. Thus participants would not have the emotional distance of simply judging a disembodied voice, but rather a person they have been told will be seeing all the results and feedback due to being the researcher. Additionally, it gave each speaker the same background and level of education, which meant that any differences in perception would be a direct result of gender and/or accent.

Data was collected from 65 participants across the U.S. using Qualtrics Survey Software. Participants were randomly assigned one of the four guises, and after completing the memory task were asked questions relating to the reliability and capability of the speaker they heard. My hypothesis was that the female speaker would be perceived more poorly than the male speaker, due to the anecdotal experiences of women outlined in chapter 2, as well as the research that has found women are likely to be perceived more poorly than men in professional settings (Eagly, 2007; Powell, 2018). Additionally I hypothesised that the Southern American English guise would be perceived more poorly than the GAE guise, due to the SLI literature discussed in chapter 2 which suggests that non-standard dialects will be perceived more poorly in formal settings. Overall, this would result in the female Southern American English guise being perceived most negatively and the male GAE guise most positively.

After being evaluated by three naïve raters (again, people who have no prior knowledge or involvement with the study), the comments and ratings left by participants were separated into a "neutral" category (participants did not feel the guise had any effect on their ability to complete the task) or a "negative" category (participants saw it as actively hindering their ability to complete the task); there

were no overtly positive responses. The overall results of this can be seen below in Figure 8 and are shown in further depth in Table 3. As predicted, GAE was rated more highly than Southern. Figure 8 would also suggest that the male guise was rated more highly than the female guise, again as predicted.
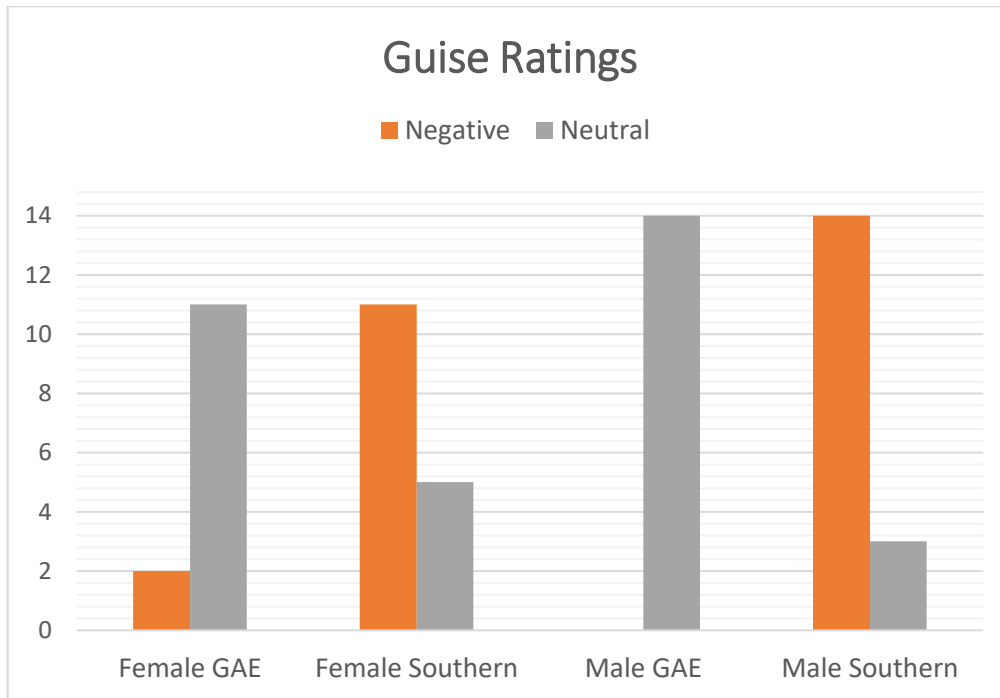
Guise Ratings

Figure 8. Gender Effect Pilot Study - Guise Ratings

However, as laid out in Table 3, there was no statistically significant difference between Female GAE and Male GAE, nor was there any significance between Male Southern American English and Female Southern American English. In light of this finding, I chose to use only male speakers in the full version of my study in order to make it more comparable to previous studies' results.

Female GAE did differ significantly in its ratings from Female Southern American English by $p < 0.05$ and from Male Southern by $p < 0.01$.

	Estimate	Std. Error	z value	P	Sig.
(Intercept)	-1.704	0.768	-2.218	0.026	*
Female Southern	2.31	0.921	2.509	0.012	*
Male GAE	-16.861	1743.248	-0.01	0.992	
Male Southern	3.245	0.997	3.242	0.001	**

Table 3. Gender Effect Pilot Study - Guise Ratings

(Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)

As the neutral/negative dependent variable was tested on a binary scale of 0-1, coefficients below 0 predicted no effect, while coefficients above 1 predicted hindrance. Thus, the coefficients in Table 3 show that, in comparison to the Female GAE guise, the probability that one of the guises would be considered a hindrance increased with both Southern guises. These results would suggest that when directly compared, GAE and Southern American English are treated differently in some contexts. It will be of interest to learn whether this extends to the courtroom and my full study using mock juries. Further analysis into these results can be found in Appendix I.

4.1.2. Stimulus Type Effect Pilot Study

I also piloted the use of a video stimulus versus only an audio one. I used the same methodology of a memory task as the previous pilot study. Due to the lack of differences found in respect of gender in the first pilot, there were only two conditions in this study: male Southern American English audio only and male Southern American English with video. Participants would either see the "researcher" while he gave them instructions and later again when he read the word list aloud, or they would only hear the speaker in both instances. This was a matched guise test and the same male actor from the first pilot study provided the video. I stripped the audio from the video in order to create the audio only condition.

Once again I used Qualtrics to run this pilot study in the U.S. and I collected responses from 25 participants. I hypothesised that participants would be less likely to give negative responses when given the video condition as I believed that seeing the "researcher" would make the person more real to them than simply hearing the

69

person. I expected this would make them more self-conscious in their responses and feedback, and thus less inclined to be negative or harsh. However, after conducting analyses on the results, I did not find any statistically significant differences between the two conditions. Despite this, in my full study I still chose to use video stimulus in order to better simulate the experience of a real juror, and therefore achieve greater ecological validity.

In this way I worked to discover which of the flaws identified in the previous research done on this topic I ought to strive to fill. Further details on the justification for the choices I made in my full study, and information on the exact methods used, can be found in the background and methodology sections of chapters 5 and 6.

4.2. Research Questions and Hypotheses

As previously outlined in chapter 1, below are the questions addressed in this dissertation. The overall question this research seeks to answer is whether the presence of language attitudes leads directly to accent discrimination. My specific RQs are as follows:

Research Question I: Will a defendant be perceived differently by a jury based on whether he speaks in a standard or regional accent?

Research Question II: If a defendant is perceived differently, will this go as far as affecting the verdict?

Research Question III: Does collecting individual responses lead to reliable results on how a jury would function or does the group condition make a large enough difference to invalidate research results that used only individual participation?

Research Question IV: Are the results of the first two research questions similar enough across western nations, such as the United Kingdom and the United States, as to be valid outside of the country the study was conducted in?

RQs I and II seek to directly answer the overall question. RQs III and IV seek to answer questions brought up by the flaws left in previous studies.

Due to the findings of previous studies and the fact that it is possible to identify incidences of accent discrimination in actual court cases, I predicted that juries would perceive the defendant differently based on whether he testified in a regional or standard accent. However, while I thought it was possible this could affect the verdict, I did not expect that it would do so consistently, since accent discrimination is not an automatic or universal occurrence and there may be a range of factors that contribute to it. Rather, I hypothesized that juries would rate the regional accent guises more negatively, and judge the defendant more harshly when it came to recommended punishments, but that this would not always lead to juries also being more likely to find the regional guises guilty. I based this hypothesis on previous work that has claimed the standard is generally preferred over non-standard accents, at least in English (Collins & Clement, 2012). As a courtroom is considered a formal setting, it is possible that it will be even more expected that people speak the standard variety when testifying. Therefore, for them to speak the non-standard is less expected and perhaps less socially acceptable.

The above predictions are simplified. It is just as possible that jurors do not expect *complete* standardness in a courtroom, and that actually the use of a regional dialect could signal a level of authenticity. The type of crime may have an unforeseen effect on my results, as Seggie (1983), Dixon et al. (2002), and Dixon and Mahoney (2004) hypothesised, even if they did not find strong effects on how that would then affect a verdict. As previously mentioned, there are numerous variables that come into play in trials, and these are only two possible ones. Due to this, and the fact that there is not a great deal of prior evidence or information into this topic to build upon, the predictions were simplified and throughout the study I remained open-minded with regards to what results I would find, and what conclusions could then be drawn from them.

Another point of consideration is the possibility of null results. Just as it is possible for a variable other than accent to affect the results, it was equally possible that I would not find discrimination to be present. This would still be a finding of interest, as it would add knowledge to the field regarding when discrimination is not present, and therefore lead to future avenues of discovering which variables must be involved in order to trigger discrimination. Additionally, it would add balance to studies published on the topic. It is clear through chapters 2 and 3 that accent

71

discrimination does exist. However, due to the media reports, court transcripts, and study results those chapters could equally lead to a mistaken implication that accent discrimination is a common and regular occurrence if one speaks a non-standard accent. Among the studies listed, only Cargile (2000) found null results while examining the possibility of accent discrimination in employment. This could be due to the difficulty of getting null results published (Jena, 2017; Munafò & Neill, 2016), or perhaps of all those who have studied accent discrimination, it truly is only Cargile (2000) who has not found evidence of accent discrimination. Regardless of the cause, it is unlikely that accent discrimination is as unavoidable as the previous chapters may indicate. For this reason, although I did predict that the non-standard dialects would be perceived more poorly than the standard dialects, any result is of interest and will further the field.

As for the group verdicts versus the individually rendered verdicts, I predicted that the process of working in a group, even briefly, would significantly affect the results. This is due to the large body of research in psychology looking at group dynamics and the effect they can have in the moment as well as even after the group is dispersed (Bornstein & Greene, 2011; Forsyth, 2017; Greene and Bornstein, 2003). Therefore, I hypothesized that the data I collected from the mock juries would differ significantly from the data collected from individual jurors. Specifically, I predicted that group polarization, where the process of working as a group, and knowing others agree with you, would cause the mock jury participants to become more extreme in their beliefs than the Qualtrics participants who completed the study alone (Myers & Lamm, 1976). Beyond that, however, I did not make any further predictions regarding what differences would appear between the group responses and the individual responses. Because this is not something that, to my knowledge, has been directly compared before using matching methodologies run at the same time, I had no previous literature to suggest in what way the individuals would differ from the groups. Therefore, I remained open-minded to any possible outcome in this aspect.

Finally, I did not expect to find large differences in results between the United Kingdom and the United States. While they do have different cultures and their justice systems are not identical, I predicted that they would not differ significantly in every analysis question. Yet if it was only their verdicts that

distinctly differed, this would suggest that trial outcomes significantly differed and therefore, when considering legal contexts, these two nations would need to be studied separately. Even if this was not the case, however, and geographical location did not have a strong effect, I still felt it was an important variable that had been overlooked in previous research and therefore ought to be considered. On this research question I also remained open-minded to any possible outcome.

The following three chapters will examine and discuss in detail the research undertaken and the results that came out of it. Chapters 5 and 6 will look at the literature, the methodology, and the results from the two countries separately, both addressing RQs I and II, and chapter 6 additionally addressing RQ III. Chapter 7 will then analyse and cross-compare the individual Qualtrics results between both countries, seeking to answer RQ IV. If it is found that the group aspect significantly alters the data, and thus invalidates the results from the individual responses, it is still of interest to analyse the individual data to discover what, if any, differences might exist between the two countries in their potential for accent discrimination with courtrooms.

Chapter 5. UK Mock Juror Study

5.1. Introduction

This chapter will detail the research I conducted in the United Kingdom addressing the potential for accent discrimination within courtrooms. It seeks to answer RQs I and II using UK individual participant responses. The standard variety used in my UK research is SSBE. I chose Yorkshire English (YE) as the regional, non-standard variety due to its contrasting stereotypes with SSBE and lack of association with a minority ethnic group. In order to answer the RQs, videos of a cross examination between a prosecuting lawyer and a defendant were used as the stimuli and mock jurors were recruited to make a verdict.

The initial section outlines the literature review conducted on the accents used and their respective stereotypes. This is followed by a description of the methodology with justifications and explanations for each of the steps taken. The results have been analysed and laid out before discussing what they could imply for the larger picture. The defendant testified in either SSBE or YE, depending on which video the mock juries watched. I expected that when testifying in YE, the defendant would be perceived more poorly by the jurors. While this would not necessarily lead the jurors to give significantly more guilty verdicts to the YE guise than the SSBE guise, I also predicted that the suggested punishments would be harsher for the YE-accented speaker.

The following section is a brief literature review of these two accents, where they are spoken, their most commonly used and distinctive features, and the indexical linguistic stereotypes associated with them. As with any discussion of stereotypes, it must be understood that these are by no means held to be true by

everyone. Additionally, there is no evidence that everyone who does believe these stereotypes then acts on them in any discriminatory manner. Rather, a discussion of stereotypes allows for a generalised understanding of how popular culture may portray these language varieties and what biases participants may approach the study with.

Furthermore, the variables and language attitudes reported here are those that have been generally found across numerous studies over multiple decades. Eckert (2008) rightly notes that one variable can carry multiple indexical and social meanings, and SSBE and YE are no exception to this. Different people in varying contexts will ascribe distinct and separate meanings to these language varieties. Therefore, it must be understood that the discussion on these dialects that occurs in the following section is a generalisation based on what research has found over the years, rather than dealing with the specific nuances of a single time in a single context.

5.2. Yorkshire English and Standard Southern British English

Yorkshire is the largest traditional county in England, covering roughly 1/8 of the whole country. Even after it was formally split into four smaller counties in 1974—North Yorkshire, West Yorkshire, South Yorkshire, and the East Riding of Yorkshire—North Yorkshire remains the largest county in the country (*Local Government Act 1972*). Today, however, while locals might use the cardinal directions to be more specific regarding what part of Yorkshire they live in, most would still consider it one county and themselves all "Yorkshiremen" or "Yorkshire people." Even in academic writing Yorkshire is often considered a single county, and strong claims are made about it, such as Yorkshire "typifying 'Northern-ness'" and being "the most famous county in England" (Fletcher, 2012, p. 227). Thus I chose to similarly refer to these counties collectively as "Yorkshire," and the accent described here is broadly YE without the supra-regional phonetic markers. Just as there are multiple Yorkshire counties, there is also a great deal of variation in accent across the regions and cities of Yorkshire (Hughes, Trudgill, & Watt, 2012; Petyt, 1985). However, I chose to use only the phonetic features noted across the majority of Yorkshire and not any of the restricted features specific to a particular location.

Accents in Northern England are distinguished from southern varieties through several phonetic features, mainly in the pronunciation of vowels. Northern English uses the short [a] vowel in words such as *bath*, *grass*, and *laugh* rather than the southern long [ɑ:] (Evans & Iverson, 2004). Northern English generally pronounces the FACE vowel as a monophthong, typically [ɛ:] (or near the third cardinal vowel) instead of the standard diphthong form of [eɪ] (Williams & Escudero, 2014). In Northern English the monophthong GOAT vowel [ɔ:] is produced rather than the standard diphthong [əʊ] (Watt & Tillotson, 2001). Additionally, production of the happY feature is lax in some areas in the north, including Yorkshire, so the vowel at the end of words such as *city*, *funny*, and *coffee* is [e], or [ɛ] rather than the more standard [i] or [ɪ]. However, this is only produced when the feature occurs utterance finally. Furthermore, this feature has been found to be strongly indexed with social class. Research has noted that it is stereotyped as being used by working class speakers and those who are "common." Possibly in response to this, it has been observed that some Yorkshire speakers, in particular middle class women, have been moving away from producing a lax back vowel and instead using the diphthong [eɪ] (Beal, 2008; Kirkham, 2013).

In many varieties including SSBE, there is the split of the /ʊ/ vowel into /ʌ/ and /ʊ/ vowels, creating a FOOT / STRUT vowel distinction. While many English varieties have adopted this split, Northern England maintains the conservative feature [ʊ], leaving no difference in vowel between words like *luck* and *look* or *putt* and *put* (Evans & Iverson, 2004; Williams & Escudero, 2014). In some instances [ə] or [u:] are used in addition to [ʊ], thus allowing for some distinction between words. An example of this is *book* [bu:k] and *buck* [bʊk]. However, as this is not a feature used regularly across Yorkshire and Northern England (Hughes, Trudgill, & Watt, 2012; Petyt, 1985), this research focused on the general phonemic difference of the /ʊ/ vowel in the north and the split of it in the south.

In addition to the features described above that make most northern varieties of British English distinct from ones in the south, there are also a number of features that are specifically associated with Yorkshire. H-dropping, though stigmatised and perceived as a working class feature, has been found around Yorkshire (Beal, 2008). In Bradford voiced plosives can become unvoiced when they are immediately preceding an unvoiced consonant (e.g. "bed time" may be pronounced [bɛt taɪm])

76

while devoiced plosives may be glottalised when preceding an unvoiced consonant (e.g. the town "Bradford" may be pronounced [bræʔfəd] (Hughes, Trudgill, & Watt, 2012; Wells, 1982, pp. 366-367). Finally, secondary contraction is possible in YE, where words such as *can't* and *don't*, are pronounced as [kʰãː?] and [dẽːʔ] (Hughes, Trudgill, & Watt, 2012, pp. 109-110; Richards, 2008, pp. 120-125).

YE, like all accents, is subject to social evaluation (Coupland, 1988). Historical records show that YE and its speakers have been viewed as "other" and substandard since at least the fourteenth century (Hickey, 2015). One notable examples comes from John of Trevisa (ca. 1385), who wrote, "all the language of the Northumbrians, and especially at York, is so sharp, piercing and grinding, and unformed, that we Southern men can that language hardly understand" (cited in Foulkes, 2006, p. 625). Until the eighteenth century a large portion of the north was fairly difficult to reach from the south, and was generally perceived by southerners as wild and uncivilised. The advent of the Industrial Revolution made travel easier, but reinforced notions of "otherness" as the North became associated with dirty mines, mills, and factories, along with poverty in the working class. This negative perception was not then helped when those industries were closed and abandoned in the twentieth century, leading to further poverty (Beal & Cooper, 2015).

Today perceptions of Yorkshire and the language attitudes held towards YE are mixed, with YE receiving higher *solidarity* characteristic ratings but lower *status* ratings. This has been found across multiple traditional language attitude studies spanning multiple decades (Giles, 1970; Hiraga, 2005; Milroy, 2000; Strongman & Woosley, 1967). YE is generally viewed positively with regards to friendliness but negatively in more formal contexts. YE speakers are commonly stereotyped as being resilient, hard-working and friendly, renowned for being open, warm, and hospitable. In fact, YE scores so highly on characteristics pertaining to friendliness that some businesses have moved their call centres and customer services departments to cities within Yorkshire in order to take advantage of those stereotypes (Wales, 2006); although there is also evidence that would imply the companies moved to the north and Yorkshire for economic reasons, rather than any potential linguistic advantages (Cameron, 2012). Regardless of whether this was in fact due to cheaper overheads in the north or the language variety, it shows the prevalence of this stereotype that PR departments would use it to bolster their image.

Due to its history, YE and working class are generally perceived as going hand-in-hand. While this is not negative in and of itself, some people have been willing to take elocution lessons in order to learn how to use the standard variety to increase their professional standing (Wales, 2006). YE has been associated with a lower education level (Giles, 1970). This could be due in part to the Standard Language Ideology held in regards to education, believing that a "good" education would include learning to write *and* speak using the standard. This in turn may further reinforce the stereotype that YE speakers predominantly belong in the working class rather than in professions that require a high degree of education (Beal & Cooper, 2015).

SSBE has contrasting stereotypes to that of YE and is fairly easily distinguishable as a different variety of English in its linguistic features (Hughes, Trudgill, & Watt, 2012). SSBE uses the standard features listed above, of the long [ɑː] in the BATH vowel, the diphthongs for the FACE and GOAT vowels, happY tensing, and the FOOT / STRUT vowel distinction (Evans & Iverson, 2004; Williams & Escudero, 2014). SSBE syntax generally follows Standard English syntactic constructions (Cheshire & Milroy, 2013). As with all dialects, there is variability in the manner in which SSBE will be produced, as speakers—largely dependent on their origin, class, and age—will vary in the way they produce it. For instance, it has been noted that upper class speakers of SSBE may have less aspiration in their pronunciation of word initial voiceless plosives than speakers of SSBE from other classes (Hughes, Trudgill, & Watt, 2012). However, the features that are of interest are the ones listed above, as they are salient (unlike aspiration) and act as indexical markers regarding *status* characteristics due to the perception of these features being "standard" (Cheshire & Milroy, 2013; Evans & Iverson, 2004; Hughes, Trudgill, & Watt, 2012; Mugglestone, 2003; Williams & Escudero, 2014).

For as long as the north of England has been perceived as "rough" and "other" there has been a pervading preference for the south and anything associated with it, including the accent (Hickey, 2015). While the north was seen as uncivilized, the south was often painted in literature as the land of the genteel, with rolling green hills, thatched cottages, and good health (consider *North and South* by Gaskell published in 1855). The south was where "true" English culture and society was retained. Due to this, it follows that the variety of English spoken in the south would

have these same attributes attached to it. The accent became associated with education and the way any proud and "true" Englishman would speak (Wales, 2006).

The prestigious southern variety spoken by the upper class became known as Received Pronunciation and was taught in schools as the "correct" way to speak. Since the 18th century, RP has been the standard variety for England, serving as the perceived uniform ideal, despite intraspeaker variation and the changes it has undergone over the centuries (Cheshire & Milroy, 2013; Mugglestone, 2003). However, recent research has observed a steady decline of RP. It was estimated at the start of the twenty-first century that only 3-5% of the entire English population spoke it, and some linguists expect that number to remain stable at best, if not continue to decrease (Hughes, Trudgill, & Watt, 2012). This could be due to increasingly negative attitudes towards RP. Where once it was perceived as the variety of the elite and therefore highly desirable, it is now often viewed as artificial, stiff, or even a sign that someone is trying to show themselves as socially superior (Mugglestone, 2003; Przedlacka, 2005). As an acknowledgement of RP's changing status and the difficulty in defining its features, some linguists have begun using the term SSBE when referring to the standard UK variety (Hughes, Trudgill, & Watt, 2012), as seen in this dissertation. Regardless of the standard accent title, Standard British English grammar is still taught in schools to native and non-native speakers alike under the oversimplified heading of "English" (Cheshire & Milroy, 2013). In some instances SSBE pronunciation will also be taught to speakers of non-standard varieties, though it is unclear to what degree of success (Davies & Ziegler, 2015). Regardless, this reinforces the notion that the standard variety is how the educated English would speak.

Although SSBE is associated with being a Southern variety, it is equally the *prestigious* variety (as determined by income, education, and/or profession) regardless of which region of England the speaker comes from. This may also be true in Wales, Northern Ireland, and Scotland to varying degrees but overall with a weaker correlation than in England as SSBE is still perceived as a distinctly English variety (Hughes, Trudgill, & Watt, 2012). Language attitude studies have found a distinct correlation between SSBE and high levels of *status* characteristic ratings. Not only that, but SSBE has consistently been rated highly in *status*, above other national standard varieties as well as above regional varieties (Creber & Giles, 1983;

Giles, 1970; Giles 1972; Hiraga, 2005; Strongman & Woolsey, 1967). Coupland and Bishop (2007) found that SSBE rated more highly on both *status* and *solidarity* characteristic traits, though this finding with regards to *solidarity* has not been found in the previously mentioned attitudinal studies. In the past there was an expectation that if speakers had a good level of education or wanted to work in professional, higher paying jobs, they would learn to use RP (and later SSBE). However, the pressure to adapt one's speech to the standard accent in professional environments has been gradually lessening, and increasingly researchers are finding those with a higher level of education will use Standard English syntax while retaining their regional pronunciation (Cheshire & Milroy, 2013).

While the pressure to adapt to SSBE may be lessening, it was due to the general perception of it as a more formal and professional variety that led me to predict it would be rated more highly than YE by participants in this study. I hypothesized participants would consider court a formal setting and therefore would expect formality in any testimonies. Similar to how anyone involved in the trial is expected to dress "appropriately" in formal attire (Brodsky, 2013), it may be expected that testimonies would be spoken in a formal manner. Participants may feel that the speaker is not taking their testimony or court appearance seriously enough if non-standard language is used. The following sections examine in depth this potential for negative perceptions based on the language varieties used as the UK study is outlined and the results analysed and discussed.

5.3. Methodology

This is a mock trial study which sought to answer whether a defendant would be perceived differently by jurors depending on his speech variety, and if so would this go so far as affecting the verdict. A script was created of a cross-examination between a prosecutor and defendant and this interaction was then recorded in a moot courtroom using actors. Two video recordings were made, one where the defendant testified in SSBE and a second where he testified in YE. These recordings were then played to individual mock jurors who were asked to render a verdict and answer a series of questions regarding their perceptions of the prosecutor, defendant, and crime. These responses were analysed using SPSS in order to answer RQs I and II.

While the nature and interests of this study are primarily in (socio)linguistics, the study itself was designed to look like a psychological study of trials and jury decision making. This was done, following the successful pilots outlined in chapter 4, in order to ensure participants would not be primed for the linguistic components. Additionally, this allowed me to test covert attitudes. It was important participants be unaware of the linguistic concerns of this study so that any reactions with regard to the voice guises were more likely to be genuine than if they had been consciously activated. For this reason participants were told that the researcher was interested in how jurors conduct their decision making and the effectiveness of the prosecutor's examination style. This explanation allowed me to draw participants' attention elsewhere, and in so doing mask my interest in their perceptions of SSBE and YE, and whether their attitudes towards the defendant altered depending on which accent he testified in.

5.3.1. Accent Guises

As discussed earlier, I chose to use SSBE and YE for the U.K. version of the mock trial study. This was due to the clear phonetic differences between the two varieties, as well as for the contrasting stereotypes and levels of prestige associated with them, as described in section 5.2. They were also suitable for this study as neither are associated with a specific ethnic group. The presence of such associations could bring confounding variables into play, particularly racial prejudice, and thus make it more difficult to determine whether attitudes were formed based on the speech or perceived race (Hurwitz & Peffley, 1997). Additionally, these two accents serve as a suitable example of the cultural North/South divide (Wales, 2006). As noted in the previous section, there is considerable variation between speakers of the Yorkshire counties. However, the supra-regional features were not included in the accent guise, enabling me to focus on a more general "Yorkshire English" accent for the study.

This study followed the technique developed by Lambert et al. (1960) and employed the matched-guise method. Using this technique meant it was possible to control for any major differences in voice quality between SSBE and YE, and also to avoid any discrepancies in the mannerisms or appearance of the actor. Consequently,

any significant differences noted by participants could be attributed solely to phonetic, or pronunciation, differences. I sought two actors to create my stimuli. The first actor played a SSBE speaking prosecutor. The second actor played the defendant, testifying in SSBE and YE to create the two guises. Both actors were middle-aged white men and were currently living in the York area. The defendant actor spoke SSBE natively and the prosecutor actor spoke modified near-SSBE. Neither man was a professional actor. However, the prosecutor actor had an extensive background in linguistics and training as an expert court witness. The defendant actor had experience in amateur acting and I, along with my supervisors, coached him on YE features. After creating the stimuli, the defendant actor's YE guise was tested for authenticity by 20 participants, recruited through colleagues, and all labelled it as a type of YE accent and considered it natural.

Ideally the defendant actor would have been a bi-dialectal speaker, using both SSBE and YE natively, or near-natively. However, after three months of searching, no one was identified who was able to produce both guises natively and was willing to be involved in this study. One person who was initially identified and approached was personally willing. However, as they were a part of an actor's guild, they were unable to assist at a fee I could afford due to a pre-existing contract. A second person was approached, but they were uncomfortable with being involved and declined (it was never made clear what about the study made them uncomfortable). Due to this, we approached the man described above, as he was willing, had acting experience, and with some coaching was able to produce a believable YE guise.

The study had two conditions: in the first condition the defendant testified in SSBE and in the second condition the defendant testified in YE. Apart from the change in accent, the testimony otherwise remained nearly identical between the two conditions. In both conditions the prosecutor cross-examined the defendant using SSBE. The actors were given a script, included in the following section, which was used to control content. They were asked to generally follow the script, but allowed some spontaneous style in order to make the speech sound natural, and not in fact scripted. Because of this, there were a few small differences in word choices used by the defendant actor in each of his guises. However, the story and overall testimony stayed the same between the two conditions. Overall, Standard English grammar was used in both guises. When speaking in the YE guise, the defendant actor would

occasionally use "were" instead of "was" (e.g. "I *were* driving home"). However, this was not consistent and otherwise only the phonetics were altered between the guises. Therefore, from this point onward, they will be referred to as *accent* guises rather than dialects.

The YE phonetic features present in the YE accent guise are listed in Table 4 and examples are all broadly transcribed. Of the features noted in section 5.2, the YE accent guise had the short [a] vowel, the long monophthong GOAT vowel [ɔ:], and no FOOT/STRUT split, thus producing only /ʊ/ in those cases. The FACE vowel was inconsistent, in some instances it was the monophthong variant and in other instances the actor produced the diphthong variant. There was not an instance in the stimulus where there was an environment appropriate to produce a lax happY utterance-final feature, and so this feature is missing from the phonetic analysis. No h-dropping or secondary contractions were evident in the accent guise. Through the features noted here that the actor was able to produce consistently, it is clear that he was able to produce a northern accent. When asked, participants were also able to note that he was northern, with many placing him in Yorkshire. Thus, while the accent guise produced is not a strong Yorkshire accent, it can still be labelled a Yorkshire accent, as supra-regional markers are absent but it is still identifiable as both northern and Yorkshire. The marked northern and Yorkshire features that the actor produced for the YE guise were absent from his production of the SSBE accent guise.

Within the guise, there is intraspeaker variation evident. Although this did not affect the perceptions of the guise, and is a common phenomenon for speakers to vary in their own speech, it is still worth noting. For instance, in the word "hitting," the actor varied in his production of intervocalic /t/. In one occurrence at the start of the stimulus he pronounced the glottal stop variant and in the second occurrence towards the end of the stimulus he pronounced the voiceless alveolar plosive variant. A second instance of variation was when he said "happened." In the first occurrence he produced it as [hapnd̥], with the final alveolar plosive devoiced. In the second occurrence of this word the actor produced [hapənd], adding a schwa between the voiceless bilabial plosive and the alveolar nasal. The third time the actor produced this word it was [hapnd], without a schwa. In both this and the second occurrence the final alveolar plosive was not devoiced.

83

Yorkshire English	YE Accent Guise	
	Production	*Example*
Short [a] vowel	[a]	"afternoon" [aftəˈnuːn] "fast" [fast]
Monophthong FACE vowel [ɛː]	Inconsistent	"came" [keɪm] "braked" [brɛːkt] "brake" [brɛːk]
Long monophthong GOAT vowel [ɔː]	[ɔː]	"home" [hɔːm] "don't" [dɔːn] "know" [nɔː] "no" [nɔː]
Utterance final happY-laxing	*Environment not present*	*n/a*
No FOOT / STRUT split [ʊ]	No split	"just" [dʒʊst] "enough" [ɪnʊf]
H-dropping	No H-dropping evident	"home" [hɔːm] "hitting" [hɪʔɪn], [hɪtɪn] "happened" [hapnd̥], [hapənd], [hapnd]
Secondary contraction	No secondary contractions evident	"don't" [dɔːn] "wasn't" [wɒzən]

Table 4. YE Accent Guise Phonetics Features

5.3.2. Mock Trial Stimulus

The scenario for the trial in this study comes from the U.S. court case *Haley v. Brown* (2006). On August 27, 2001 eight year old Joshua Haley was playing on his bike with two friends of a similar age. The driveway they were cycling on had tall shrubbery which obstructed the view of it from the road. Haley's two friends cycled on to the road, in front of Gary Brown who was driving a small lorry. He swerved and avoided hitting them. However, Haley then also cycled into the road. In the process of swerving to miss the first two boys, Brown hit and killed Joshua Haley. When it came to an inquiry and brought to court, it was found that Brown had

been following the posted speed limit and therefore was without fault. Due to this the judge acquitted him. This ruling was appealed by Haley's family and again brought to court, resulting in the case *Haley v. Brown*, in which Brown's acquittal was ultimately confirmed.

Haley v. Brown was suitable for the purposes of this study as the case dealt with a vehicular incident, which, unlike other types of crime—such as embezzlement or store robberies—is not racially or socially stereotyped (Hurwitz & Peffley, 1997). The victim involved was a young boy on a bicycle, which also does not appear to have strong stereotypes associated with it. This was a necessary consideration as I desired that participants approach the crime from as neutral a position as possible, removing other potentially confounding variables.

This case also fit the requirements for this study as it could be easily adapted in order to allow juries to "justifiably" give either a verdict of guilty or not guilty. As I was testing whether accent can make a difference in how a jury perceives a defendant, and the potential to affect the eventual verdict, it was necessary that the case not be clear cut one way or another. Due to this, I altered some of the information from the original court case. I added the detail that a speed camera had caught the defendant going 11 miles per hour (mph) over the posted speed limit. However, the speed camera was five miles before the scene of the crime, therefore leaving the possibility that the defendant's speed had changed in the interim. I conducted a verbal survey, asking drivers on average how much over the speed limit they generally drove. As many of the drivers considered it fairly common practice to drive 10 mph over the speed limit, I chose to say that the defendant had been driving 11 mph over the speed limit. Due to the survey it appeared that 11 mph over is close to what many drivers may do on average, but is pushing the boundaries on what may be considered acceptable. I predicted I would see a difference in results between participants who are drivers and those who are not drivers due to this addition to the case.

Another detail I changed is that I did not have the victim die. Instead participants are told that he went to hospital with "severe injuries," leaving it unclear whether or not he recovered. I did this as I wanted to make the crime less serious. I was concerned that if participants heard that a child had died, they might

automatically find the defendant guilty. However, I still made it clear that an eight year old child was injured, as I wanted it to be a grave enough crime that participants take the trial seriously. I analysed the data to see if parents or participants in caregiving roles responded in a harsher manner than other participants as I predicted this could make a difference.

A final detail I changed was the name of the victim and date of the incident. While the defendant's name was never given to the participants, the victim's name was changed from Joshua to Peter and the date of the incident was altered from August to April. I did this to protect the anonymity of the court case and make it more difficult for participants to find information on *Haley v. Brown*. I did not want them knowing that I had changed details, as I wanted them to believe the trial they were watching was a real one and not staged using actors. I also did not want them to know the verdict of *Haley v. Brown*, as that could have affected their responses in this study. Although I did not expect participants to put in the amount of effort it would take to find the court case, I changed these details so that I could be confident it would not affect the results and impede the research.

Using all of this information, I put together a script for the actors. In order to check the appropriateness of the language used within it, I sent the script to two British barristers. They provided feedback on how barristers form their questions, helping to make the interchange in my script more authentic. I also had it looked over by my (then) supervisors, Paul Foulkes and Andrew MacFarlane, who were able to make some corrections based on my accidental usage of American English phrasing (e.g. "went to *the* hospital" v "went to hospital," "no sir" v "no"). The final version of the U.K. script given to the actors is included on the following pages. Due to the impromptu nature of the interaction, designed to ensure naturalness, minor changes were made by the actors to the script. However, neither I, nor my supervisors who were present throughout the recording, felt that the changes were significant enough to be retained in the official script.

Prosecutor stands to begin questioning defendant

Prosecutor: Could you please describe to the court, in your own words, the
events of the 23rd of April?

Defendant: I was driving home that afternoon in my lorry. As I came down Spencer Street two kids on bikes suddenly rode onto the road. I swerved to miss them. But another kid riding a bike came out of nowhere. I was already swerving to avoid the other kids, there was nowhere for me to go… I braked but…

Prosecutor: So you hit 8 year old Peter?

Defendant: I did everything I could to avoid hitting the kids!

Prosecutor: Because of your actions he went to hospital with severe injuries, didn't he?

Defendant: … I … I don't know what to say … I'm just so sorry it happened…

Prosecutor: Now tell me, at the top of the hill there was a sign with a 30 miles per hour speed limit on it, wasn't there?

Defendant: I don't remember.

Prosecutor: Well, a speed camera – just 5 miles away from the scene of the incident – clocked your speed at 41 miles per hour. That is 11 miles over the legal limit, is it not? Do you recall going so fast?

Defendant: No.

Prosecutor: By your own admission, you didn't brake when the first children came out in front of you, did you?

Defendant: No… I swerved … it all just happened so fast!

Prosecutor: You only began to brake when Peter came out in front of you, didn't you?

Defendant: I was just really focused on not hitting the first kids – I tried to swerve. And then when the lad cycled in front of me, I just slammed on the brake with both feet, but it wasn't enough.

Prosecutor: Sadly, it wasn't. No further questions.

Prosecutor sits down.

Filming of the mock trial was conducted in the University of York's moot courtroom at the law school. It was set up as if occurring in the Magistrates' Court, as I had been unable to acquire a gown and wig for the prosecutor actor to wear. These items are used in Crown Court proceedings, and therefore the lack of them may have raised participants' suspicions regarding the authenticity of the mock trial if I had said it was in the Crown Court. However, juries are not used in the Magistrate's Court. In this instance I was relying on people's general lack of knowledge regarding specific courtroom procedures. As no participant mentioned the lack of juries in a Magistrate's Court as an issue in any of their comments, and in particular their feedback to me as the researcher at the end of the study, I have no reason to believe that any participant was aware of this disparity. Therefore, the setup as a Magistrate's Court and the use of mock jurors does not appear to have been a problem.

Two Panasonic V160 video cameras were used to record both actors simultaneously. I requested lower quality, non-professional recording equipment as I wanted the finished product to look like something that had come from CCTV, and thus perpetuate the impression that the trial clip the participants watched came directly from a real trial (and that I had been given special permission to use it). In addition to the video cameras, I also collected recordings of the audio using radio lapel mics attached to the actors' ties and connected to Zoom H4n Handy Recorders. I chose to do this as a type of backup should something go wrong with the video cameras. I also thought it would make the audio clearer, as while I did not want high video quality, I did want good audio quality. However, ultimately the video camera audio was of a higher quality than the zoom recorders' and therefore both the video and audio used came from the Panasonic V160 cameras.

Prior to and throughout filming, both I and my supervisors provided instructions and guidance to the actors in order to achieve the naturalness we desired. The actors were directed to use the script as a guide, but to speak naturally and adapt the script as needed. Initially, both actors spoke too quickly. They were instructed to speak slower and take time to think through their questions and answers. I did this because in a real courtroom prosecutors take time to consider the best questions to ask and defendants need time to consider the question and formulate their answers. Additionally, I asked them to pause, repeat themselves, and, with regards to the

defendant, even stumble through their words at times. These instructions helped the dialogue to sound unscripted and achieve the perceived naturalness I was striving for in the recordings. Because the defendant actor was not a native speaker of YE, he required some coaching, specifically in the use of the monophthong GOAT and FACE vowels used in YE. He received this coaching from myself and my supervisors. I continued to film takes of the mock trial until everyone involved was satisfied with the results. In total I had nine takes, five of the SSBE cross examination and four of the YE examination.

During the filming, I asked friends to sit behind the prosecutor. I wanted it to appear as though there were some members of the public in the courtroom on the video. Moreover, prior to filming, I was unsure of exactly what angles would be visible in the camera. There are some portions of the moot courtroom at the University of York which look more like a classroom than an official courtroom. Due to this, I needed to find ways to blocks views of those parts from the video recording. I did that through carefully considering the camera angles during filming and having people there ready to act as an obstruction to any of those views. However, with rearrangement of the moot courtroom, this did not end up being an issue, and in the final cut of the videos, only one person was visible at times behind the prosecutor, as seen in Figures 9 and 11.

I edited the footage using Adobe Premiere Pro CC (2017). I went through all the takes of the prosecutor, choosing which sections in which takes were the best in regards to naturalness and perceived authenticity, and combining them to form a single shot. I did the same for the defendant takes, in both the SSBE guise and YE guise. While using this method made it clear that the footage had been edited, particularly the prosecutor's, it was necessary as it allowed me to use the most authentic takes. It would also make it appear to participants that the entire cross-examination was much longer – as a real cross-examination would be – as if excerpts of it had been pulled out and compiled to show to participants. As a result, the clear edits themselves would make the footage appear more credible. The lines within the script above note where the edits in the video footage are obvious. In order to further impress upon participants that the breaks in footage were due to large portions of the examination being taken out, I added a timestamp to the bottom right corner of the

screen, designed to look like many of the timers cameras automatically include. With each break, I had the clock jump to a later time. This can be seen in Figures 9-11.

I then stripped the audio from all three videos in order to edit only the audio in Audacity version 2.1.0. I opened the prosecutor audio as track 1 and the defendant SSBE audio as track 2. Within the prosecutor audio, I erased all instances of speech other than the prosecutor's and did the same with the defendant SSBE audio so that only the defendant's speech remained. However, I chose to keep any ambient noise, with the aim that it sound more like a true courtroom and not a sound studio. Because I could see both tracks at once, I could make sure that the two speakers' dialogue did not overlap. Whenever I moved the audio on one track to make sure it did not overlap, I made sure to make similar edits to the corresponding video so that mouth movements remained synchronised with the audio. Finally, when the speech was correctly aligned in both audio tracks, I combined them into a single audio track. I then went through this same procedure when matching up the prosecutor's audio track to the defendant's YE audio track. In this way, I created an audio file for when the defendant was examined and testified in SSBE and another audio file of the same examination but with the defendant testifying in YE.

Moving back into Premiere Pro, I cropped the prosecutor video and SSBE video so that they could be placed side-by-side in the same screen. I set both to grey-scale and used the "venetian blinds" effect, which causes fine lines to run through the video. This was done to make it appear more like a CCTV recording. With the "fast blur" effect, I blurred the defendant's face in order to add a level of anonymity. Finally, I added the audio created in Audacity combining the prosecutor and the SSBE testimony. This completed the SSBE video stimulus. Once this was saved, I then removed the SSBE audio and replaced it with the YE audio. I was able to use the same video with both audio guises due to the defendant's face being blurred as it was not possible to tell when his mouth did not exactly match up with the words. This created the YE video stimulus, and meant that the only difference between the two videos was the defendant's speech. The videos ran for a total of 1 minute 58 seconds.

Many of these editing decisions can be viewed in Figures 9-11. Additionally, the videos can be watched here:

Figure 9. UK Mock Trial. L: Defendant, R: Prosecutor. First Segment

Figure 10. UK Mock Trial. L: Defendant, R: Prosecutor. Second Segment

Figure 11. UK Mock Trial. L: Defendant, R: Prosecutor. Third Segment

These videos served as the two conditions in this study and the two main independent variables. Through these I sought to answer my research question of whether juries react differently to a defendant based on the accent. Additionally, would that difference in reaction go so far as to affect the verdict?

5.3.3. Ethics

To conduct this study, I completed an ethics application for the Department of Language and Linguistic Science at the University of York. In this form, I laid out my study design and justified the use of deception in leading my participants to believe what they were watching came from a real trial, not a simulated one. As previously mentioned, I did this in order to encourage the mock jurors to take the study more seriously, and thus improve the ecological validity of the results. Additionally, I acknowledged that the topic of an injured child could be hard for some people. To mitigate that, participants would be warned at the start of the study of the trial topic and reminded that they were free to withdraw from the study at any point. I also assured the anonymity of each participant by not collecting any identifiable information through Qualtrics, such as names. With this, the ethics committee approved this study for both the U.K. and the U.S.

5.3.4. Questionnaire

This study was conducted completely online, using Qualtrics Survey Software. This software was suitable for the purposes of this study for multiple reasons. Qualtrics allowed me to recruit participants from across the United Kingdom, rather than from just one region. This software is password protected, so only I have access to the data obtained through it. Thus, participants' information is protected and confidentiality is ensured. Qualtrics allows for complete anonymity, while also providing the researcher with participants' IP addresses; consequently, if desired, I could see where participants were responding from. The official University of York logo was at the top of every page participants saw as they completed surveys, as seen in Figure 12.

Before participants began the study they were presented with an information and consent page, which can be viewed in Appendix II. This was followed by a series of demographic questions regarding their age group, where they were from, and what their profession was. In addition to these questions, participants were also asked if they had any children, how often they drove a vehicle, and whether they had previously served as a member of a jury. These questions were included to gain a better understanding of how participants' personal experiences may shape their perceptions of, and reactions to, the mock trial. As previously discussed, because the trial dealt with an injured child and a vehicular accident, it was possible participants would respond differently given their personal history. Additionally, as this was a simulated jury experience, it was of interest whether those who had previous (and authentic) jury experience would respond differently from those who had no jury experience and knowledge.

I also asked participants which political party they would be most likely to vote for in an election, with the option of refraining from answering if desired. Research has suggested that liberals and conservatives are likely to react differently, and that the process of deliberation actually increases the polarization between the two ideologies (Schkade, Sunstein & Hastie, 2010). This connects back to the psychological concept of group polarization, mentioned in chapter 3. Additionally, some research has noted that those who work in government (and therefore most

likely have strong political affiliations) may reflect that affiliation in their speech (Hall-Lew, Friskney, & Scobbie, 2017; Hall-Lew, Starr, & Coppock, 2012). It is possible that strong political ties or beliefs may also affect speakers' language attitudes. Due to this, I included the question on political stance as it was possible that, just like the other demographic questions, it could account for any potential differences in results between individual participants. All the demographic information collected provided additional independent variables alongside the study conditions. The demographic questions can be found in Appendix III.

After the demographics questions, there were two parts to this study: the mock trial video stimuli and the questionnaire. The mock trial began with a brief explanation and instructions for the participants, seen in Figure 12. The information in this was purposefully misleading, in the hopes that participants would believe the video clips came from a real trial, and therefore give serious consideration to the responses they gave as a potential juror. This was done in an effort to make an online survey more ecologically valid.

UNIVERSITY
of York

The clip you are about to watch contains segments of a cross-examination between a prosecutor and a defendant.

You will act as a mock juror. Please watch the clip carefully, paying close attention to the facts of the case as they are presented. Then answer the following questions, giving them the serious consideration you would if you were actually one of the jurors on this trial.

For purposes of confidentiality, the defendant's face has been blurred. I was given permission to use the following clip for the specific purposes of this research.

>>

Figure 12. Explanation and instructions for participant's role

After the instructions, participants were randomly assigned, by Qualtrics, one of the two conditions. As previously mentioned, this study had two conditions, the SSBE guise and the YE guise, corresponding to the video stimuli. Each condition was an independent variable in the statistical analysis. Due to the way videos in Qualtrics are set up, participants could watch whichever video they were assigned as many times as they pleased before continuing on to the questionnaire.

The first question, immediately following the video, asked participants to render a verdict of either guilty or not guilty. They were then asked, on a three point scale of "unsure" to "somewhat certain" to "certain," how confident they were in their verdict. The following five questions all gave participants text boxes to fully explain their answers, asking them what caused them to reach that verdict, the most and least convincing features of the prosecutor's argument, and the most and least convincing aspects of the defendant's testimony. Images of these questions on Qualtrics can be viewed in Appendix IV. During the statistical analysis of these questions, I paid close attention to whether accent, speech, or language was mentioned in any way as well as for other common themes that might emerge in participants' responses.

The subsequent two questions used a Likert scale of 1-6, with 1 being "not at all" and 6 being "extremely," one of which can be seen below in Figure 13. The first question asked to what extent participants found the prosecutor trustworthy, reliable, and believable. The same was then asked regarding the defendant. The scale was done from 1-6 in order to disallow any true central—and thus neutral—rating, forcing participants to take a stance, however slight, in one direction or another. These specific characteristics were used based on the recommendation of Dr. Tarika Daftary-Kapur, an expert in social psychology and jury decision making. It also allowed for an additional level of comparison with the studies listed in chapter 3 as they also used Likert scales. These questions served two purposes: firstly it helped fulfil the perception that I was interested in the effectiveness of the prosecutor's examination style; and secondly it offered another way for me to collect perceptions on the defendant and test whether these ratings would differ between guises. Questions directly asking about accent or speaking style were purposely kept until the end in order to not prime participants. Therefore, I needed to find other ways of asking about the defendant.

To what extent did you find the defendant:

	Not at all					Extremely
Trustworthy	○	○	○	○	○	○
Reliable	○	○	○	○	○	○
Believable	○	○	○	○	○	○

>>
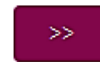
Figure 13. Defendant Likert scale

After this, five questions followed concerning the crime for which the defendant was accused. The purpose of the questions on crime was to determine if, regardless of the verdict participants had given, they would be inclined to answer more or less harshly depending on the guise conditions. Participants were asked a yes/no question on whether they could tell what crime the defendant was being tried for (question included in Appendix IV). They were then presented with a series of options, shown in Figure 14, with the correct answer in English law being "causing serious injury by dangerous driving." However, the term "alleged" was used in order to remind participants that no judgment, beyond the verdict they had given, had yet been rendered. I did not want participants to assume that speaking of the type of crime the defendant was accused of signalled that he was in fact guilty. This question helped monitor participant understanding of and attention to the mock trial.

What type of alleged crime would you consider this to be?

○ Driving without insurance
○ Driving without a license
○ Driving while under the influence
○ Careless driving
○ Dangerous driving
○ Causing serious injury by dangerous driving
○ Unsure

>>

Figure 14. Type of crime

After choosing the type of crime, any participant who had given a guilty verdict was then asked their opinion on what a fitting punishment would be. Their choices were: prison, community service, a monetary fine, or a fill in the blank option. In a real trial, it is the responsibility of the judge to sentence any defendant who has been convicted, not the jury's. However, I included this question as an additional way to measure participants' reactions to the accent guises. I hypothesised that while some participants might not change their verdict due to the accent, they might still react more negatively in other ways towards the non-standard accent guise than the standard accent guise. Allowing them to choose varying levels of punishment was one such way that I could elicit additional covert attitudes from participants. Furthermore, this was a question that had been included in previous studies (as discussed in chapter 3) and therefore added a potential point of comparison between my study and previous ones.

Following the question on recommended punishments, participants were asked a yes/no question regarding whether they associated this type of crime with any particular demographic. If they answered yes, they were requested to specify

with which demographic(s) they associated this crime. These questions can be viewed in Appendix IV.

The final three questions dealt specifically with the accents. Participants were directly asked how they would describe the prosecutor's accent, the defendant's accent, and their own accent. They were given text boxes to answer these questions in as much detail as they desired. These questions were left until the end so that participants would not be primed when they answered any of the previous questions. I was interested in the type of responses participants would give when directly solicited about accent. The question regarding their own accent was included in the hopes of gaining some knowledge of the participant's background and language attitudes when it comes to accents. Depending on how this question was answered, it could serve as a potential baseline to then compare against any other responses they gave concerning speech and accents.

Once participants had finished answering the questions, they were given the option of leaving any additional comments or concerns that may have come up in the process of completing the questionnaire. While I did not expect all participants to fill this in, it gave them the opportunity to mention something that a question did not cover, or more fully explain an answer they had given if they so chose. Participants were then thanked for their participation and given the option of leaving their email address if they wanted to learn more about this study once all the data had been collected and analysed. Although it clearly specified that they would not receive the results of their own responses due to the anonymity requirement, they could learn more about what the data showed as a whole.

I piloted this study for a week and collected information from seven participants I recruited through friends. As it was only a pilot, I was not as careful about finding people who would be completely unaware of the true purpose of this as I was for the main study. Therefore, to err on the side of caution, their data was saved but not used in the final analysis. Instead, their responses were only used to improve the study. Due to comments they left either in the study itself, or later reported back to me, there were a few changes I made to the final version. After being given the chance to think through the court case further, with the assistance of the questionnaire, there were a few participants who wanted to change their verdict

by the end. Due to this, after the questions regarding the type of crime, but before the accent questions in order to continue keeping them last, I added a text summary of the "facts" of the case, included in Figure 15. It was written with the goal of keeping the case balanced, not showing any personal bias, and making either verdict plausible. I then asked if, after additional time for consideration, they would like to change their verdict. Any participant who answered "yes" was then asked what caused them to change their verdict, and given a text box in order to adequately explain themselves. These additional three questions can also be viewed in Appendix IV.

The facts of the case are as follows:

In the afternoon of 23 April, 2013, the defendant was driving home in his lorry. Shortly after a bend in the road, two young boys on bikes rode directly onto the road. The defendant swerved to miss them. However, a third boy then also rode straight onto the road. The defendant hit the third boy, who ended up in hospital with severe injuries.

The posted speed limit was 30 miles per hour. It is unknown at what speed the defendant was travelling at the time of the incident. He was clocked going 41 miles per hour 5 miles before the scene of the incident by a speed camera.

The defendant was accused of the offense of "causing serious injury by dangerous driving" and the case was tried in the Magistrates' Court. The defendant entered a plea of "not guilty," saying that although he did hit the third boy, it was only due to having previously swerved to miss the other two boys, leaving him with nowhere else to go, and thus was not his fault. The prosecution argued that the defendant was speeding at the time of the incident and that he should have begun braking as soon as he saw the first two boys.

>>

Figure 15. Summary of the Facts of the Case

The biggest problem I noted in the pilot data did not come from the participants' responses, but rather from the way Qualtrics collected the data. Because the pages that had the mock trial video stimuli on them were not coded as questions, Qualtrics did not report whether participants had watched the SSBE video or the YE

video. As the videos were the main independent variables, it was necessary to alter the questionnaire in such a way that Qualtrics would report which video participants watched. Due to this, I changed each page with the videos into a question. Directly below each video, I asked participants to indicate how they found the quality of the playback after watching the video, giving them an option of "high quality" or "low quality," as shown in Figure 16. In this way I was then able to see which video participants watched. Furthermore, this question of video quality question could reveal additional information about how participants perceived each video. Once these changes were made I then began to recruit participants.

Figure 16. Updated Stimulus Page

5.3.5. Participants

As I conducted this study during the summer months, I did not have the option of recruiting students solely through universities. It was also important to me to achieve balance in the demographic sampling, and not solely rely on students; thus I sought out alternative approaches to recruiting participants. I utilised my British connections to distribute the Qualtrics survey link. The link was primarily spread through social media (e.g. Facebook, Twitter), with my contacts "sharing" the link across their personal accounts. The contacts distributing the link were aware of the participant requirements – namely that participants be from the United Kingdom or had lived there for a minimum of two years and that they be unaware of the true nature of the study—and circulated the link on these grounds. The popular 2017 Channel 4 show "The Trial: Murder in the Family" had come out only a few weeks before I began sending out my survey. Due to this, I suggested to my contacts that they advertise taking my study in connection to "The Trial," saying that it would give people a similar chance to try out the role of a juror.

However, due to both my actors being based in York and linked to various overlapping social groups in my own network, I avoided recruiting participants from York. Again, I was seeking to have participants believe the trial they were watching was real. Therefore, if anyone knew either of my actors, they would immediately know this was staged. As this could skew my results, I instead chose to have fewer participants rather than take people from York.

The study ran for three months, with a total of 54 participants. There was an even split between conditions, as 27 participants had the SSBE condition and 27 had the Yorkshire condition. The majority of my participants were female, though the margin was small enough that I do not believe it adversely skewed my data, with 33 female participants and 21 male participants. I also asked participants to report their ethnicity, with the option of "prefer not to specify" offered. Two participants were Asian, seven did not specify, and the rest identified as white.

One participant lived in Northern Ireland, three participants lived in Scotland, and the rest were scattered across England. Participants were asked which UK country they lived in, but not the specific region within each country. This was an oversight. If I had asked them, I could have broken down the participants in England into regions, offering more detailed information. Additionally, it would have been of

interest to note if participants from the northern regions reacted differently to the accent guises than participants from the southern regions of England. This is something that could be rectified in future studies.

Further breakdown of the participant demographic information can be seen in Tables 5-7. I had participants in all four age ranges, though 35-64 was the largest age group.

Age of Participant	No. of Participants
18-24	5
25-34	18
35-64	27
65+	4

Table 5. U.K. Participant Age

I had participants in every driving category, from those who never drove to those who drove daily. The largest group was daily drivers, followed by those who drove multiple times a week.

Average Driving Regularity	No. of Participants
Never	7
Rarely	6
A couple times a month	3
A couple times a week	15
Daily	23

Table 6. U.K. Participant Average Driving

Except for UKIP, I had participants in each of the political categories I provided. The majority affiliated themselves with Labour and the rest of the categories were fairly equal in numbers.

Participant Political Affiliation	No. of Participants
Conservative	7
Liberal Democrat	5
Labour	20
Green	9
UKIP	0
Unsure	7
Prefer not to specify	6

Table 7. U.K. Participant Political Affiliation

Lastly, 33 participants out of 54 reported having children and five participants had previously served on a jury.

5.3.6. Analysis

In the analysis of this data, I used a mixed methods design in order to investigate the effects of accent on perceptions of confidence in verdict, characteristics of the defendant in comparison to the prosecutor, verdict of the mock trial, and any recommended punishments. A mixed methods design was suitable as it combined both quantitative and quantitative analyses. Quantitative research data is numerical and is analysed using statistical tests. By contrast, the qualitative approach is suited to gathering exploratory, descriptive data. This data includes the descriptive answers participants gave regarding the accent of the defendant and their own accent.

This study used several types of statistical tests in order to study the relationships between the dependent and independent variables of the study. Prior to analysis, the quantitative data was coded to enable analysis and checked for normality as well as outliers. Normality was checked by inspection of skewness and kurtosis statistics and outliers were assessed through standardized scores. No outliers were removed from the data for the full analysis. Independent t tests were used to test for differences in means between two independent groups. Chi-square tests of

association were used in order to test for relationships with two categorical variables and Fisher's Exact test was used when the categorical variables were unequally distributed. Additionally, multinomial logistic regression tests were used when there were two or more independent variables (i.e. whenever a demographic variable was included with the main independent variable of the video guise). The quantitative data was analysed using R Studio Version 1.1.383 and the qualitative data answers were analysed using NVivo 12 Pro for Windows.

The analysis sought to answer the first two main RQs laid out in chapter 4: Will a defendant be perceived differently by a potential juror based on whether he speaks in a standard or regional accent? If a defendant is perceived differently, will this go as far as affecting the verdict? A list of analysis specific questions was put together in order to address these two main RQs. They can be seen below:

Did the particular video (SSBE or YE) lead to a difference in:

Q1. Participants' verdicts?

Q2. How confident were participants in their verdict?

Q3. Whether participants were more likely to rate the defendant lower in characteristics?

Q4. What level of punishment was suggested?

Q5. How long participants took to finish the questionnaire?

Q6. The quality rating participants gave the videos?

Q7a. Whether participants changed their verdict after reading the text summary.

Q7b. If so, why?

Q8. Whether participants were more likely to offer negative opinions on the defendant?

Q9. How participants spoke about the accent of the defendant?

Questions 1-7a were analysed quantitatively while questions 7b-9 were looked at qualitatively. In addition to these, the demographic independent variables below were included in the statistical tests. These were analysed using fixed effects multinomial logistic regression.

- Gender of participant

- Ethnicity of participant
- Age of participant
- Which region the participant lived in
- Whether the participant had children
- What the participant's politic affiliations were
- How regularly the participant drove
- If the participant had previously served on a real jury

The results of this mixed methods analysis are examined in the following section. The p values are reported for all non-significant test results. The full table of results for those can be found in Appendix VII.

5.4. Results

Question 1: Did participants' verdicts differ significantly depending on the video?

Overall the SSBE guise had more *guilty* verdicts while the YE guise had more *not guilty* verdicts, shown in Table 8 below. This is contrary to the prediction that SSBE would be viewed more positively, and therefore have fewer *guilty* verdicts, than YE. I used a chi square test here in order to determine if there was a relationship between verdict and type of video watched.

	verdict		
	Guilty	Not Guilty	*total*
SSBE	16	11	27
YE	11	16	27
total	27	27	54

Table 8. UK Verdict

Despite the apparent trend of the YE guise having fewer convictions, these differences in distributions of verdict were not statistically significant (p = .174). Additionally, after running multinomial logistic regression tests, the participant demographics were shown to have no significant effect.

Based on analysis question 1 regarding the verdict, RQ II has been answered: there is no evidence that the accent guise caused the participants to perceive the defendant differently enough that it would affect the verdict.

Question 2: How confident were participants in their verdict?

The results relating to this question are summarised in Table 9. There were no distinct differences in confidence levels, though there is an apparent trend in that participants who watched the SSBE video skewed slightly more towards Somewhat/Certain while participants who had the YE video tended to be more Unsure/Somewhat. I used Fisher's Exact test to determine if there was a relationship between level of confidence and type of video watched.

	confidence level			
	Unsure	Somewhat	Certain	*total*
SSBE	8	13	6	27
YE	12	12	3	27
total	20	25	9	54

Table 9. UK Confidence

The Fisher's Exact test confirmed that these differences in distributions of confidence were not statistically significant, at $p = .447$. The demographic variables were tested as an independent variable in addition to video guise, using multinomial logistic regression, and the confidence levels were still statistically insignificant regardless of the specific demographic.

Question 3: Were participants more likely to rate the defendant lower in characteristics?

As can be seen in Figure 17, the mean rating scores for how the defendant and prosecutor were perceived are mostly in the 3-4 range (further details in Appendix V). As the Likert scale was 1-6, these hover in the middle range and show no sign of strong judgements between either the speaker (prosecutor or defendant) or the video guise (SSBE or YE). There is a trend in that the SSBE guise has a slightly higher mean rating than the YE guise, regardless of speaker or type of characteristic.

I used Independent t tests in order to determine if the type of video (SSBE or YE) had a significant effect on the mean ratings of trust, reliability, and believability of the prosecutor and defendant.

<figure>
Perception Ratings of the Prosecutor and Defendant

Legend: ■ U.K. Mean   ■ U.S. Mean

Bar chart with y-axis from 0 to 6. Categories along x-axis:
- prosecutor trust: U.K. ≈ 3.2, U.S. ≈ 4.4
- prosecutor reliable: U.K. ≈ 3.2, U.S. ≈ 4.5
- prosecutor believable: U.K. ≈ 3.4, U.S. ≈ 4.6
- defendant trust: U.K. ≈ 3.6, U.S. ≈ 3.3
- defendant reliable: U.K. ≈ 3.2, U.S. ≈ 3.15
- defendant believable: U.K. ≈ 3.9, U.S. ≈ 3.45
</figure>
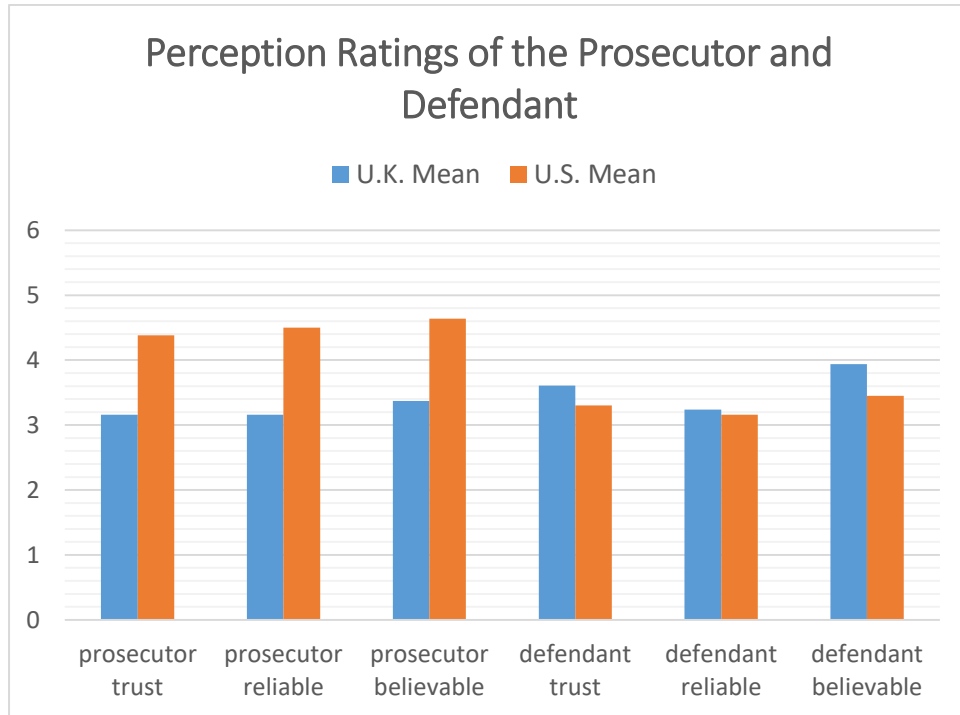
Figure 17. UK Question 3 - Perception Ratings

However, there was nothing statistically significant about that trend. Table 10 shows the results when all the mean ratings for the SSBE guise were compared against all the mean ratings for the YE guise. Overall, there were no statistically significant differences in mean scores between SSBE and YE video types ($p > .05$ in all cases). How reliable the prosecutor was perceived to be neared significant levels ($p = .063$), but there is no evidence in any of the other questions that this near significant result is meaningful.

	t	df	**Sig. (2-tailed)**	Mean Diff.	Std. Error Diff.
prosecutor - trust	.726	47	**.471**	.238	.328
prosecutor - reliable	1.902	47	**.063**	.565	.297
prosecutor - believable	.933	47	**.356**	.312	.334
defendant - trust	.739	47	**.464**	.220	.298
defendant - reliable	1.035	47	**.306**	.317	.306
defendant - believable	.594	47	**.555**	.205	.345

Table 10. UK Perception Ratings – Independent T Test

The participant demographics were added to the analysis to test whether they had any effect on the ratings given. However, after running multinomial logistic regression tests, no statistical significance was found.

Question 4: What level of punishment was suggested?

27 participants did not answer the question on recommended punishment due to their verdict of *not guilty*. Therefore, there is only a total of 27 responses in Table 11. The majority of participants recommended something other than prison, such as community service or a monetary fine, regardless of the video guise. There were five participants who recommended no punishment. This could suggest that these five changed their minds on the verdict, as they did not feel the defendant ought to be punished in any way. However, none of those five reported changing their minds, so it is equally possible that they felt the defendant had already suffered a consequence in some way, perhaps by having to go to court. A Fisher's Exact test was performed in order to determine if recommended punishment was associated with video type.

	punishment				
	None	Other	Prison	Prison+	*total*
SSBE	2	10	3	1	16
YE	3	7	1	0	11
total	5	17	4	1	27

Table 11. UK Recommended Punishment

The results of the test was not significant (p = .836). The demographic independent variables were also tested using multinomial logistic regression, but no statistical significance was found.

Question 5: Did the video guise make a difference in how long participants took to finish the questionnaire?

The mean length it took for participants to finish the questionnaire after watching the SSBE video was 1083.81 seconds (18 minutes) while the mean length after watching the YE video was 1218.56 seconds (about 20 minutes). There is only a difference of 135 seconds in mean length between the two guises, suggesting that on average participants completed the questionnaire slightly faster after hearing the SSBE guise. I used an independent t test to determine if the length of time to complete the survey differed based on video type.

Results of the independent t test were not significant, at p = .561. Thus there is no evidence to suggest that participants were quicker to judge one guise than the other. Despite the apparent trend mentioned above, the difference in length was simply too small. The demographic independent variables were also tested using multinomial logistic regression, but again the tests returned no significant findings.

Question 6: Did the quality rating participants gave the videos differ significantly depending on which video?

Regardless of video guise, majority of participants rated the video quality as being "high quality," although there is an apparent trend in that the YE guise has more "low quality" ratings than the SSBE guise. This can be seen in Table 12. A chi square test was used to determine if there was a relationship between video quality and type of video watched.

	video quality		*total*
	High	Low	
SSBE	22	5	27
YE	18	9	27
total	40	14	54

Table 12. UK Video Quality Ratings

These differences in distributions were not statistically significant, at p = .214. Again the demographic independent variables were tested as before, but there was no significant levels.

Questions 2-6 have sought to answer RQ I. The results have shown no evidence of any statistically significant differences in how participant answered dependant on the accent guise. RQ I will continue to be looked at in the following qualitative analyses of the responses participants gave in regards to analysis questions 7-9.

Question 7: Did participants change their verdict after reading the text summary? If so, what reason did they give?

An overall number of 10 participants changed their minds regarding the initial verdict they gave. Most changed from a *guilty* verdict to a *not guilty* verdict after reading the text summary, with only a single participant switching their verdict to *guilty* after reading the summary.

	Changed to "not guilty"	**Changed to "guilty"**
SSBE	6	1
YE	3	0

Table 13. UK Question 7a

A chi square test was run to analyse this, but unsurprisingly did not return significant (p = .359). Too few participants changed their minds to show any significant effects due to the accent guises.

The three YE all changed due to the speed limit evidence being five miles prior to the accident. They noted that five miles is actually a long period of driving time and the defendant very well may have slowed down at some point and no longer been going over the speed limit. Three of the SSBE participants also said they changed their minds due to the speed limit evidence being misleading as it was five miles away from the scene of the accident. One of the SSBE participants did not say why they changed their mind. The final two SSBE participants said they changed their minds due to realizing the children had more culpability than they had previously noted. They discussed how the children cycled straight into the road and it was already difficult for the defendant to have missed the first children before than hitting the final child. The only participant who switched their verdict to *guilty* also reported that it was due to the speed limit evidence, only they found it more damning of the defendant, rather than recognising how far away five miles was. These are all laid out below in Table 14.

	speed limit	reckless children	no answer	*total*
SSBE	4	2	1	7
YE	3	0	0	3

Table 14. UK Question 7b

However, a difference of seven SSBE participants to three YE participants is also not large enough to denote significance. Additionally there were no discernible patterns in the demographics of the 10 who changed their verdict. This does suggest, however, that participants missed pertinent information in the video and having a text summary provided them with additional relevant, and possibly crucial, information.

Question 8: Were participants more likely to offer negative opinions on the defendant in the Yorkshire guise?

When participants were asked what they found the most convincing about the defendant, the positive comments mainly centred on his emotions (e.g. "he seemed remorseful," "that he was sorry") and his actions (e.g. "it seemed like he did what he

111

could," "he swerved and braked later"). A few participants also mentioned his memory (noting that it is understandable he would not remember the incident clearly) and two participants sympathised with the defendant over how quickly accidents happen. Still other participants found nothing positive to say about the defendant. This can all be viewed in Table 15. One participant refrained from answering this question. Although participants had less negative things to say about the standard guise, focused less on his emotions and more on the facts of the case, and offered more sympathy, it is not by a large enough margin to claim significance. The ratings only differed by a maximum of two participants in any category. The four participants who did not have anything positive to say all had children; however, this is the only visible pattern within the demographics.

	Actions	Emotions	Memory	Quick	All negative	total
SSBE	10	9	5	2	1	27
YE	10	10	3	0	3	26

Table 15. UK Most Convincing

Participants' responses to the question regarding what they found *least* convincing about the defendant differed slightly more. This is shown below in Table 16. Again the defendant's actions, emotions, and memory of the event were all mentioned, though in negative contexts. Some participants did specifically mention the defendant's speech, with the SSBE guise receiving more negative comments on this than the YE guise, though only by a matter of one participant. Six participants refrained from answering this question. Unlike in Table 15, here one category differs by a total of four participants and another category by a total of five. The standard guise had four more participants than the non-standard guise claim they had nothing negative to say. Additionally, when speaking in the standard guise, five fewer participants questioned an aspect of his memory. However, again it is difficult to say that this is a large enough difference to be significant.

	Actions	Emotions	Memory	Speech	All positive	*total*
SSBE	6	4	7	2	6	*25*
YE	6	2	12	1	2	*23*

Table 16. UK Least Convincing

In order to better analyse this data, I combined the two "convincing" questions and collapsed all the categories into either *positive* or *negative* comments. I did this by adding together the numbers in the four SSBE positive categories in Table 15 with the "all positive" category in Table 16 to create the "SSBE Positive" category below in Table 17. Similarly, I added together the numbers in the four SSBE negative categories in Table 16 with the "all negative" category in Table 15 to create the "SSBE Negative" category in Table 17. I then did the same thing with the YE categories in order to create the overall "YE Positive" and "YE Negative" categories below. As can be seen in Table 17, SSBE was more likely to have positive comments, whereas YE had a nearly even 50% split between positive and negative ratings. A Chi-square test was performed and returned no significant difference (p = 0.287).

	Positive	Negative	*total*
SSBE	32	20	*52*
YE	25	24	*49*

Table 17. UK Combined Overall Convincing

There were also no discernible patterns within the participant demographics dependent on category response.

Question 9: How did participants speak about the accent of the defendant?

Overall the comments were fairly similar between the two guises and mainly focused on the emotions of the defendant. For Yorkshire, participants primarily made comments along the lines of: upset, nervous, emotional, and remorseful. For SSBE, the comments were primarily: Apologetic, clear, nervous, and normal. Of note is that SSBE was described as "normal" and "clear" while Yorkshire was not.

One participant went so far as to write: "Not well-formed sentences, giving the prosecutor something to use against the defendant" in regards to the Yorkshire guise.

In order to better understand the language attitudes participants were bringing to this study, particularly question 9, I looked at how they described their own accents. 45 participants in total chose to answer this question. They were given an open textbox in order to respond with as much or little detail as they felt appropriate. After analysing their responses, I organised them into categories, which can be seen in Table 18. Eight participants took on the non-accent myth, saying they themselves either did not have an accent, spoke normally, or were neutral. One of them further described it as an "educated accent" and another said theirs was "easily understood." 18 participants described their speech in terms of being "Southern" or "RP." Most of them did not give a location within the South, and some went so far as to say it was "unplaceable." Of those 18, six said they had a "tinge" or "a bit" of another thing, but were mainly Southern, and one said they "actively avoided" having that trace of a non-standard accent come out. Three participants said they had a type of "posh" northern accent, separating themselves out from those with a general Northern accent. 10 participants claimed a northern accent, with one participant specifying Geordie and another specifying Yorkshire. Only four participants said they had an Irish or Scottish accent, but all four specified the county or city and said their accent was not very strong. Two participants did not give enough information, one simply put "mixed" and the other put multiple question marks.

Non-accent myth	Southern	"Posh" Northern	Northern	Scots or Irish	Other
8	18	3	10	4	2

Table 18. UK Self-reporting Accents

Overall, 29 of the 45 participants sought to present their accent in a positive light. This was done either by claiming not to have an accent, saying it was southern and unplaceable (and therefore could be perceived as more standard), or by saying that although they had a northern accent, they qualified it as "posh." To an extent even

the Scottish and Irish participants could be seen as doing something similar in saying none of their accents were "strong."

Between the participants own self reporting on their accents and the way they spoke about the defendant's accent it is clear they held language attitudes towards various dialects and accents of English. However, due to the lack of statistically significant findings, the presence of these language attitudes does not appear to have affected their judgments throughout the study. Therefore the answer to RQ I in this data is *yes*, the mock jurors did perceive the defendant differently based on his accent. Yet there is no evidence that this caused them to treat the defendant differently in any way.

5.5. Discussion

This study sought to answer RQ I "Will a defendant be perceived differently by a jury based on whether he speaks in a standard or regional accent?" and RQ II "If a defendant is perceived differently, will this go as far as affecting the verdict?" in the context of the United Kingdom. The analysis addressed these questions through the specific questions outlined in section 5.3.6. The answer for RQ I was *yes*, the defendant was perceived differently, as there was clear evidence of language attitudes present in the qualitative analyses. However, there was no evidence in the quantitative results that these language attitudes caused the defendant to be treated differently or in any way affected the verdict, making the answer to RQ II a clear *no*.

I found no evidence to support my initial my hypotheses and predictions when, after analysing all 9 questions, no significant effects dependent on the accent condition were found. There were some apparent trends, but they were not consistent. SSBE had more *guilty* verdicts while YE had more *not guilty* verdicts in question 1, showing a trend that would suggest the non-standard guise was actually favoured. However, in question 3 SSBE received slightly higher ratings than YE across every category. Due to the lack of consistency in the trends, and the general lack of statistical significance, I have no reason to believe these apparent trends were meaningful.

The notable result found in the data was the clear presence of language attitudes and ideologies held by the participants. This was shown in the qualitative

answers to question 9. It was evident that some participants held to the non-accent myth, believing themselves or the SSBE guise to be "normal" and "neutral" in speech. In some cases they went far enough to say outright that they or the SSBE guise did not have an accent. Previous research into perceptual dialectology contexts has found similar results in how the standard is perceived and commented on (Campbell-Kibler, 2009; Preston, 1989). This finding suggests that the lack of significant results in the quantitative data was not due to this participant group being unaware of linguistic stereotypes. Rather the qualitative answers would suggest that the participants were not only aware of these stereotypes, but when directly asked about accent they were able to repeat those stereotypes, revealing their own language attitudes. However, due to the absence of statistically significant results, it is equally clear that the presence of these language attitudes had no overall effect on their responses.

The importance of that finding cannot be overstated. As discussed in chapter 2, language attitudes and ideologies have long been studied. However, it has not been clear whether people regularly (and maybe even purposely) acted upon these attitudes and ideologies when interacting with others, or if these attitudes are held and known, but are not automatically used in interactions. The results of this research would suggest that the fact people have language attitudes may not be relevant in every situation. People are not automatically inclined towards discriminating solely due to someone's speech. While accent discrimination does occur, as outlined in chapters 2 and 3 both through real life cases and in research, it may be that accent discrimination is not prevalent. Previous research has found evidence of language attitudes causing discrimination, but this could be due to many of the experiments' methodologies foregrounding the linguistic variable, and therefore priming the participants. More research on this topic ought to be conducted to discover what must be present in order for an environment to be open to accent discrimination. Recommendations and further discussion on this is discussed in chapter 8.

While the results of this study did not find any significant effects regarding accent, this can still be understood in light of previous research. Many of the studies listed in chapter 3 found either weak results in terms of accent or did not find any significant differences. Seggie (1983) showed people held language attitudes, but not

if those same people would have acted upon them and therefore make any difference in a conviction. Dixon et al. (2002), along with Dixon and Mahoney (2004), did find that the non-standard guise was more likely to be perceived negatively, but again that does not say whether a conviction would have followed. Maerowitz (2014) found similar results and Grey (2010) did not find accent to have any strong effects on verdict. Sobral Fernández and Prieto Ederra (1994) and Frumkin (2007) both looked at witness testimonies rather than suspect or defendant testimonies. They did find accent to have an effect, in that participants were more likely to give higher ratings to the standard or those they were more familiar with. However, once again this does not answer whether this would have had an effect on a conviction or if participants were simply revealing the language attitudes they held. Therefore, the findings in this study align with the previous research into this topic in that language attitudes were found to be present. Furthermore, this study adds more nuanced knowledge in that these same language attitudes do not automatically cause accent discrimination.

This was an experimental design that purposefully used a neutral environment. The crime type was not heavily stereotyped, and could be perceived as either an unfortunate accident or reckless driving. The regional guises used had both negative and positive stereotypes associated with them. Finally, in every way possible my interest in linguistics was hidden from the participants. This was all done in order to give participants options in how they reacted and what responses they gave, rather than pushing them towards a specific response. It is possible that the lack of significant results is due in large part to this neutral design. However, as stated before, this in no way invalidates the results. Rather, it shows that not all non-standard varieties will be automatically discriminated against, as some media reports might lead people to believe (Eligon, 2019; Owens, 2019). This is not to suggest that accent discrimination is not real and does not occur; chapters 2 and 3 offer numerous examples in how accent discrimination does in fact negatively affect people across all aspects of life. However, it is also not as simple as creating a more biased environment. Dixon et al. (2002), Dixon and Mahoney (2004), and Maerowitz (2014) all used more heavily stigmatised non-standard varieties, but found weak results with regards to how they were perceived in comparison to the standard varieties used. Therefore, this study, due to the use of a neutral environment, offers a baseline for future research to work from, gradually adding in more variables or

adapting the current variables, in order to understand what variables need to be present in order to open the possibility for accent discrimination to occur.

I chose to use a defendant cross examination as I wanted to test directly whether accent would make a difference in conviction rate. Although the trials outlined in chapter 3 show that some defendants are convicted or discriminated against due to their language use, the results suggest that this was not the case here. This could be due to participants being highly aware of the negative consequences to another person's life if they made a judgment based on anything but the evidence. It is possible that the same care and consideration is not given to a witness when deciding how believable their testimony is. It does not seem to be an exaggeration to suggest that juries receive much less information about a witness than they do a defendant. When on trial, a defendant's entire life may be uncovered and examined in the process of trying to prove guilt or innocence. A witness may only be heard from once and given the barest introduction. This gives a jury much less to go on when deciding the importance of a witness testimony and it may be in this context where accent plays a larger role. This is something that could be looked into in future studies.

Beyond the use of a defendant rather than a witness, it is possible that the lack of significance arose due to the methodology and its structure. Although I purposely noted that the speed limit evidence occurred five miles before the incident, many participants may not have noticed that detail, as evidenced by the 10 participants who changed their verdict due to that realisation. Others who may have noticed that it was five miles away from the incident may not have considered, or been aware of, just how long a distance that is. Additionally I may have made the speed too high. Going less than 10 miles over the speed limit may have been deemed more acceptable, and therefore forgivable, than going 11 miles over. Although I surveyed drivers and found that going 10 mph over the limit was generally deemed acceptable, I did not specify whether that was just as much true in a 30 limit zone as a 60 limit zone. Instead I asked a question that may have been too general. Overall, it is possible that I made the negative speeding evidence too strong without fully making it clear that the speeding evidence was irrelevant due to the speed being registered five miles before the location of the incident.

Another potential issue in the methodology was the choice of Yorkshire as the non-standard accent. While YE may be viewed as lower educated than the standard, it does not have any associations with being unintelligent or criminal. Rather it is considered friendly and hospitable. It could be that YE is not viewed negatively enough to offer a distinct difference between itself and the standard in a courtroom scenario. This result does suggest, however, that simply being a non-standard accent is not enough for accent discrimination to arise. It may be that it requires a heavily stigmatised non-standard accent, or rather a heavily stigmatised community where one of the ways to discriminate against them is through their speech. Equally, because my stimuli still used Standard English syntax, it could be that the YE phonetic features were not strong enough alone and I should have used dialect guises rather than accent guises.

Overall, many of the results were fairly evenly split. *Guilty* and *not guilty* verdicts were completely 50/50 and the Likert scale ratings mainly hovered in the neutral 3-4 area. The results were not heavily skewed in favour of a single result, suggesting that despite any potential methodological missteps, participants were not primed and did not view only a single answer as viable. In this way, the even split in answers may increase the ecological validity of the results, strengthening the conclusion that accent discrimination is not an automatic occurrence.

5.6. Conclusion

This chapter examined the perceptions of SSBE and YE and then sought to discover whether those perceptions would be acted upon, causing discrimination. After analysing the data, it was clear that accent had no effect, suggesting that the perceptions found in previous research are simply perceptions and may not regularly play an active part in people's decision making and judgments within the United Kingdom. This generally aligns with the findings of previous research, in that language attitudes were found to be present. However, they do not automatically lead to strong reactions and overt accent discrimination.

Chapter 6. US Mock Trial Study

6.1. Introduction

This study investigated whether the low status stereotypes of Southern American English have an active effect on juries and whether that can have an effect on an eventual verdict. This chapter will detail the research I conducted in the United States addressing this topic, some of which is a replication of what was done for the UK version (chapter 5). It seeks to answer RQs I, II, and III using US individual participant responses on Qualtrics and US focus groups set up as mock juries. This chapter describes the accents used and their respective stereotypes, followed by a description of the methodology with justifications and explanations for each of the steps taken. The results have been analysed and laid out before discussing their potential implications.

It was predicted that when testifying in Southern American English, the defendant would be perceived more poorly by the jury. I did not make a prediction regarding whether this would lead them to convict the Southern American English guise significantly more regularly than the GAE guise. While research has found that the perceptions differ between Southern American English and GAE (as will be discussed in the following section), it is unclear whether these perceptions are pervasive enough to affect participants' judgments, particularly in the context of a court with serious consequences. However, I did predict that the suggested punishments would be harsher for the defendant when speaking in Southern American English, as that would only come after a participant had decided on a guilty verdict, at which point they may be more open to being influenced by perceptions. Additionally this study sought to answer RQ III through comparing the

results of the individual Qualtrics responses with those of the mock juries'. I predicted that working as a group would cause clear differences from when participants completed the study individually on Qualtrics, though I did not make any predictions regarding how those differences would appear.

The literature review conducted on Southern American English and GAE will now be outlined, describing where these accents are spoken, their linguistic features, and the linguistic stereotypes associated with them. As in chapter 5, it must be noted that what is reported here are the *generalisations* of what has been found over multiple decades, and therefore may not be true for every person and in every context. Instead the following section offers a broad discussion on the variants and language attitudes of the dialects.

6.2. Southern American English and General American English

Southern American English is spoken within the U.S. south-east, roughly spanning from Texas in the west to Virginia in the northeast, as shown in Figure 18. The dialectal line for the south generally follows the region the U.S. census bureau has designated as "South" (view Figure 24 for a map of the U.S. regions). There are variations of Southern American English across this large region, with specific enclaves of accents marked on the map in Figure 18 in the Appalachian Mountains and most of the state of Louisiana. As with any dialect and the variation inherent in speech, those who live in the south can vary from having all the markers of a Southern American English speaker to only some of the features to speaking GAE, or another dialect if they are migrants to the area (Lippi-Green, 2012).
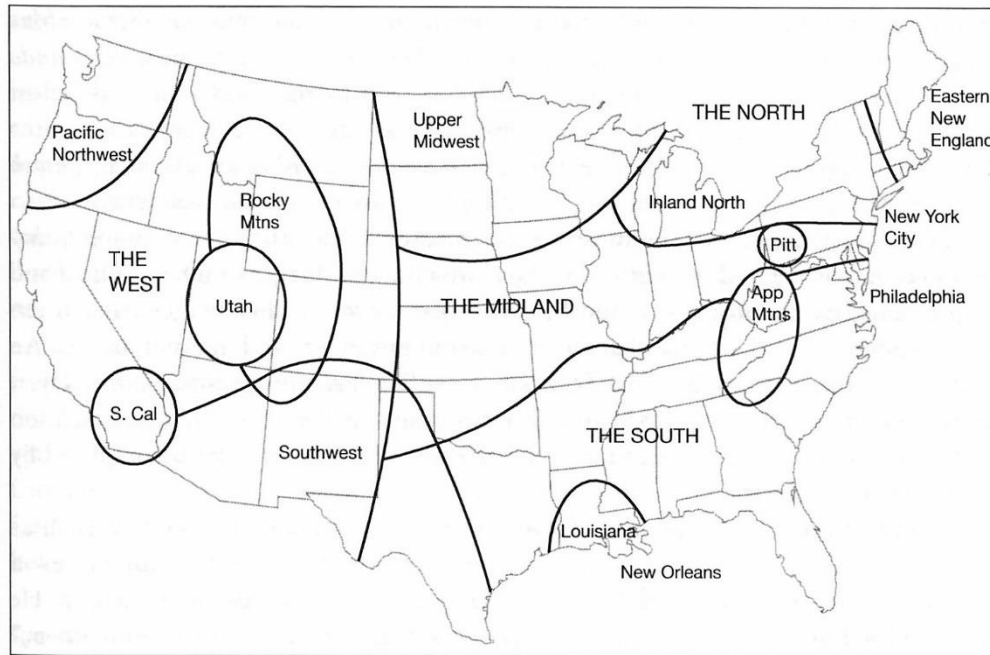
Figure 18. Dialect Atlas of the U.S. (Tamasi & Antieau, 2015, p. 107)

Therefore, for this study I used the variety of Southern American English as described and researched by Allbritten (2011). Her work looks at the most common features found across all or most of the southern varieties, the perceptions of them, and any changes currently occurring. She compiled this information and used the data to describe the overarching dialectal features of the south, which she called Southern American English. I chose to follow suit in order to simplify the methodology by avoiding the numerous variations in accents across the southern region. This also allowed me to discuss the south as a whole, with its general stereotypes and perceptions.

Southern American English speech is often associated with being non-rhotic, although this feature is currently disappearing (Lippi-Green, 2012; Thomas, 2008). It is most famous for its "drawl," produced by monophthongising and lengthening the PRICE vowel /aɪ/ to [a:] in words such as *mine*, *like*, etc. (Allbritten, 2011), as well as the standard /oi/ to [ɔ:] in words such as *oil* (Clopper & Pisoni, 2004). The stereotypical drawl also includes "breaking," where it diphthongizes (or even triphthongizes) long monophthong vowels and adds a glide. This commonly occurs with the TRAP vowel /æ/, where words such as *cat* and *past* are pronounced [kæᶦjət] and [pʰæᶦjəst] (Allbritten, 2011). Older research found a similar process occurring

with the KIT and DRESS vowels, where *hill* became [hɪjəl] and *bet* was pronounced [bɛjət] (Habick, 1980), though this appears to be less prevalent in current speakers (Allbritten, 2011; Thomas, 2008).

A vowel shift is currently underway in the south. It appears to have begun towards the end of the 19th century and is still ongoing (Thomas, 2008). Studies have found that speakers of Southern American English tend to front /o/ and /u/ and have /e/ centralization (Clopper, Pisoni, & de Jong, 2005). Additionally, some speakers may also have /i/ centralisation (Clopper & Pisoni, 2004) and /ɪ/ and /ɛ/ have been found to be raised in some cases. Due to this, in some Southern speakers the words *him* and *hem* are indistinguishable (Feagin, 1996). This is known as the PIN/PEN merger and generally occurs in these vowels when they precede a nasal (Thomas, 2008). This vowel shift can be seen in Figure 19 below.
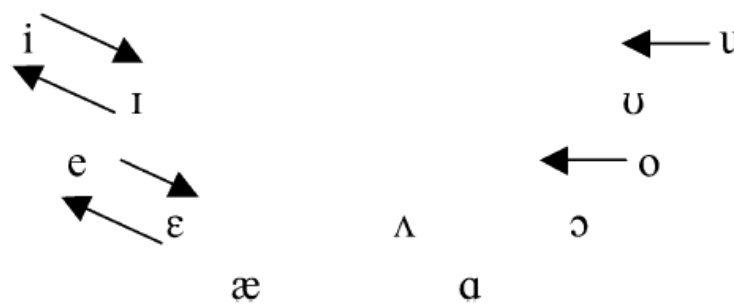
Figure 19. Southern Vowel Shift (Clopper & Pisoni, 2006)

In addition to the vowel variables listed above, there are also a few consonant variables that are associated with a Southern American English dialect. When the unvoiced alveolar fricative [s] occurs between vowels, it becomes voiced [z] in Southern American English, in words such as *greasy* (Clopper & Pisoni, 2004). When the consonant cluster "shr" begins a word, many Southerns will pronounce [sɹ] rather than the standard [ʃɹ]. This can be observed in words such as *shrimp* and *shrub* (Thomas, 2008).

This dialect also has a distinct lexicon, with such terms as *y'all* for a plural second person pronoun; *lightening bug* rather than the standard term *firefly* and *snap bean* for *green bean*. These terms are heavily associated with the south and are not commonly used elsewhere. In other cases, some words used elsewhere hold a different meaning in the South, such as *mess*, meaning "a large amount of

123

something," and *carry,* as in "to drive someone somewhere" (Bailey, 1997). Southern American English is also characterized by differences in its syntactic structure. It utilizes double modals, such as *might could*: "I might could go there." It has perfective *done*, as in "you done went there already." Southern American English has developed *liketa* (nearly) and *fixin' to* (preparing to), and is known to use personal dative constructions such as "you got you a new car" (Tamasi & Antieau, 2015). In many cases Southern American English speakers will also use double negatives, as in "I never did nothing" (Wolfram & Schilling-Estes, 2016).

In comparison, GAE, while not clearly defined itself, is often viewed as being the "correct" variety to speak within the United States (Lippi-Green, 1994). GAE is not associated with any one social group or one specific area, nor is there an agreed upon definition, making it difficult to accurately describe its phonetics (Clopper, Pisoni, & de Jong, 2005). In fact, some experts argue that there is no evidence of a consistent standard for the U.S. and have concluded that GAE is a "mysterious, non-existent variety" (Coupland, 2000, p. 10). This is partly due to that fact that, unlike many European countries but similar to the U.K., no official standard was ever created and decided upon in the U.S. While dictionaries and mandatory education standardised spellings, syntax, and to an extent semantics, areas such as pronunciation were never well demarcated (Tamasi & Antieau, 2015).

Americans cannot consistently point to the same area where the standard is spoken and derived from because GAE is arguably more a language ideology than an accent (Lippi-Green, 2012; Tamasi & Antieau, 2015). When American non-linguists are asked who speaks the standard accent or where that variety comes from, some may say that the standard originates somewhere in the Midwest. However, many on both the East Coast and the West Coast will argue their area also speaks GAE. As not everyone will index the same features as ones that are "standard," this leads to a great amount of interspeaker variation within those who claim to speak "the standard" (Kretzschmar, 2008). Only New York City and the southern states are regularly pointed out as not speaking the standard (Preston, 2002b). GAE is viewed as "mainstream" and (erroneously) what the majority speak. In fact, another name for it is *Mainstream* American English (MAE) (Wolfram & Schilling-Estes, 2016). It is described by some Americans as "colourless" and "characterless," a description that fits with the non-accent myth, as previously discussed in chapter 1 (Milroy,

124

2000, p. 58). This is seen in how many Americans, particularly those who speak a variety of GAE, believe that they themselves do not have an accent, and instead it is everyone else who speaks in a distinctive manner (Lippi-Green, 2012).

Kretzschmar (2008) argues that any pronunciations that are not marked with a specific regional or social meaning can be considered standard by speakers. For instance, across the west, midwest, and parts of the northeast of the US [ɔ] and [ɑ] have merged, creating the low-back merger where *caught* and *cot* are pronounced the same (Wolfram & Schilling-Estes, 2016). This merger is still ongoing, with some speakers adopting it, and others still differentiating between *caught* and *cot* in their phonology. Yet either system can be indexed as "standard," because they are not directly associated with any specific region or community of speakers (Kretzschmar, 2008). This would never be the case with the Southern monophthising of /ai/ because it does carry meaning with regards to both a specific region and type of community. Due to this, some linguists have suggested that it is more beneficial (and more accurate) to describe the dialect of the region of the speaker, rather than try to describe the idealized, and highly variable, form of GAE itself (Clopper, Pisoni, & de Jong, 2005). Although my prosecutor actor was originally Canadian (detailed in section 6.3.1) but has previously lived in Montana and Colorado, I will describe the dialectal features of the western region.

In contrast to Southern American English, within the vowel system, the west uses the standard diphthongs [ai] for the PRICE vowel and [oi]. The KIT and DRESS vowels are also the standard monophthongs. Studies have noted that the western pronunciation of /u/ is fronted, but unlike Southern American English, this is constrained only to /u/ and /o/is not also fronted (Clopper & Pisoni, 2006). With regards to this, however, Clopper and Pisoni (2004) note that the [u] fronting does not appear to be noticed by listeners and is therefore not salient enough to allow listeners to accurately place the origin location of the speaker. /æ/ is commonly raised when preceding nasal consonants. There is variation in the pronunciation of the –ING variable, though it appears that this is stylistic intraspeaker variation, with the alveolar nasal variant used in informal settings while the velar nasal variant will be used in formal contexts. Lastly, there is some distinction between /w/ (witch) and /ʍ/ (which) though this is reportedly dying out as younger speakers retain only the /w/ for both (Gordon, 2008). Except for the –ING variable, none of these carry any

type of indexicality. Even in the case of –ING, the standard velar nasal variant is still used in formal contexts. Additionally, I have not found any evidence of syntactic variation associated with the west. Due to this, it is likely that all of these variants generally go unnoticed by listeners and a general western accent could be considered a form of GAE according to Kretzschmar (2008).

In studies that directly compared the language attitudes held towards dialects, once again the two major categories that emerge when defining how these dialects are perceived are *solidarity* (comprised of characteristics relating to community identity and unity) and *status* (containing characteristics pertaining to levels of prestige) (Preston, 2002b). Participants have rated the South high in *solidarity*, with regards to being friendly, approachable, and welcoming (Bayard et. al., 2001; Heaton & Nygaard, 2011). Southerners are generally believed to be supportive of one another and there is a feeling that they are more willing to help others than Northerners are. However, Southern American English is regularly viewed negatively in terms of *status* characteristics, particularly in professional environments (Lippi-Green, 2012). Southern American English speakers are stereotyped as having a low level of education, being lower class, and as speaking much less "correctly" than others. A common indexical stereotype is that southerners are easy going and lazy, and this is reflected in "their lazy, drawled vowels" (Preston, 2002a, p. 40). These speakers have often been portrayed in movies, shows, comics, etc. according to such stereotypes, which has the potential to reinforce them (Lippi-Green, 2012).

Conversely, the same studies suggest that GAE is rated highly when it comes to *status* (Bayard et. al., 2001; Heaton & Nygaard, 2011). It has often been reported as the dialect employers prefer when hiring or promoting employees (Cargile, 2000). The news media have historically required their reporters to speak GAE (Lippi-Green, 1994). People who speak GAE are more likely to be perceived as educated and of high social standing (Preston, 2002b). However, GAE is regularly ranked lower in characteristics linked with *solidarity* than any non-standard American dialect, including Southern American English (Tamasi & Antieau, 2015). Perhaps because GAE cannot be linked to a specific region, it is not viewed as having a cohesive or collaborative community who speak it, unlike non-standard dialects. Therefore, GAE speakers are not usually identified as offering the same level of

support and assistance other dialect speakers do and so are consistently perceived as having low solidarity but high status.

Similar to my hypotheses in chapter 5, I predicted that U.S. participants would also consider the court a formal setting and thus would prefer hearing testimonies in the standard accent, here GAE. Just as Rachel Jeantel was reprimanded for using non-standard slang and had her own non-standard sentences repeated back to her with standard syntax, I hypothesized that participants would equally take the use of a non-standard accent in a courtroom less seriously (Carodine, 2015; Sullivan, 2016). The following sections investigate this hypothesis and offer discussions on RQs I, II, and III.

6.3. Methodology

The methodology for the Qualtrics portion of this study is nearly identical to the U.K. mock juror methodology outlined in chapter 5. Therefore, some of this section will be truncated and will contain references to the explanations and procedures in chapter 5. However, this methodology will detail in full the study design for the group jury deliberations used in multiple locations around the United States and how the results were analysed.

The nature and goals of the U.S. study were in general terms identical to those of the U.K. study. The addition of the group jury deliberations was done for two reasons: First, real jury deliberations always occur as a group and thus the verdict does not solely depend on one individual. While I did not expect the group deliberations to perfectly represent what occurs at trial, it is still "likely that the same values … and sensibilities that engage the research group will resonate in the jury room" (Keene, 2013, p. 1). Secondly, this allowed me to directly compare the group deliberations and verdicts to the individual Qualtrics verdicts and responses. If the two were generally consistent with one another, this would lend credibility to the studies mentioned chapter 3 that only used individual responses when researching juror decision making. However, if the responses between the groups and the individuals were substantially different, this would call into question the trial research results that used the individual method.

As I did not have any formal background or practical experience regarding jury decision making research and conducting mock trials, I sought to work with experts in that field who could advise me through the process as well as provide the necessary resources. Dr. Tarika Daftary-Kapur at Fairleigh Dickinson University (FDU) in New Jersey and Dr. Edie Greene at University of Colorado at Colorado Springs (UCCS) both agreed to work with me and assist in this research. Through them I was also given permission to recruit students from their respective universities for this study.

6.3.1. Accent Guises

I chose to compare GAE with Southern American English as they have both been widely commented on and the stereotypes associated with them have been well established in previous research, as mentioned in section 6.2. Additionally, both varieties have a large number of speakers. It can therefore be safely assumed that American participants would be familiar with both. They both have distinct phonetic/phonological differences and the stereotypes associated with them are generally contrasting (Tamasi & Antieau, 2015), making any potentially differing effects more straightforward to decipher. Similarly to the U.K. accents, neither GAE or Southern American English are associated with a specific ethnic minority group which could cause confounding variables, and thus worked well for the purposes of this study (Hurwitz & Peffley, 1997). Due to the stereotypes associated with them, these accents are also as comparable to the accents used in the U.K. version as is feasible.

This study, following the successful application of the UK version, employed the matched-guise method. Once again I sought two actors, one to play the prosecutor who spoke GAE and one to play the defendant who could speak both GAE and Southern American English. Both actors were white men and, vitally, were currently living in the York area. This was a necessary factor as the recording had to be conducted in York, as that was where I was based; and making the recording there gave me time to fully edit and create the stimuli before going to the US to conduct the study. Neither man was a professional actor. The defendant actor was in his 30s, originally from Texas, and spoke both GAE and Southern American English

natively. The prosecutor actor was in his 40s and Canadian, but had spent years living in the western region of the US and as a result spoke GAE natively. His GAE guise was piloted to check for authenticity, and none of the listeners identified him as Canadian.

The study had two conditions: in the first condition the defendant testified in GAE and in the second condition the defendant testified in Southern American English. Apart from the change in accent, the testimony otherwise remained the same between the two conditions, with the caveat that the defendant's speech was delivered as spontaneously and naturally as possible, and is thus not identical in the two conditions. As in chapter 5, the defendant actor used Standard English in both guises, with a single exception. In the Southern American English guise, instead of saying "the kids" he said "them kids." However, because this is the only syntactic change, and there were no lexical changes, I considered the stimuli accent guises and not dialect guises.

The Southern American English phonetic features present in the Southern American English accent guise are listed in Table 19. Of the features noted in section 6.2, this accent guise had the main indexically marked features associated with the "southern drawl." The long monophthong PRICE vowel [a:] was produced, there was consistent *breaking* of the TRAP vowel /æ/, and inconsistent breaking of the KIT vowel /ɪ/. The accent guise was rhotic, though this is unsurprising as research has noted that non-rhoticity is disappearing from the south (Lippi-Green, 2012). Additionally, the diphthong [oi] was produced rather than [ɔ:] and there was no evidence of a PIN/PEN merger in the accent guise. I was unable to determine whether the actor would produce a /z/ rather than /s/ when found between vowels as this environment did not occur in the stimulus. Although it is clear that not all of the indexically marked southern features were produced, the features that are mainly associated with the "drawl" were present. Furthermore, participants consistently identified this accent as being "southern." Therefore, this can be reliably considered an authentic Southern American English guise. These same marked southern features were absent from the GAE accent guise.

Southern A.E.	Southern A.E. Accent Guise	
	Production	*Example*
Rhoticity	rhotic	"afternoon" [æftɹnun] "remember" [ɹəmɛmbəɹ]
Long monophthong PRICE vowel [aː]	[aː]	"bikes" [baːks] "riding" [ɹaːdɪn] "bike" [baːk] "tried" [tɹaːd] "trying" [tɹaːɪn] "biked" [baːkt] "my" [maː]
Long monophthong [oi] → [ɔː]	[oi]	"boy" [boi]
"Breaking" – involving TRAP vowel [æ]	Both with and without a glide	"can't" [kæən] "fast" [fæjəst] "that" [ðæəʔ]
"Breaking" – involving KIT vowel [ɪ] and DRESS vowel [ɛ]	Inconsistent breaking involving KIT vowel	"kids" [kɪdz] or [kɪədz]
PIN/PEN merger	No merger	"kids" [kɪdz] "remember" [ɹəmɛmbəɹ]
[s] → [z] / v__v	*Environment not present*	*n/a*

Table 19. Southern A.E. Accent Guise Phonetics Features

To ensure comparability, in both conditions the prosecutor cross-examined the defendant using GAE. The prosecutor's GAE accent was from the western states and is as described phonetically in section 6.2. The actors were given a script, included in the following section, that was used to control content while they spoke in spontaneous style. The process of filming, the materials used, and the coaching given to the actors are all described in the following section.

6.3.2. Mock Trial Stimulus

The concept for the trial in this study and the adaptation of it is identical to the one in the U.K. trial, which is fully outlined in chapter 5. The only difference is in the script, which was written using American English spelling, syntax (UK: "…went to hospital" vs. US: "went to *the* hospital") lexis (UK "lorry" vs. US "truck"), and the American convention of the day following the month. The script can be viewed below. Because the interaction was meant to be extempore, and thus more natural, the actors made small changes to the script as they spoke, just as the U.K. actors did. However, the changes were insignificant and therefore are not marked in the script.

Attorney stands up to begin questioning defendant.

Attorney: Could you please describe to the court, in your own words, the events that occurred on April 23rd?

Defendant: I was driving home that afternoon in my truck. As I came down Spencer Street two kids on bikes suddenly rode onto the street. I swerved to miss them. But another kid riding a bike came out of nowhere. I was already swerving to avoid the other kids, there was nowhere for me to go… I braked but…that was it.

Attorney: So you hit 8 year old Peter?

Defendant: I did everything I could to avoid hitting the kids!

Attorney: Because of your actions he went to the hospital with severe injuries, didn't he?

Defendant: … I … I don't know what to say … I'm so sorry that it happened…

Attorney: Tell me, there was a sign at the top of the hill with a 35 miles per hour speed limit, wasn't there?

Defendant: Uh, I can't remember?

Attorney: A speed camera – only 5 miles away from the scene of the incident – logged your vehicle as travelling at 46 miles per hour. That is 11 miles over the speed limit. Do you recall travelling so far above the legal speed limit?

Defendant: No sir.

Attorney: Now by your own admission, you didn't brake when the first children came out in front of you, did you?

Defendant: No… it all happened so fast…

Attorney: You only began to brake when Peter came in front of you, which was too late, wasn't it?

Defendant: … I was just trying not to hit the other kids – I tried, I swerved as hard as I could. And then when that boy biked out in front of me, I just remember slamming on the brakes with both of my feet, but it wasn't enough.

Attorney: Sadly, no it wasn't. No further questions.

Attorney sits down.

Filming of the mock trial was again conducted in the University of York's moot courtroom. The same two Panasonic V160 video cameras that were used to record the U.K. version were used here to record both actors simultaneously. I did not use Zoom recorders this time because they had not been necessary when editing the U.K. version.

Throughout filming, both I and my supervisor, Paul Foulkes, provided instructions and guidance to the actors in order to achieve the results we desired. From the start, the actors were directed to use the script as a guide, but to speak naturally and adapt the script in minor ways as needed. I told them to speak slowly, to take time to think through their questions and answers, to give time between each other's turns. I suggested they pause, hesitate, or even stumble over their words in order to make their speech appear unrehearsed. Due to the prosecutor actor having lived in the U.K., he had picked up some British English syntax, such as saying

"…went to hospital," leaving out the definite article that would immediately precede the noun. As this phrasing is not generally present in American English, I reminded him to use the article in that sentence. However, this was not a problem in the final product as his guise was confirmed as authentic, as previously mentioned. I continued to film takes of the mock trial until everyone involved was satisfied with the results. In total I had five takes, three of the GAE cross examination and two of the Southern American English examination.

Unlike the U.K. version, I did not ask friends to attend the video shoot, as I knew from the U.K. videos that only one person would be seen in the background, and then only occasionally. Therefore, simply to match the look, I asked Paul Foulkes to sit behind and to the left of the prosecutor. In both versions, the actors were dressed in suits as appropriate to what would be expected in a real courtroom.

I edited the audio and footage using exactly the same methods as described in chapter 5. Once again, the lines in the script above note where there are purposeful breaks in the video footage. This also corresponds to when the timestamp in the bottom right corner jumps forward in order to give the impression the cross-examination lasted much longer. The videos ran for a total of 1:59 and the three segments of the video can be viewed in Figures 20-22. Furthermore, these videos can be watched by following this link:

https://drive.google.com/drive/folders/1Fo_2pSmhQ2no0F2U6J6X-y12htAqD5qU?usp=sharing

Figure 20. US Mock Trial First Segment

Figure 21. US Mock Trial Second Segment

Figure 22. US Mock Trial Third Segment

6.3.3. Ethics

To conduct this study, I was required to complete an ethics form for the department of Language and Linguistic Science. I included both the Qualtrics and group deliberations in the same ethics form I filled in for the U.K. version, the details of which can be found in chapter 5. In addition to this, while UCCS did not require it, I did have to complete IRB ethics paperwork and CITI certification for FDU in order to run studies there and use their participant pool. The same justifications and explanations were made for FDU's ethics as for the University of York. FDU also approved my research.

6.3.4. Qualtrics Questionnaire

Individual responses were collected online, using Qualtrics Survey Software. The questionnaire was formatted and run identically to the final U.K. version of the study, as described in chapter 5, with three minor changes. First any British English wording was changed to reflect American English culture (e.g. which state the participant lived in rather than which U.K. country, the specific political parties, and the types of alleged crime), and syntax (e.g. the day following the month when writing dates). Secondly, the videos used were, of course, the U.S. videos where the

defendant testified in either GAE or Southern American English. Thirdly, the question regarding how confident participants were in the verdict they had given was changed from three multiple choice answers to a six point Likert scale. This was changed due to a recommendation made by Tarika Daftary-Kapur as it would provide more data points and was more in line with the style of questions usually given in jury research. The updated question can be seen below in Figure 23.

Figure 23. New Confidence Question

6.3.5. Qualtrics Participants

Similarly to the UK version of this study, participants were recruited for the online Qualtrics portion of this from across the United States, using social media and word of mouth. This improved the ecological validity as I was able to get multiple demographics from every region of the U.S. However, using this method only gave me 16 participants after four months of running the study. I wanted a minimum of 55 in order to be comparable with the U.K. version. Therefore, I used Amazon Mechanical Turk "MTurk" to recruit another 40 participants, for a total of 56 overall. MTurk allows researchers to pay for participants, making it easier to gain non-university student participants and to do so at an affordable cost. MTurk has been found to be a reliable data generator by numerous studies as long as restrictions are

136

put in place, such as length of time spent on the survey or the geographical location of the participants (Buhrmester, Kwang & Gosling, 2011; Buhrmester, Talaifar & Gosling, 2018; Goodman, Cryder & Cheema, 2012; Paolacci & Chandler, 2014). With that in mind, I was able to set specific parameters that limited my MTurk participants to those who lived in the United States, completed the entire survey, gave "real" answers to the questions (not just a string of letters in order to progress to the next question), and took at least 15 minutes to complete the survey (which would suggest they took the necessary time to think about their answers). Anyone from MTurk who did not fulfil one or more of these requirements had their responses deleted and was not paid. The 40 MTurk participants I accepted were each paid $0.50 for their time and were gathered within a two day period.

In total, the study ran for four months, reaching the goal for 56 total participants. Unfortunately, due to having to delete multiple MTurk participants responses who did not fulfil the location and/or time spent on the survey requirements, there was a difference of ten participants between the two accent guises. GAE had 23 participants while Southern American English had 33 participants. This difference in numbers was accounted for in the analysis. I had a fairly even split between participant gender, with there being slightly more females (32) than males (28). Further breakdown of the participant demographic information can be seen in Tables 20-24.

As before, participants were asked about their ethnicity, with the option given to not specify. The majority were white, but as seen in the table below, there were more ethnicities represented in the US Qualtrics dataset than the UK dataset.

Ethnicity of Participant	No. of Participants
Asian	7
Black	8
Hispanic	2
White	34
Prefer not to specify	5

Table 20. U.S. Qualtrics Participant Ethnicity

In order to study perceptions based on regions as well as from the whole of the United States, based on where participants reported living, I organised the participants into one of the four sections recognized and defined by the U.S. Census Bureau: Northeast, Midwest, South, and West. This regional divide can be seen in Figure 24.
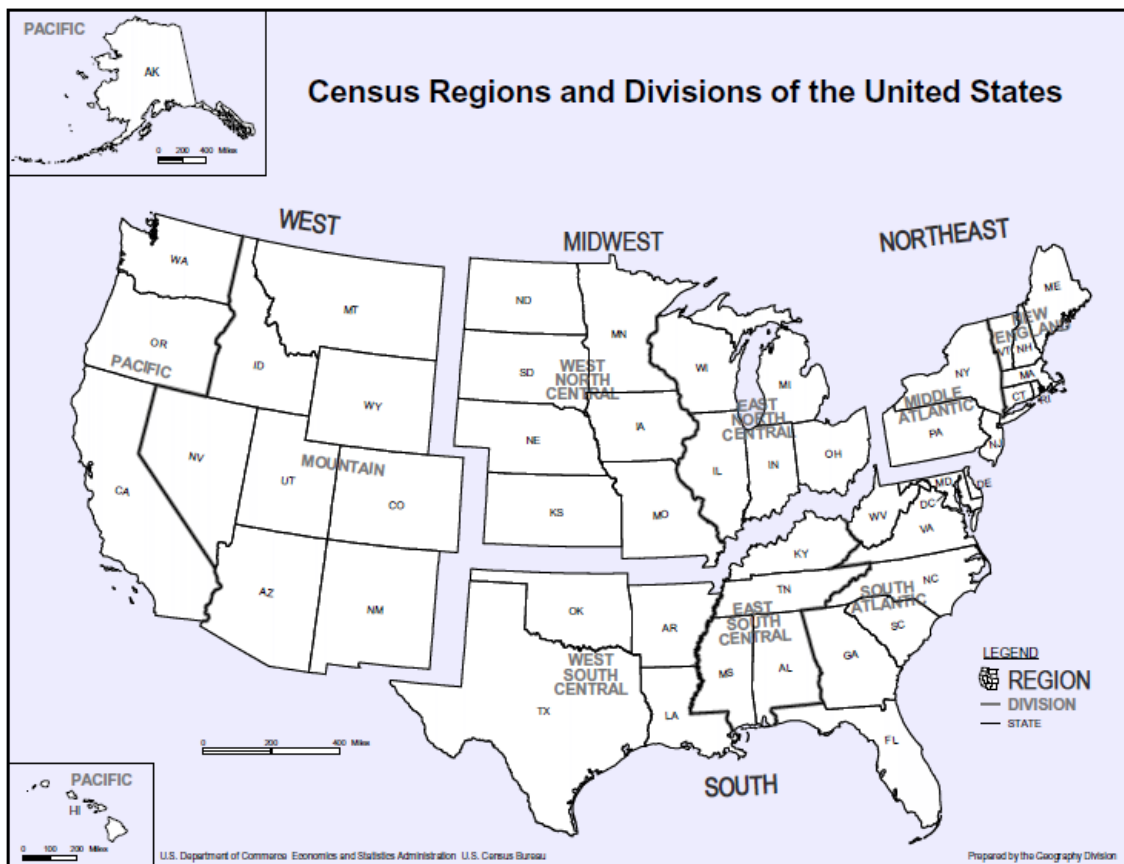
Figure 24. U.S. Census Regions and Divisions (U.S. Census Bureau, 2016)

There was roughly an even number of participants from each region, with the largest number coming from the south. It will be worth noting whether southern participants respond to the accent guises, in particular Southern American English, differently from the other participants. The remaining six participants were Americans who currently lived in a foreign nation. Although I did also collect information regarding the participants' regions of origin in the US before moving abroad, I felt it was more relevant to mention where people currently lived, as that multi-cultural life experience could have affected their language attitudes in unpredictable ways. In this way I was accounting for potential outliers.

Location of Participant	No. of Participants
Northeast	13
Midwest	11
South	17
West	9
Foreign	6

Table 21. U.S. Qualtrics Participant Location

The majority of participants were between the ages of 25-34, followed by participants between 35-64 years old. I did not have any participants over the age of 65.

Age of Participant	No. of Participants
18-24	10
25-34	27
35-64	19
65+	0

Table 22. U.S. Qualtrics Participant Age

Although I had participants in every driving category, from those who never drove to those who drove daily, the largest group by far was daily drivers, followed by those who drove multiple times a week.

Average Driving Regularity	No. of Participants
Never	4
Rarely	3
A couple times a month	2
A couple times a week	12
Daily	35

Table 23. U.S. Qualtrics Participant Average Driving

The majority labelled themselves as "liberal," but the "conservative" group had nearly as many participants. No participant chose not to specify their political affiliation.

Participant Political Affiliation	No. of Participants
Conservative	19
Centrist	8
Liberal	25
Unsure	4
Prefer not to specify	0

Table 24. U.S. Qualtrics Participant Political Affiliation

Lastly, 25 participants out of 56 reported having children and nine participants said they had previously served on a jury.

6.3.6. Mock Jury Design

Within the US, trial juries (also known as petite juries) can be made up of anywhere between 6-12 jurors (United States Courts, 2018). While 12 has long been the traditional size of a jury, courts have gradually been moving to using smaller sizes in order to save time, money, resources, or other issues such as too many jurors being rejected for reasons such as perceived biases (The National Center for State

Courts, 2004). This includes highly publicized court cases such as *Zimmerman v. State* (2013), where only six jurors were used. Due to this, when running the focus group mock trials, I sought to recruit six participants per mock jury. I chose the smallest size as it maintained ecological validity while also making the group dynamics more manageable, requiring fewer participants to recruit, and using less resources (e.g. amount of paper questionnaires, size of room needed, etc.). However, in some cases when I was not able to reach that number or participants did not show up to their assigned time, I did run mock trials with only four or five participants in the mock jury. In this way I was still able to get the desired group deliberation despite the lower level of participant involvement. Overall I had a total of 82 participants across 18 mock jury groups.

An hour was scheduled for each mock jury, even though on average they took 35 minutes. However, by scheduling an hour, it gave time between mock juries to prepare for the next group by resetting the room (getting all the participants forms ready and the video set to play) and tidying from the previous group when necessary. It also meant there was no need to rush participants as they deliberated or filled in each questionnaire. When participants arrived for their mock trial they were informed it was unlikely to take the full hour.

At FDU I was given permission to use a classroom within the Criminal Justice, Political Science & International Studies department. I arranged the participants to be sitting in a semi-circle, making it possible for them all to watch the video at the front of the room but also be facing one another when they began to deliberate. This setup can be viewed below in Figure 25 when it was arranged for five mock jurors. Everywhere else I had a long table that I had students sit at, which still allowed them to face one another while deliberating and to watch the video.
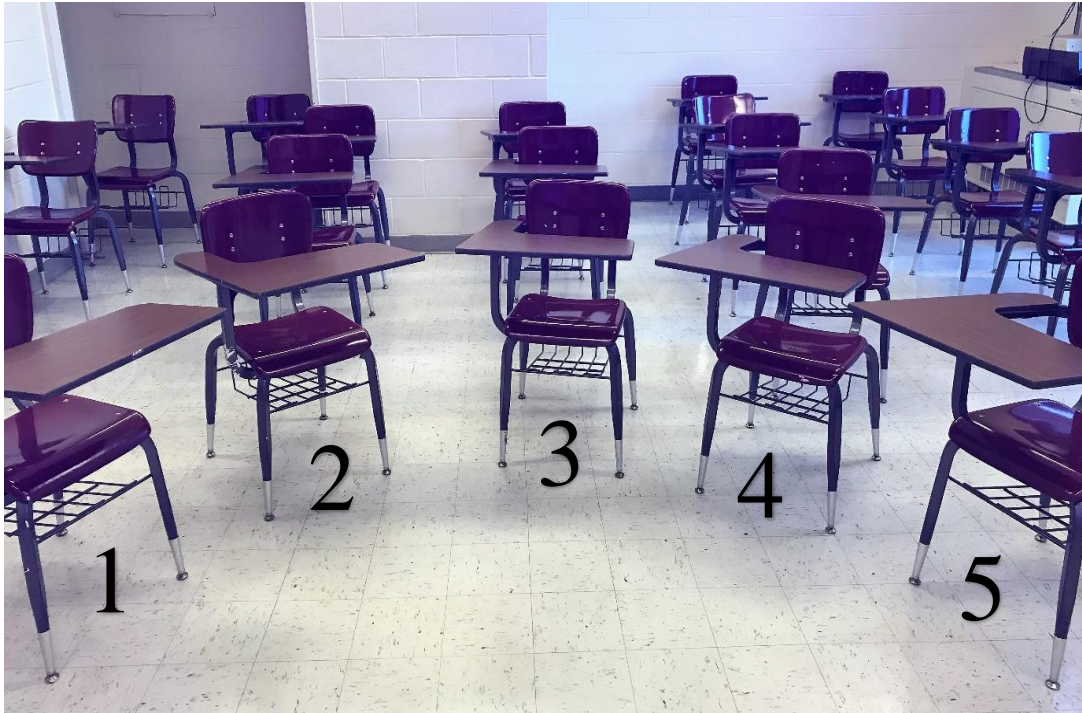
Figure 25. FDU Juror Seat Arrangement

I gave each participant a name badge with a juror number on it and a jury letter. This allowed me to keep track of each participant response while still maintaining anonymity. Additionally, assigning each jury a letter made it easier to keep track of which participants had been a part of which jury, as well as which juries had rendered which verdicts. The numbers underneath the chairs as seen in Figure 25 shows how each participant would be assigned a juror number based on where they sat.

Once participants were seated and given their juror badges, I handed out the information and consent sheets (again found in Appendix II). I asked them to read the texts carefully, and if they were still willing, to then sign and date the paper before giving it back. I also reminded them verbally that all the information they gave was completely confidential and everything would be kept anonymous. I told them they were free to leave at any point, in which case any information they had provided so far would be destroyed and not used in any way. I warned them that the trial clip they would watch dealt with the topic of an injured child and that if anyone thought they would find that difficult to deal with, they were free to leave. I informed them that during their deliberation they would be filmed, but their deliberation would be transcribed and their names would not be used. No participant

chose to withdraw at any point. When participants signed the information and consent sheets, I collected them back and put them away in a folder separate from the rest of the responses.

At this point I handed out the demographic questionnaires (included in Appendix III). I used the same questions in the same order as the online Qualtrics version, with the only difference coming in that participants were reading them on a printed out page and therefore could see more than one question at a time. I reminded participants to write their juror number at the top of the first page of the demographic questionnaire. Once they were finished, I collected them, making it clear I was putting them in a folder different from the one that had their signed information and consent sheets.

I then brought their attention to the screen where I had one of the two video stimuli ready to go. I told them what they were about to watch came from a real trial and I had been given special permission to use the video for the purposes of my doctoral research. I said I had chosen the cross-examination for them to watch but that I had edited it down so that it would only contain the bare minimum information necessary to understand the key elements of the case. Additionally, I told participants that I had blurred out the defendant's face for reasons of anonymity. I asked them to pay close attention and be ready to discuss the particulars of the case once it was finished. I alternated which video stimulus I played with each mock jury.

After the video was done playing, I gave each participant a short pre-deliberation verdict questionnaire. I did this in order to measure whether participants' opinions would change or be affected in any way by the deliberation process. I asked for their initial verdict of the defendant based on what they watched. I then asked, on a Likert scale of 1-6, how confident they were in this verdict. As I ask this confidence question again in the individual questionnaire after the deliberation process, it gives a way to test the effects of the deliberation. Once again I reminded participants to write down their juror number on each of their sheets. This pre-deliberation verdict questionnaire can be seen on the following page:

143

After some consideration, how would you as a juror find the defendant based on the video clip?

> Not guilty

> Guilty

How confident are you in that verdict?

	Not at all					**Very**
Confident						

From this point, I collected their pre-deliberation verdicts and began the video recording in order to capture their deliberation. In the first mock jury deliberation I ran I recorded using both my iPhone 7 IOS 11.03 and a Canon Elura 65 camcorder. However, after reviewing the footage and audio, my iPhone 7 had better quality for both. In light of this, in all subsequent mock jury deliberations I recorded only using the iPhone 7.

After starting the recording on my phone, I once again reminded participants that their deliberation would be recorded, but anonymity would be maintained. I requested that the first time they spoke that they would say their juror number. In this way, during the transcription, I could mark down the speaker according to their juror number and compare their verbal comments to their written responses. I then told them they would have 10 minutes to deliberate, with a maximum of 15 if they found it difficult to come to a consensus. Each time I assigned the participant who was Juror #1 as the "Foreman," giving him or her the jury verdict form, seen directly following:

Verdict   **Jury #**

After deliberation, how did you as a jury find the defendant?

> Not guilty

> Guilty

I included a space on the form for me to write their jury letter so I could keep track of each jury's verdicts. I then left, telling them I would be just outside the closed door and to come get me when they were finished deliberating. I chose to leave them alone to deliberate as, even though they knew they were being recorded, I believed they were likely to have a more honest debate, and to be more willing to get into the debate, if they did not feel like I was directly watching and analysing them. All of the deliberations were transcribed for analysis and the specific guidelines followed when transcribing can be found in Appendix VI.

Once the mock juries finished deliberating and gave me their verdict sheet, I handed out the individual questionnaire. Participants were informed that this was the final piece of the study and asked to not read ahead but instead do each question as they came to it before moving on. In this way I hoped to mitigate the potential for priming if they read the questions on accent before completing the rest of the questionnaire. All the questions in this were formatted and organised the same as in the online Qualtrics study except for the addition of two questions at the start. These questions dealt with the deliberation process and working as a group to come to a verdict. They are included directly below.

1. If your verdict changed after deliberations, can you explain why?

2. Was it difficult to work as a group to come to a verdict?
   o Yes
   o No

Occasionally participants asked me for clarification on questions, such as what I meant by "accent" when they reached the questions asking them to describe the prosecutor's, defendant's, and their own accent. Each time this happened I made a note of what they asked and regarding which question(s), and then included it in my data sheet to be a part of the eventual analysis.

The final page of the questionnaire gave participants the opportunity to leave their email address if they wanted to learn more about this research. It gave them a second chance to leave their email address if they wanted to be signed up for the

$200 raffle, which had been offered as an incentive to recruit participants. If participants left any personal information on this page, they were asked to tear this page off from the rest of the questionnaire and hand it in separately. In this way anonymity was maintained and I stored these pages with the signed information and consent sheets.

6.3.7. Mock Jury Participants

The mock juries were recruited through Fairleigh Dickinson University and University of Colorado at Colorado Springs. This was done by posting flyers around both campuses; the flyer can be viewed directly following this paragraph. Additionally, both universities permitted me to incentivise students through extra course credit. At FDU a number of Dr. Daftary-Kapur's colleagues were willing to offer extra credit to their students for participating in my study. At UCCS I was able to be added to a university wide extra credit system known as SONA, which allowed any student at the university to receive a set amount of extra credit to be applied to the course of their choice by participating.

Mock Jurors Needed

What: Research is being conducted into jury decision making and which variables in courtrooms make the biggest impact.

Who: Participants of 18 years and older to act as mock jurors.

Time: 45 minutes – an hour

Participants' names will be added to a **$200** raffle, to be given at the end of the research period.

The study involves working as a group to come to a verdict as well as filling in a questionnaire.

Interested?

Please contact <u>Grace Wood</u>

email: glw525@york.ac.uk

Due to these being the main methods of recruitment (with the occasional participant gained through word of mouth), the majority of participants were students. Although students are a single demographic and therefore not a balanced representation of the U.S. as a whole, the diverse locations these mock juries were run in does improve the likelihood that the students themselves came from varied backgrounds. Additionally, research into how reliable student participants are in jury research found that there were no significant differences between student participant groups and non-student groups (Bornstein, 1999; Bornstein et al., 2017). Therefore, it is possible that this study is just as ecologically valid as one that had a more diverse demographic.

Mock jury participants were not paid for their involvement. However, they were given the opportunity at the end of each mock trial to leave their email address to be added to a raffle for $200. This raffle, and participants chance to win it, was

part of the advertisement used to recruit participants. I chose this method for two reasons: Firstly, it was less expensive to offer a large sum of money to one person than to pay every single participant a smaller sum of $5; secondly, I believed that more people would be motivated to participate by the chance to win a larger sum of money than for a guaranteed small sum that may not even cover transportation costs. After all my data collection was completed, each email address given was inputted into http://www.randomresult.com/pick.php in order to choose a random winner of the $200. The winner was then emailed and informed that they would be receiving the $200, sent via PayPal.

The mock jury version ran for three months, with a total of 82 participants across 18 mock juries. Nine groups watched the GAE stimulus video and Nine groups watched the Southern American English stimulus video, giving an even split between the two guises. Out of 82 participants, only 19 (23%) were male. The rest were female. This difference in numbers was accounted for in the analysis. When asked about occupation, 63 participants self-reported themselves as students, while the other 19 wrote a type of full time employment. Due to that, despite having recruited through universities, I considered those 19 as "professionals" and not students in the analysis. Further breakdown of the participant demographic information can be seen in Tables 25-29.

As seen in Table 25 below, there was a near equal number of both mock jury groups and participants between the two universities where the study was run.

University	No. of Mock Juries	No. of Participants
UCCS	8	40
FDU	10	42

Table 25. Mock Jury Numbers

Participants were again asked to self-report their ethnicities if they were willing. The majority ethnicity who participated were Hispanic, followed by those who were white. Six participants identified as being "mixed race" or having a "mixed ethnicity heritage." One participant reported themselves as "native" but did not specify in what way they were "native." It is possible they meant Native American.

148

Ethnicity of Participant	No. of Participants
Asian	1
Black	10
Hispanic	31
Middle East	1
Mixed Heritage	6
Native	1
White	24
Prefer not to specify	8

Table 26. Mock Juries Participant Ethnicity

The vast majority of participants were between the ages of 18-24, unsurprising due to having recruited through universities and therefore gaining mainly university students as participants. I did not have any participants over the age of 65.

Age of Participant	No. of Participants
18-24	74
25-34	6
35-64	2
65+	0

Table 27. Mock Juries Participant Age

Regarding participant's frequency of driving, the largest group by a significant margin was those who drove daily. No one reported driving only a few times a week.

Average Driving Regularity	No. of Participants
Never	5
Rarely	10
A couple times a month	10
A couple times a week	0
Daily	57

Table 28. Mock Juries Participant Average Driving

The majority labelled themselves as "unsure" of their specific political leanings, though the "liberal" group had the second largest number of participants. This majority could be due to how relatively young most of the participants are, in that they may still be forming their own opinions and beliefs. However, it could also be a reflection of the apathy some people hold towards politics.

Participant Political Affiliation	No. of Participants
Conservative	10
Centrist	3
Liberal	24
Unsure	33
Prefer not to specify	12

Table 29. Mock Juries Participant Political Affiliation

Lastly, seven participants reported having children and three participants said they had previously served on a jury.

6.3.8. Analysis

The US version of this study also used mixed methods, for the same reason as was explained in chapter 5.3.6. The online US Qualtrics data and the questionnaire the mock juries completed individually were coded and analysed in the

same manner as the online UK Qualtrics data. RQs I and II were again looked at, with the same specific analysis questions as the UK version:

Did the particular video (GAE or Southern American English) lead to a difference in:

Q1. Participants' verdicts?

Q2. How confident were participants in their verdict?

Q3. Whether participants were more likely to rate the defendant lower in trustworthiness, reliability, and believability characteristics?

Q4. What level of punishment was suggested?

Q5a. Whether participants changed their verdict after reading the text summary.

Q5b. If so, why?

Q6. Whether participants were more likely to offer negative opinions on the defendant?

Q7. How participants spoke about the accent of the defendant?

As the nature of the data collection for the online Qualtrics questionnaire and the mock juries were not the same, I also had analysis questions specific to the dataset:

US Qualtrics Questionnaire specific:

Q8. Did the duration of how long participants took to finish the questionnaire differ significantly depending on which video they watched?

Q9. Did the quality rating participants gave the videos differ significantly depending on which video?

Mock jury specific:

Q10. Did the deliberation length differ significantly depending on which video mock juries watched?

Q11. Were mock jurors more likely to find it difficult to come to a consensus depending on which video they watched?

Q12. By working as a group, did mock jurors' confidence in their pre-verdict match their confidence in their group verdict, increase in confidence, or decrease in confidence?

Q13. What clarification questions were asked by participants and was there any correlation to which video they had watched?

Questions 1-5a, 8-9, and 10-12 were all analysed quantitatively while the rest of the questions, 5b-7 and 13, were analysed qualitatively. Along with these variables, the demographics of the participants in both datasets were again considered in the analysis using fixed effects multinomial logistic regression.

- Gender of participant
- Ethnicity
- Age of participant
- Which region the participant lived in
- Whether the participant had children
- What the participant's politic affiliations were
- How regularly the participant drove
- If the participant had previously served on a real jury

The mock jury analysis also considered the specific occupations of the participants, as there were both students and professionals in that dataset, whereas the Qualtrics dataset had only self-reported professionals.

In addition to analysing the questionnaires used in both Qualtrics and the mock juries, the deliberation transcriptions from the mock juries were also analysed. As this was qualitative data, I used NVivo to code and examine it for trends in the topics discussed. I coded the transcriptions for five themes:

1. Discussion of the defendant's emotions and whether decisions were made based on how participants felt the defendant appeared emotionally (i.e. apologetic, sincere, defensive, etc.).
2. If 'facts' of the case were inaccurately discussed and whether any false information was corrected by other mock jurors.
3. How regularly participants interrupted one another.
4. When participants went off topic, to what extent, and how long it took them to return to relevant discussions.
5. Whenever participants requested additional information or noted the lack of information while coming to a decision on a verdict.

I analysed these themes by considering to what percentage each of them happened in every deliberation and whether any of them were more likely to occur in one video guise over another. As I read through each transcript, coded the themes, and then analysed them I kept consistent notes in NVivo regarding any trends I detected, whether that was in the themes I was specifically looking for or other observations particular to mock jury deliberations.

RQ III was also addressed in the analysis. This was done by cross-comparing the results of analysis questions 1-7 in both datasets to see if and how any of the statistical outputs differed. Although questions 8 and 10 are similar in regards to studying duration, these questions cannot be cross-compared. The duration of time for the Qualtrics data comes from how long it took each participant to finish the entire study, whereas the duration of time for the mock juries comes from how long they deliberated. The overall amount of time each group spent participating in the study was not recorded. Therefore comparing the results of questions 8 and 10 could lead to misleading conclusions.

The results of all the analyses will now be outlined. The p values are reported for all non-significant test results. The full table of results for those can be found in Appendix VII.

6.4. Results

This section is organized into three parts. Part one looks at the analysis of the Qualtrics individual responses data, part two looks at the mock juries analyses, and part three looks at the cross-comparison of the two datasets in order to answer RQ III. In each section all of the quantitative data results were considered first, and then followed by the qualitative analyses results.

6.4.1. Qualtrics

This initial portion of the results section looks at the data collected from individual participants using Qualtrics. It seeks to answer RQs I and II using individual mock jurors, as was used in all but one of the previous studies conducted

on this topic, as discussed in chapter 3. In each of the following tests, there was a total of 56 participants who contributed to the data.

Question 1: Did participants' verdicts differ significantly depending on the video?

As seen in Table 30, there was a near 50/50 split in the verdict given for the GAE guise while Southern American English had majority *guilty* verdicts. I used Chi square test in order to determine if there was a relationship between verdict and type of video watched.

	verdict		
	Guilty	Not Guilty	*total*
GAE	13	10	23
Southern A.E.	25	8	33
total	38	18	56

Table 30. US Qualtrics Verdict

These differences, however, were not significant (p = .129). Additionally, after conducting multinomial logistic regression tests, there was no evidence that the demographics had any effect on the verdict given. Therefore, the accent guises did not cause the participants to view the defendant differently enough that it affected the verdict, answering RQ II.

Question 2: How confident were participants in their verdict?

On a Likert scale of 1-6, the mean values for both accent guises are above neutral ratings, with the GAE mean at 4.61 and the Southern A.E. mean at 5.06. This suggests that participants felt fairly confident in their verdicts in general, with participants who watched the Southern American English video feeling particularly confident. I used an Independent t test in order to determine if there were differences in mean confidence based on video type. However, the level of confidence was not statistically significant (p = .112). The participant demographics were added to the analysis as independent variables using multinomial logistic regression, but none of them were significant.

154

Question 3: Were participants more likely to rate the defendant lower in characteristics?

As can be seen in Figure 26, the mean characteristic scores for the prosecutor are in the 4-5 range while the mean characteristic scores for the defendant are in the 3-4 range (further details in Appendix V). As the Likert scale was 1-6, the defendant's ratings hover in the middle range and show no sign of strong judgements, positive or negative, towards the video guise (GAE or Southern American English). The prosecutor's ratings are also fairly stable regardless of video guise; however, overall his are more positive than the neutral ratings given to the defendant. There is a trend in that the GAE guise has a slightly higher mean rating than the Southern American English guise, across every category except how believable the defendant is perceived to be.
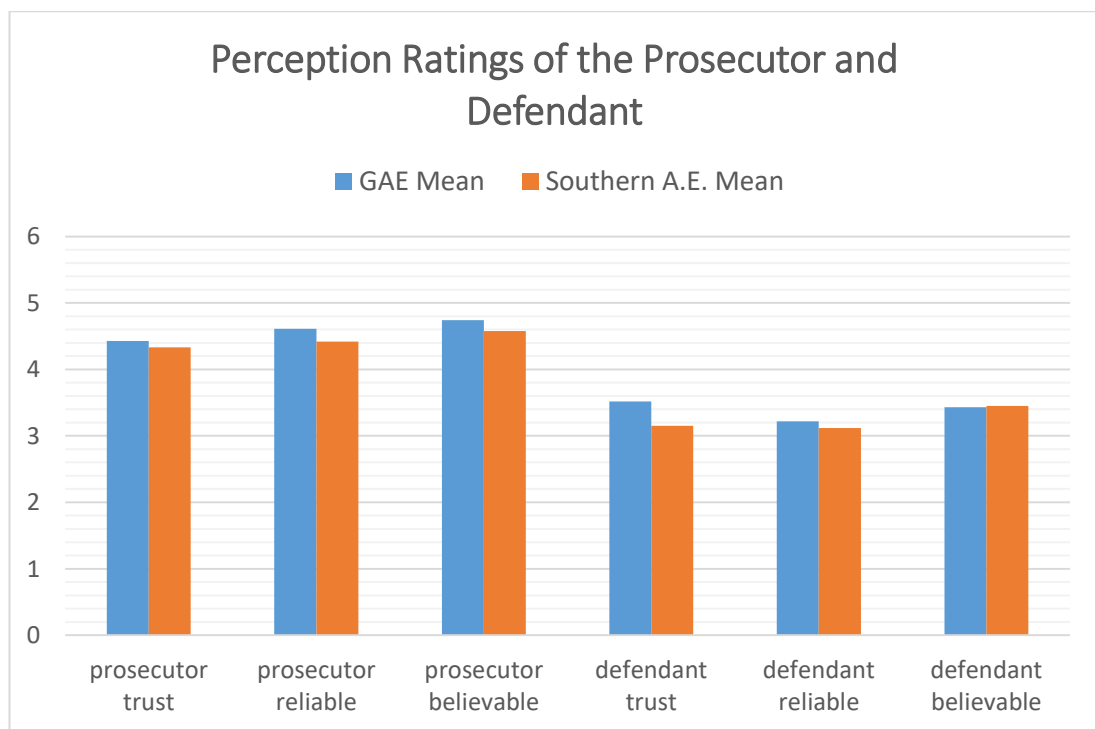
Figure 26. US Qualtrics Question 3 - Perception Ratings

I used Independent t tests in order to determine if the type of video (GAE or Southern American English) had a significant effect on the mean ratings of trust, reliability, and believability of the prosecutor and defendant. There were no statistically significant differences in any of these characteristics for either the prosecutor or defendant based on video type ($p > .05$). The trend of the standard

guise having slightly higher ratings than the non-standard guise was not strong enough to denote significance. However, considering this trend was also observed in the U.K. results outlined in chapter 5, it is possible this trend is still meaningful and reveals a general standard language ideology held by the participants.

Multinomial logistic regression tests were conducted to determine if there were significant differences in trust, reliability, and believability based on demographic categories. However, none of them showed statistical significance.

Question 4: What level of punishment was suggested?

There was a near even split in the Southern American English recommended punishments. Similarly, there was no large variation between the punishments recommended for the GAE guise. No one recommended prison in addition to another punishment. The 18 participants who rendered a *not guilty* verdict were not asked to recommend a punishment. Due to that there are only 38 responses included in this question. This can be seen in Table 31.

	punishment			
	Other	Prison	Prison+	*total*
GAE	8	5	0	13
Southern A.E.	12	13	0	25
total	20	18	0	38

Table 31. US Qualtrics Recommended Punishment

Fisher's Exact test was performed in order to determine if recommended punishment was associated with video type. The results of the Fisher's Exact test were not statistically significant (p = .506). Therefore, despite the apparent trend of the non-standard guise receiving harsher recommendations for punishment, there is no significant association between punishment and video guise.

As before, multinomial logistic regression tests were used to analyse whether any of the demographic independent variables had a significant effect on what punishment was recommended. Again, as before, no significant effects were observed.

Question 8: Did the video guise make a difference in how long participants took to finish the questionnaire?

I used an independent t test to determine if the length of time to complete the survey differed based on video type. Descriptive statistics were first assessed, where the length of time was measured in seconds. The mean length of time for the GAE condition was 835.74 seconds (about 14 minutes) while the mean length for the Southern A.E. condition was 912.76 (about 15 minute). There was less than 100 seconds difference in the means given for both guises, suggesting only a slight trend in which participants completed the questionnaire faster after hearing the GAE guise. Due to this small difference, it is unsurprising that the test result was not significant, indicating that the mean duration to complete the survey for the GAE was not significantly different from the Southern American English group (p = .598). Multinomial logistic regression was used to test whether the demographics had an effect, but there was no evidence in the statistics that they had a significant effect.

Question 9: Did the quality rating participants gave the videos differ significantly depending on which video?

Keeping in mind the additional ten participants in the Southern American English group, there is a fairly even split, as both guises received more "high quality" ratings than "low quality." This is shown in Table 32.

	video quality		
	High	Low	*total*
GAE	15	8	23
Southern A.E.	26	7	33
total	41	15	56

Table 32. US Qualtrics Video Quality Ratings

A chi square test was used to determine if there was a relationship between video quality and type of video watched. As suspected due to the similar ratings, there were not significant differences between what level of quality both video guises were assigned (p = .259). The demographics were also tested using multinomial logistic regression, but no significant effects were observed.

There is no evidence in the quantitative data that the accent guise caused participants to react differently to the defendant, answering RQ I. While trends have been observed, there have been no strong effects. What now follows are qualitative analyses of the responses participants gave in regards to questions 5-7 in order to continue answering RQ I.

Question 5: Did participants change their verdict after reading the text summary? If so, what reason did they give?

Five participants changed their verdict after reading, three out of the GAE video and two from the Southern American English video.

	Changed to "not guilty"	Changed to "guilty"
GAE	1	2
Southern A. E.	2	0

Table 33. US Qualtrics Question 5a

A chi square test was run, but returned no significance (p = .367). This is unsurprising, as the number of participants between the groups is too small to even show trends.

Two of the GAE participants changed their verdict from *not guilty* to *guilty*. Neither gave a reason for their change in verdict but said the defendant should face jail time. The third GAE participant changed their verdict from *guilty* to *not guilty*. They said they changed their verdict due to the lack of evidence when it came to the defendant's speed and culpability. The two Southern American English participants changed their verdict from *guilty* to *not guilty*. One said it was due to the lack of concrete evidence, while the second said that while the defendant should be punished for speeding, it was the children's fault for coming suddenly into the road. Due to the varying nature of the reasons given for their change, as well as the variation within the type of change (guilty to not guilty and vice versa) and the lack of statistical significance, these five participant responses ought to be as individual outliers. As there were not any discernible patterns within their demographics, this is further evidence that these five participants were outliers.

158

Question 6: Were participants more likely to offer negative opinions on the defendant in the Southern American English guise?

When participants were asked what they found *most* convincing about the defendant their responses mainly focused on his actions and his attempts to avoid the accident. Within both guises participants commented on the defendant's actions (e.g. "he swerved and tried to avoid the boys"), on the defendant's emotions (e.g. "he seemed sorry"), sympathized with the defendant based on how quickly accidents occur making it difficult to avoid such situations, or did not believe there was anything positive about the defendant's testimony. One SSBE participant simply noted that there are many outside variables that affect situations and these ought to be considered before a verdict is given. A Southern American English participant noted that the defendant had not received a speeding ticket, therefore calling into question the validity of the prosecution's argument that the defendant had been speeding at the time of the accident. The total breakdown of categories by guise can be seen below in Table 34.

	Actions	**Emotions**	**Quick**	**Other**	**All negative**	*Total*
GAE	12	3	2	1	5	*23*
Southern A.E.	13	7	5	1	7	*33*

Table 34. US Qualtrics Most Convincing

Although there are some slight differences in numbers based on category, and at first glance it would appear there were fewer negatives for the GAE guise, it is not by a large enough margin to be significant. Additionally, it must again be noted that there were 10 fewer respondents for the GAE guise than the Southern American English guise. Therefore, once this difference in numbers has been accounted for, the variation in the category responses depending on the guise has even less significance. The demographics of the participants did not appear to affect any of the category responses.

Similarly, the results for the question regarding what participants found *least* convincing about the defendant were too similar per category to be significant. As before, participants from both guises commented on the defendant's actions, emotions, lack of memory, and his speech as negative contributors to how convincing he was. One GAE participant said that the least convincing aspect was how poor the overall quality of the video was. This can be viewed in Table 35. Once again the demographics did not appear to have an effect.

	Actions	Emotions	Memory	Speech	Other	All positive	*Total*
GAE	7	3	8	1	1	3	*23*
Southern A.E.	14	2	8	4	0	5	*33*

Table 35. US Qualtrics Least Convincing

I chose to combine the overall positive and negative comments as an additional way to analyse participants' comments. I collapsed all of the categories into either *positive* or *negative*, which can be seen in Table 36. I combined the positive and negative categories using the same methods as described in chapter 5. However, once again taking into consideration the variation in the total numbers, there are no significant differences between the two guises. Both were slightly more likely to be viewed negatively, though not enough to be significant. Equally there were no discernible patterns within the demographics to suggest other confounding variables.

	Positive	Negative	*Total*
GAE	21	25	*46*
Southern A.E.	31	35	*66*

Table 36. US Qualtrics Combined Overall Convincing

160

Question 7: How did participants speak about the accent of the defendant?

When given the GAE guise, participants most often described it as "typically American" or said that they did not notice an accent. A few said the defendant just sounded "white" and others commented on his emotions, saying he seemed credible and apologetic, if a bit defensive. As for the Southern American English guise, a number of participants also labelled it as simply "American." However, there were additional comments saying he sounded like he was "blue collar," less educated, and that he sounded "moronic" and "vague." One participant said he sounded "weak and guilty." While some participants were able to pinpoint it as Southern American English, a few participants labelled him as a New Yorker, Latino or simply "white immigrant," showing that although they recognized it was a non-standard accent, they were unable to correctly place it.

I also looked at how participants described their own accents in order to better understand the language attitudes they brought to this study. 49 participants in total chose to answer this question, with the full breakdown shown in Table 37. 26 participants described their speech as "normal," "neutral," or said they did not have an accent. These all fit within the non-accent myth and would suggest they view their speech as being the standard dialect. A further five said they were Midwestern, which is often put forward as the location where the standard originates (Lippi-Green, 2012; Tamasi & Antieau, 2015). With this in mind, those who specified Midwestern may also believe they speak the standard. Seven participants said they were Southern, though one said they were only "slightly Southern" while another said they were "Southern but with Northern vowels." These qualifiers would suggest that the two participants were distancing themselves Southern American English. Five participants specified a particular city along the eastern seaboard of the United States; one from Boston, two New Yorkers, one from Philadelphia, and one from Pittsburgh. These cities are all known to have distinctive non-standard dialects and the participants seemed proud of it (e.g. the one from Boston specified: "Boston, every day and always"). Four of the participants said they had a type of foreign accent, and the final two participants were unsure what their accents were.

Non-accent myth	Midwestern	Southern	Eastern City	Foreign	Unsure
26	5	7	5	4	2

Table 37. US Qualtrics Self Reporting Accents

Similar to the findings in chapter 5, the results of question 7 here would suggest that participants do have specific attitudes and stereotypes towards different varieties of language. This can be seen in the Southern American English guise being called "moronic" and "less educated," among other things. Additionally, the fact that 26 participants believed the non-accent myth suggests that the standard is seen as what is normal to them, thus opening the possibility that any other variety may be abnormal. The two participants who distanced themselves from Southern American English appear to accept the overt prestige given to the standard, whereas the five participants from Eastern cities appear to be proud of their non-standard variety, showing the covert prestige they attach to those varieties. However, once again these language attitudes do not appear to have had any noticeable effect on their responses throughout the study. Therefore, the answer to RQ I is *yes*, the defendant was perceived differently, but there is no statistical evidence to show that this caused the defendant to also be treated differently, and thus the answer to RQ II is *no*.

6.4.2. Mock Juries

This second part of the results section looks at the data collected from the groups of mock juries in the U.S. It seeks to answer RQs I and II with the use of a group dynamic. In each test there was as total of 82 participants providing data across 18 mock jury groups.

Question 1: Did mock juries' verdicts differ significantly depending on the video?

Because the verdict was a group decision, Table 38 shows the verdict columns for the number of *mock juries* (not individual participants) who agreed upon *guilty* or *not guilty* and this question was analysed at the group level. It is clear that regardless of the accent guise, the vast majority of mock juries returned a guilty verdict. Only one mock jury found the defendant *not guilty*.

162

video	No. of participants	verdict		
		Guilty	Not Guilty	*total*
GAE	40	9	0	9
Southern A.E.	42	8	1	9
total	82	17	1	18

Table 38. US Mock Juries Verdict

Chi square test was used here in order to determine if there was a relationship between verdict and type of video watched. These results were not significant (p = .303). It is worth noting that although overall juries found the defendant *guilty*, the one *not guilty* verdict was given to the Southern American English guise. While this is an outlier, it is in contradiction to the results I predicted. I expected that the non-standard guise would be perceived more poorly than the standard, yet this is not what occurred in the verdicts. There were no significant differences, suggesting that participants were not affected by the guise when giving their verdicts, answering RQ II. Furthermore, the single outlier is a positive outcome for the non-standard guise rather than a negative outcome.

Because question 1 was analysed at the group level, rather than at the individual level, the demographics of the individual participants were not tested.

Question 2: How confident were participants in their verdict?

Mock juries who watched the GAE video had a mean confidence rating of 5.08 and those who watched the Southern A.E. video had a mean confidence rating of 5.29. As participants were given a Likert scale ranging from 1-6, a mean rating of 5 suggests overall they felt confident in their verdicts as it is above the neutral (3-4) range. This is regardless of the accent guise condition. I used an Independent t test in order to determine if there were differences in mean confidence based on video type. These small differences in mean confidence ratings was not significant between the accent guises (p = .335). Thus, participants felt just as confident in their verdicts after listening to the standard guise as they did after listening to the non-standard guise. Using multinomial logistic regression, the participant demographics were

added to the analysis of question 2. However, no statistically significant differences were observed.

Question 3: Were participants more likely to rate the defendant lower in characteristics?

Responses were again given on a Likert scale from 1-6. Shown in Figure 27 below, only one of the ratings were in the 3-4 neutral range. The mean scores for the prosecutor are all between 4 and 5, suggesting that overall he was viewed positively. In contrast, the defendant appears to have been viewed more negatively, as most of his scores are between 2 and 3. Of note is that the only mean score for the defendant that is higher (entering the neutral range) is the *believable* characteristic in the Southern American English guise. Overall, there is a general trend that the non-standard guise was reacted to more positively in every category except how *trustworthy* the prosecutor was perceived to be. Further details about the mean numbers in Figure 41 can be viewed in Appendix V.
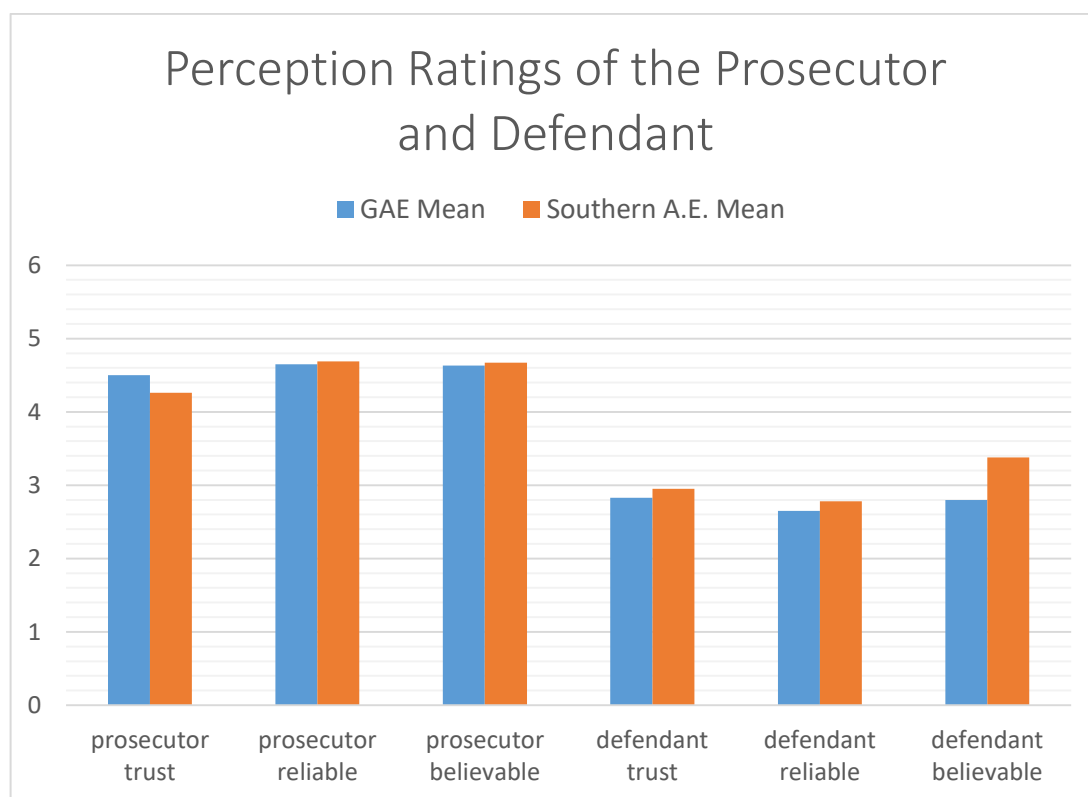
Figure 27. US Mock Juries Question 3 - Perception Ratings

I used Independent t tests in order to determine if the type of video (GAE or Southern American English) had a significant effect on the mean ratings of trust, reliability, and believability of the prosecutor and defendant. As laid out in Table 39, these trends were not statistically significant, although the rating of the defendant in regards to his *believability* approached significance at p = .095. There is no evidence, however, that this near significance score is meaningful. Participants did not mention this specific characteristic in any of their other responses. Additionally, this is not a trend that is observed in either of the previous dataset results.

	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
prosecutor - trust	.865	80	**.390**	.238	.275
prosecutor - reliable	-.169	80	**.866**	-.040	.240
prosecutor - believable	-.152	80	**.879**	-.042	.274
defendant - trust	-.446	80	**.657**	-.127	.286
defendant - reliable	-.451	79	**.653**	-.130	.290
defendant - believable	-1.687	80	**.095**	-.581	.344

Table 39. US Mock Juries Perception Ratings – Independent T Test

After running multinomial logistic regression tests, there was also no evidence to suggest that the demographics of the participants had any effect.

Question 4: What level of punishment was suggested?

Participants were less likely to recommend harsher punishments like prison and something additional, such as community service or a fine, and much more likely to recommend lesser punishments, such as only community service or a fine. Overall, the recommended punishments had a near even split between the accent guises in each category. As before, the participants in the jury who rendered a *not guilty* verdict were not asked this question. This can all be viewed below in Table 40.

	punishment			
	Other	Prison	Prison+	*total*
GAE	21	15	4	40
Southern A.E.	21	12	6	39
total	42	27	10	79

Table 40. US Mock Juries Recommended Punishment

I performed a Fisher's Exact test in order to determine if recommended punishment was associated with video type. As suspected based on the overall even split, the results of the Fisher's Exact test were not significant, with $p = .736$. Thus neither guise caused participants to react more strongly against or for the defendant. As with the previous analysis questions, the demographics were analysed using multinomial logistic regression, but there was no significant evidence they had an effect.

Question 10: Did the deliberation length differ significantly depending on which video mock juries watched?

The deliberation times were measured in seconds, with mock juries who watched the GAE video deliberating for an average of 257.13 seconds and the mock juries who watched the Southern A.E. video deliberating on average for 234.52 seconds. Overall, the means in deliberation time between the two videos only differed by a total of 23 seconds.

An independent samples t test was performed in order to determine if deliberation duration times were different depending on whether mock juries watched the GAE video or the Southern American English video. As suspected, based on the small variations in the means, the specific video guise had no statistically significant effect on how long mock juries then deliberated ($p = .627$). This suggests that mock juries did not find it more difficult (therefore taking longer) to come to a unanimous verdict after watching a specific video. As with the previous analysis questions' results, this is further evidence that the accent guise had no discernible effect on participants' judgments.

As with question 1, because question 10 was analysed at the group level, the individual participant demographics were not added to the analysis.

Question 11: Were mock jurors more likely to find it difficult to come to a consensus depending on which video they watched?

While the majority of participants reported no difficulties when coming to a unanimous verdict, the seven participants (across three mock juries) who did find it hard had all watched the GAE video. This breakdown can be viewed in Table 41.

	difficult		
	No	Yes	*total*
GAE	33	7	40
Southern A.E.	42	0	42
total	75	7	82

Table 41. US Mock Juries Consensus Difficulty

I used a Chi square test here in order to determine if there was a relationship between which video groups watched and whether the individuals considered it difficult to come to a consensus. The test was statistically significant ($p = .005$), shown in Table 42. This suggests that if participants found it difficult to come to a consensus, they were significantly more likely to have watched the GAE video than the Southern American English video. While this had no visible effect on the verdict itself, and is a result from seven out of 82 participants, this is still evidence that the accent guises did have an effect on the study results, and that the two guises were not simply interchangeable.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**8.036[a]**	**1**	**.005**
Continuity Correction[b]	5.951	1	.015
Likelihood Ratio	10.738	1	.001
N of Valid Cases	82		

Table 42. US Mock Juries Consensus Difficulty – Chi-square Test

Multinomial logistic regression tests were run to analyse what effect the participant demographics may have had in the results. However, no statistically significant differences were found.

Question 12: by working as a group, did mock jurors' confidence in their pre-deliberation verdicts match their confidence in their group verdict, increase in confidence, or decrease in confidence?

As with the final confidence scores (analysed in question 2), the ratings were given on a Likert scale of 1-6. As shown in Table 43 below, both mean scores were above the neutral range of 3-4, though the final confidence mean score was higher (indicating greater confidence) than the pre-confidence mean score.

	Mean	Std. Deviation	Std. Error Mean
pre-confidence	4.59	1.111	.123
post-confidence	5.18	.983	.109

Table 43. US Mock Juries Pre and Post Confidence

A paired t test was conducted in order to determine if the pre-deliberation confidence mean scores significantly differed from the post deliberation confidence scores. There was a significant difference between these two confidence scores ($p <$ .001), shown in Table 44. This suggests that the process of working as a group and coming to a mutual decision increased individual's confidence in the overall verdict. This is regardless of how difficult participants found working in a group to be. Furthermore it supports my hypothesis that group polarization would have an effect on the results. After running multinomial logistic regression tests, there was no evidence that participant demographics had any effect.

	t	df	**Sig. (2-tailed)**	Mean	Std. Deviation	Std. Error Mean
pre-confidence / confidence	-5.085	81	**.000**	-.598	1.064	.118

Table 44. US Mock Juries Pre and Post Confidence – Paired T Test

There has been little evidence thus far of the accent guise affecting participants' responses, beyond how difficult they found the deliberation. RQ I will

continue to be analysed through the qualitative responses mock jurors gave in regards to questions 5-7 and 13.

Question 5: Did participants change their verdict after reading the text summary? If so, what reason did they give?

10 participants changed their verdict from *guilty* to *not guilty* after reading the text summary. No participant changed their mind from *not guilty* to *guilty*, though this is not surprising considering the vast majority rendered a guilty verdict (laid out in Table 45).

	Changed to "not guilty"	Changed to "guilty"
GAE	4	0
Southern A. E.	6	0

Table 45. US Mock Juries Question 5a

A chi square test was run to analyse this, but returned no significance (p = .553). Too few participants changed their minds, with a near even split in numbers between the guises, to show any significant effects due to the accent guises.

Table 46 shows the breakdown of why participants changed their verdict per guise. The majority of participants reported that it was due to the speed limit evidence being five miles previous to the incident that changed their verdicts. They observed that a lot can happen in five miles and therefore the speeding "evidence" did not necessarily mean that the defendant had been going the same speed at the time of the incident as he had five miles previously. Two of the participants did not give a reason for their change in verdict. There were no discernible patterns within the demographics of these participants. While this result does not show any clear differences between video guises, it would suggest that mock jurors were not able to catch (or process) all the relevant information in the video, as it was also said in the video that the speed limit evidence came five miles previous to the accident.

	speed limit	no answer	*total*
GAE	3	1	4
Southern A. E.	5	1	6

Table 46. US Mock Juries Question 5b

Question 6: Were participants more likely to offer negative opinions on the defendant in the Southern American English guise?

When participants were asked what they found *most* convincing about the defendant, their responses mainly focused on his actions and his attempts to avoid the accident. Most commented on the defendant's actions (e.g. "he swerved," "he did press the brakes"), or his emotions (e.g. "he seemed remorseful"). Other participants felt he answered questions honestly (e.g. "he was willing to admit he hit the kid," "he admitted he didn't remember"), particularly in the Southern American English guise. One participant noted that it was understandable that the defendant would not remember every detail while some participants sympathized with him based on how quickly accidents occur. A total of 14 participants (nearly evenly split between accent guises) did not believe there was anything positive about the defendant's testimony. This breakdown of categories by accent guise can be seen in Table 47.

	Actions	Emotions	Memory	Quick	Honesty	All negative	*total*
GAE	18	7	1	4	4	6	*40*
Southern A.E.	17	8	0	1	8	8	*42*

Table 47. US Mock Juries Most Convincing

Although there are some slight differences in numbers based on category it is not by a large enough margin to be significant. No demographic category appeared to have an effect.

Similarly, the results for the question on what participants found *least* convincing about the defendant were too similar per category to show any significant due to the accent guise, seen in Table 48. Participants mainly found the defendant's

lack of memory regarding the specifics of the incident unconvincing, particularly regarding his inability to remember his own speed at the time of the incident. Once again participants also commented on his actions, some mentioned his emotions (e.g. "he didn't seem actually sorry for what he did"), and others commented on his perceived lack of honesty, noting his ability to evade questions. One participant's answer to this question was written illegibly and so was left out.

Two participants who had watched the Southern American English video commented on the defendant's speech as a negative quality. One of those participants wrote, "His remorse was insincere, *his bad grammar made him seem unintelligent which makes him seem like he shouldn't be operating a car* [emphasis my own]." The use of the description "bad grammar" is just as telling here as the participant also calling the defendant unintelligent. The Southern American English stimulus generally used Standard English. Therefore, there was no "bad grammar" here to ridicule; rather it is the accent alone that this participant is reacting to. Despite this harsh (and inaccurate) perception of the defendant, this specific participant did not recommend prison as a punishment and appears to have responded in every other way to the questionnaire with no overt aggression or intent to discriminate.

	Actions	**Emotions**	**Honesty**	**Memory**	**Speech**	**All positive**	*total*
GAE	7	7	6	17	0	2	*39*
Southern A.E.	13	4	6	15	2	2	*42*

Table 48. US Mock Juries Least Convincing

Participant demographics did not appear to have an effect in what was found least convincing about the defendant.

As before, I chose to combine the overall positive and negative comments as an additional way to analyse participants' comments (again using the method described in chapter 5). I collapsed all of the categories into either *positive* or *negative*, which can be viewed below in Table 49. However, there were no

significant differences between the two guises. Both were slightly more likely to be viewed negatively, though not enough to be significant. This is unsurprising, as most of the mock juries found the defendant at fault for injuring a child, and therefore would have been more likely to view him in a negative light. Equally there were no discernible patterns within the demographics to suggest other confounding variables.

	Positive	**Negative**	*total*
GAE	36	43	*79*
Southern A.E.	36	48	*84*

Table 49. US Mock Juries Combined Overall Convincing

Question 7: How did participants speak about the accent of the defendant?

When asked to describe the defendant's accent, participants who had watched the GAE video were more likely to respond regarding their perception of his emotions, whether positive (e.g. "he seemed nervous") or negative (e.g. "he was insincere"). The rest generally labelled it as "American" or said they had not noticed an accent. The vast majority of participants who watched the Southern American English video labelled it as "Southern," "Country," or in one case "Hillbilly." Most of the participants who wrote those added a comment regarding how he sounded uneducated, unintelligent, or unprofessional.

As with the previous datasets, I also looked at how participants described their own accents in order to better understand the language attitudes they brought to this study. 76 out of 82 participants chose to answer this question. 48 participants said they did not have an accent, that their accent was "neutral" or "normal," or that it was just "American," all falling into the non-accent myth category. An additional four said they had a "Midwest" accent. As this is often considered to be where the standard American dialect originates, these four participants may also believe they speak the standard and do not have an accent. Three participants said they had a twinge of a Southern accent mixed with something else. Four participants said they had an accent from the Northeast, specifically listing either Boston, New York, or New Jersey. Six participants said they had a foreign English accent. The final 11

participants did not describe an accent but rather an emotion (e.g. "confident," "friendly," "shy," etc.). These categories can be viewed in Table 50.

Non-accent myth	Midwestern	Southern	Northeast	Foreign	Emotion
48	4	3	4	6	11

Table 50. US Mock Juries Self Reporting Accents

Again it is possible to see that overall, participants did have strong and overt language attitudes towards specific varieties of language. The GAE guise was never called "unprofessional" or "uneducated" even though the Southern American English guise was perceived as that by the majority. This fits with the academic research of the language ideologies held towards these varieties (Bayard et. al., 2001; Frumkin, 2007; Heaton & Nygaard, 2011). Participants' attitudes towards their own speech, particularly in regards to seeing themselves as "neutral" or "normal," again shows the prevalence of language attitudes. However, there is no evidence in the quantitative results to suggest that these language attitudes led them to act in a discriminatory manner.

Question 13: What clarification questions were asked by participants and was there any correlation to which video they had watched?

The only clarification question I was asked while running this study was regarding the questions on accent, and specifically what did I mean by "accent"? Of the nine groups who asked this question, seven of them were groups who had watched the GAE video. This would suggest that for most of the GAE groups, the standard guise did not raise any thoughts regarding accent whereas the Southern American English guise was noticed enough that the majority of groups did not need to question what I meant by "accent." Additionally, even after explaining what I meant by *accent* (explicitly stating that I was looking for a description of how they spoke with regards to pronunciation), many of the GAE participants still described their perception of the defendant's emotions instead of specifically his accent. The Southern American English participants not only did not generally need clarification,

they also described their perception of his accent. This is evidence of the effect non-standard accents have, in being perceived as "unusual" and therefore raising awareness of a person's speech. However, as noted previously in the results, there is no evidence that this additional awareness of the defendant's accent caused the participants to act in any way differently from the participants who were not consciously aware of his accent.

6.4.2.1. Deliberations Analysis

The final part of this section is a qualitative analysis of the mock juries' deliberation transcripts. I used the five questions below to guide my analysis and took consistent notes of any trends I noticed.

1. Discussion of the defendant's emotions and whether decisions were made based on how participants felt the defendant appeared emotionally (i.e. apologetic, sincere, defensive, etc.).
2. If 'facts' of the case were inaccurately discussed and whether they were corrected.
3. How regularly participants interrupted one another.
4. When participants went off topic, to what extent, and how long it took them to return to relevant discussions.
5. Whenever participants requested additional information or noted the lack of information while coming to a decision on a verdict.

The amount of occurrence for each of these topics was calculated using NVivo. Each time a portion of speech was identified as fulfilling one of the five topics, it was highlighted and "tagged" as belonging to the specific topic. In this way I used NVivo to organise the topics. This then produced a percentage of overall occurrence for each topic, based on the total amount of dialogue across all the mock jury deliberations.

Of the five questions above, I did not find any statistically significant differences, or even trends, in their occurrences across mock juries that had watched the GAE video and mock juries who had watched the Southern American English video. On top of that, each of the above category questions were generally rare occurrences, the total percentages of which can be viewed below in Table 51.

Emotions	2.32%
Incorrect facts	0.8%
Interruptions	12.38%
Off Topic Discussions	0%
Request for more information	2.21%

Table 51. Mock Jury Deliberation Category Percentages

Overall, emotions were only discussed briefly in seven out of 18 juries. This only covered 2% of the discussions across all of the deliberation transcripts. Although many participants commented on the defendant's perceived emotions when discussing his most or least convincing characteristic, this rarely became a part of the deliberations. Some participants felt the defendant was "defensive" or not truly "sorry" for what had happened, leading some participants to perceive a strong sense of guilt. Despite this, there was no evidence in their discussions that he was convicted based on those perceived emotional traits. The discussions on the defendant's perceived emotions all arose at the start of deliberations, and each mock jury quickly moved on to discuss the actual "facts" of the trial in order to make their verdict. Furthermore, as shown in Table 52, the accent guise did not have an effect on whether mock juries were more likely to discuss the defendant's emotions.

Accent Guise	Mock Jury	% of Deliberation Time
Southern A.E.	Jury B	14.38%
GAE	Jury E	0.35%
Southern A.E.	Jury H	1.26%
GAE	Jury I	4.56%
Southern A.E.	Jury J	0.99%
GAE	Jury K	9.11%
GAE	Jury Q	7.39%

Table 52. Mock Jury Deliberations - Emotions

Facts of the case were wrongly cited in less than 1% of the total deliberation transcripts, occurring in only five of 18 mock juries. Moreover, in most of those the

false facts were corrected by a fellow mock juror. There was only one case, involving Jury J, where an incorrect fact (regarding the precise speed limit) was not accurately corrected. However, the participants were only off by a single digit and therefore the lack of correction was insignificant. There were a few instances when participants created facts about the incident. This may have been a subconscious way for them to fill in gaps of knowledge. For instance, a participant in Jury L said that the defendant should have known better than to speed in a residential area. However, never once was the area described as residential in the video. Rather, it is likely that this participant assumed it was residential due to the children cycling in that area. Table 53 shows that once again the accent guise did not have an effect on whether a jury was more or less likely to get information wrong, as there is an even split between guises.

Accent Guise	Mock Jury	% of Deliberation Time
GAE	Jury A	5.03%
Southern A.E.	Jury J	2.38%
Southern A.E.	Jury L	1.19%
GAE	Jury Q	2.84%

Table 53. Mock Jury Deliberations - Incorrect Facts

Participants tended to interrupt each other around 12% of the time across all of the deliberations. This occurred to some extent across 15 of the 18 mock juries. Only Juries A, P, and R never interrupted each other. However, across the majority of the deliberation times, they were willing to listen to one another's arguments and demonstrated turn-taking. When they did interrupt each other, it was usually to show their support and agreement with the speaker. Most of the time this was a brief interruption (e.g. "mmhmm," "yeah," "I agree"), but occasionally this would be more extensive, as the interrupter felt the need to explain why they vehemently agreed – and in rare cases vehemently disagreed. In all of these cases it was clear that participants were listening to one another and responding to each other's points, therefore taking their roles as mock jurors seriously. There was no evidence of the accent guise having an effect, as eight of the GAE and seven of the Southern American English mock juries interrupted one another to some degree.

Further evidence that the mock juries took their role (and the overall study) seriously, is that not a single group, no matter how long they deliberated, went off topic. Only Jury J came close, and that was because they had decided to draw a diagram of the incident to assist in their understanding of what occurred and therefore make a more informed decision. This group would occasionally interrupt themselves to mention an addition that needed to be made to the diagram. Since that was still relevant to their decision making processes, however, I did not count it as going off topic. Thus every group stayed fully focused on their goal to come to a unanimous verdict and deliberated until that had been accomplished.

Finally, participants mentioned a need for additional information in 2% of all the deliberation transcripts. As participants were basing their verdict off of a two minute video, I was surprised that they did not request additional information more frequently. Furthermore, the lack of information did not appear to have an effect on their deliberations and eventual verdicts. While most juries did make at least one comment regarding the lack of information or a general desire to have had a longer video or more background on the case from the researcher (myself), 17 out of 18 juries still returned a *guilty* verdict. Jury J gave a *not guilty* verdict for the very reason that there was a lack of evidence and they could not be sure beyond reasonable doubt that the defendant was at fault. In many of the deliberations, other juries noted that the defendant openly admitted to hitting the child and found that to be evidence enough. This suggests to me that there was a general misunderstanding that their job was to discover whether the defendant was at *fault* for the incident, not whether the incident occurred.

It is unclear whether the accent had an effect on whether participants requested additional information. Table 54 shows that of the nine mock juries who mentioned this, six of them had watched the GAE video. However, when the percentages are added up, the three groups who watched the Southern American English video spent longer discussing the need for more information, and brought it up as an issue more regularly. It is possible that participants who watched the GAE video were more likely to bring it up at least once. However, because this occurred to such a small degree across all of the deliberations (2%) and it generally did not have an effect on the eventual verdict, I do not believe the accent guise had any meaningful effect on whether participants requested further information.

Accent Guise	Mock Jury	% of Deliberation Time
Southern A.E.	Jury B	13.9%
GAE	Jury C	1.21%
GAE	Jury E	2.02%
GAE	Jury I	0.94%
Southern A.E.	Jury J	4.76%
GAE	Jury K	1.37%
Southern A.E.	Jury L	3.42%
GAE	Jury M	3.57%
GAE	Jury Q	8.56%

Table 54. Mock Jury Deliberations - Requests for Information

In addition to the five categories above, I made notes as I read through transcriptions regarding whether groups realised a speed camera five miles away was in fact a long distance. All the groups discussed the speed limit evidence and used it to make their final decision. 17 out of 18 mock juries found the evidence of speed to be quite damning for the defendant, as it showed careless driving on his part, putting him at fault. Only Jury E discussed just how far away five miles was, including how speed limits can often change within a period as long as that. As a group, they decided that the speed limit information was irrelevant and therefore could not be used when coming to a verdict. They repeatedly mentioned a desire for more information, but eventually chose to go with a *guilty* verdict because someone had been hit and therefore they felt there must be a consequence.

Another observation I made was how 15 out of 18 mock juries started off their deliberation unanimously, with each mock juror already feeling that a guilty verdict was the correct one. However, despite that, every mock jury still felt the need to discuss why they thought the defendant was guilty and justify that verdict. Here is additional evidence that the participants took their role as a mock jury and sought to discuss the case as a group regardless of their individual opinions on the verdict. Of the three juries who did not start off unanimous, it was only one mock juror who did not agree with the group, and in each case that one person was eventually swayed to agree with the majority verdict.

6.4.3. Research Question III

This final results portion seeks to answer RQ III regarding the use of individual participants in comparison to participants working together in groups. By cross-comparing the U.S. Qualtrics dataset analysed in section 6.4.1. and the U.S. mock juries dataset analysed in section 6.4.2. it will become clearer if the results were significantly affected by the group dynamic. There is a difference of 26 total participants between the datasets, as the U.S. Qualtrics dataset had a total of 56 participants while the mock juries dataset had a total of 82.

Question 1: Did participants' verdicts differ significantly depending on the video?

As commented on previously, there was a larger spread in verdicts rendered in the Qualtrics data than in the mock juries data. The mock juries overwhelmingly convicted the defendant, whereas roughly a third of the Qualtrics participants acquitted the defendant, as shown in Table 55. For this test the mock juries were analysed at the group level, as they did not give 82 individual verdicts but rather 18 group verdicts.

	verdict		
	Guilty	Not Guilty	*total*
U.S. Qualtrics	38	18	56
Mock Juries	17	1	18
total	55	19	74

Table 55. RQ III Verdict

In order to determine if the two datasets were likely to give the same verdict a Chi-square test was performed. As seen in Table 56, there was a significant difference between the frequencies of Qualtrics responses and mock jury responses ($p < .05$). The mock jury participants were significantly more likely to convict the defendant than the Qualtrics participants.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**5.046**[a]	**1**	**.025**
Continuity Correction[b]	3.749	1	.053
Likelihood Ratio	6.253	1	.012
N of Valid Cases	74		

Table 56. RQ III Verdict – Chi-square Test

Question 2: How confident were participants in their verdict?

The confidence levels for the mock juries were taken from the ratings participants reported individually after they had deliberated as a group. Both datasets had mean ratings above the neutral range (3-4), the US Qualtrics average confidence rating being 4.88 and the mock juries' average confidence rating being 5.18, indicating that participants in both datasets were fairly confident in their verdicts. I performed an independent t test in order to determine if there were significant differences in verdict confidence between U.S. Qualtrics participants and the mock jury participants. The results of the t test were not statistically significant ($p = .08$). However, they did near significance, with the mock jury participants reporting higher levels of confidence in their verdicts than the Qualtrics participants. This could show a trend towards group polarization, further supporting my prediction that group dynamics would be evident.

Question 3: Were participants more likely to rate the defendant lower in characteristics?

Between the two datasets, the mean ratings given for the prosecutor's characteristics are near identical. However, there is more variation in the ratings of the defendant. Figure 28 illustrates that the mock juries perceived the defendant more poorly than the Qualtrics participants. As the mock juries mainly found the defendant guilty, while the Qualtrics participants were more evenly split between verdicts, it is unsurprising that the mock juries therefore gave lower ratings to the defendant. Further information about the mean ratings in Figure 28 can be found in Appendix V.
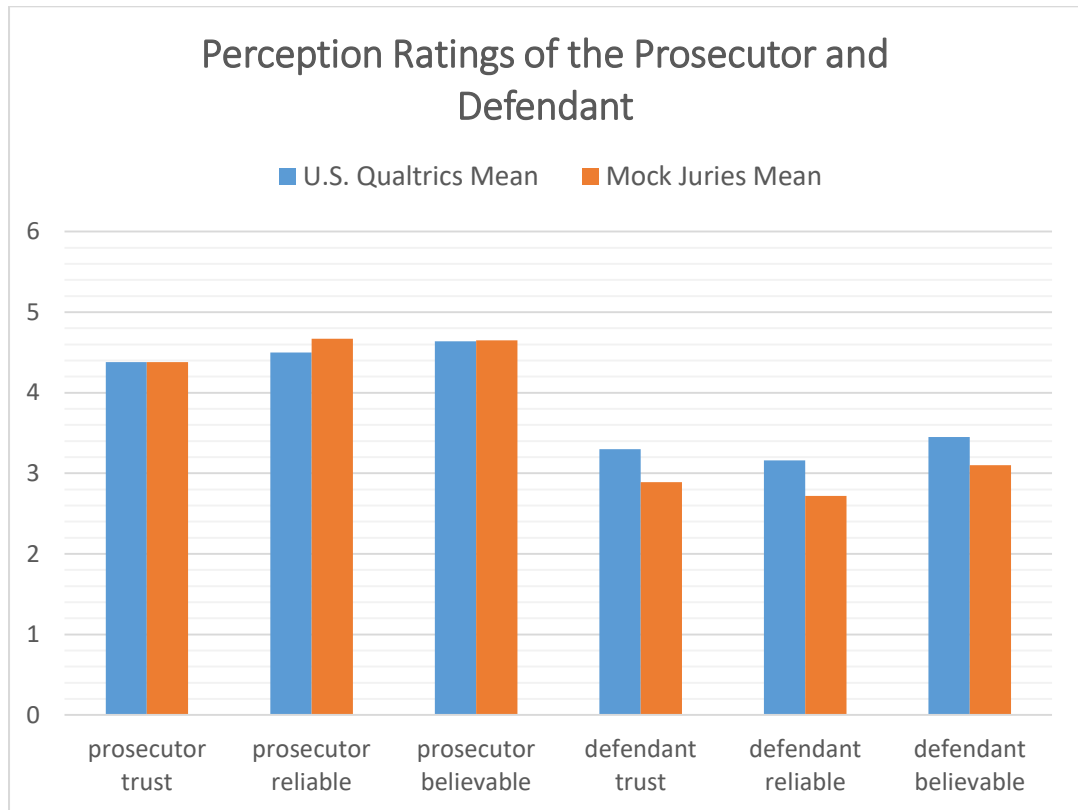
Figure 28. RQ III: Question 3 - Perception Ratings

Independent t tests were conducted in order to determine if there were any statistically significant differences in trust, reliability, and believability of prosecutor and defendant between the Qualtrics dataset and the mock jury dataset. The results of the t test did not reveal any statistically significant differences in the mean ratings for the characteristics of the prosecutor and defendant between the two datasets. However, as shown in Table 57, two results neared significance. The defendant's characteristics of *trustworthiness* and *reliability* were close to being significantly different between the two datasets, in that the Qualtrics ratings are all higher than the mock jury ratings. Although this trend does not encompass all of the defendant's qualities, it still suggests that the difference in verdicts given between the datasets may have caused participants to view the defendant differently.

	t	df	**Sig. (2-tailed)**	Mean Diff.	Std. Error Diff.
prosecutor - trust	-.015	136	**.988**	-.003	.200
prosecutor - reliable	-.937	136	**.350**	-.171	.182
prosecutor - believable	-.017	136	**.986**	-.003	.204
defendant - trust	1.725	136	**.087**	.413	.240
defendant - reliable	1.744	136	**.083**	.445	.255
defendant - believable	1.260	136	**.210**	.349	.277

Table 57. RQ III Perception Ratings – Independent T Test

Question 4: What level of punishment was suggested?

Seen in Table 58, the mock juries recommended prison in addition to something else (e.g. driving lessons, community service, monetary fine, etc.). The individual Qualtrics participants had a near even split between "other" and "prison" recommendations, but none of them recommended prison in addition to another penalty. This may be another reflection of the effects of group polarization, causing those who had worked in groups to feel more confident and therefore be more willing to give "extreme" responses. As before, participants who acquitted the defendant did not offer recommended punishments and therefore are not included in this test.

	punishment			
	Other	Prison	Prison+	*total*
U.S. Qualtrics	20	18	0	38
Mock Juries	42	27	10	79
total	62	45	10	117

Table 58. RQ III Recommended Punishment

I ran a Fisher's Exact test in order to determine if recommended punishment was associated with dataset type. The test returned significance, at p < 0.001 in Table 59. This result confirms that participants gave a different recommended punishment depending on whether they worked as a group in a mock jury or if they completed the study individually on Qualtrics.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	35.787[a]	3	.000	.000
Likelihood Ratio	45.568	3	.000	.000
Fisher's Exact Test	**39.433**			**.000**
N of Valid Cases	138			

Table 59. RQ III Recommended Punishment – Chi-square Test

However, I suspect that this significant difference is due to the 10 mock jury participants who recommended prison in addition to another penalty. It is arguable that any type of prison sentence (whether there is an additional punishment or not) is more severe than recommending the defendant be sentenced to community service or forced to pay a fine. Therefore, I chose to collapse the punishment categories into only two: *lenient* (containing the "other" recommendations) and *harsh* (containing the prison and prison+ recommendations). Once this was accomplished, both datasets had a near even split between lenient recommendations and harsh ones, shown in Table 60.

	Punishment		
	Lenient	Harsh	*total*
U.S. Qualtrics	20	18	38
Mock Juries	42	37	79
total	62	55	117

Table 60. RQ III Recommended Punishment – Condensed

I performed a Chi-square test, which did not return significance (p = .957). Although the mock jury dataset did have a larger degree of separation from the Qualtrics dataset in the harsh punishment recommendations category, once the difference in overall participant numbers was accounted for, this difference was not significant.

There has been some evidence, in both trends and statistically significant results, that working as a group caused participants to react differently from those who completed the study individually. What now follows is a cross comparison between the individual questionnaire and the mock juries results of the qualitative

analyses to the responses participants in both datasets gave in regards to questions 5-7.

Question 5: Did participants change their verdict after reading the text summary? If so, what reason did they give?

The five Qualtrics participants who changed their verdict all did so for separate reasons. Two of the participants changed from *not guilty* to *guilty*, while the other three changed their verdicts from *guilty* to *not guilty*. There was thus no systematic pattern here. This was described in Table 33 in section 6.4.1. As for the mock juries, 10 participants changed their verdict after reading, all from *guilty* to *not guilty*. Of the eight participants who gave reasons all said it was due to the lack of actual evidence regarding the defendant's speed at the time of the incident, outlined in Tables 45 and 46 in section 6.4.2. It could be suggested that the mock jurors' similar responses resulted from working together in groups and having all discussed the speed limit as evidence of the defendant's guilt. However, it is difficult to say with any confidence that it was the group dynamics versus working as an individual that caused changes this small.

Question 6: Were participants more likely to offer negative opinions on the defendant in the Southern American English guise?

After analysing both datasets in Tables 34-36 and 47-49 it was found that accent did not appear to make a significant difference in whether participants attributed positive or negative characteristics to the defendant. The only visible difference between the two datasets is that the mock jurors spoke about the defendant's honesty, or in some cases apparent lack of, as either a positive or negative while no Qualtrics participant mentioned their perception of the defendant's honesty.

Question 7: How did participants speak about the accent of the defendant?

Again, the responses between the two datasets were fairly similar. Both often said they did not notice an accent when given the GAE guise, or labelled it as "normal" or simply "American." When given the Southern American English guise participants in both datasets were more likely to place it as being Southern, and often labelled it as uneducated or unintelligent. The biggest difference came in that many of the mock jurors misunderstood the questions on "accent" when given the GAE guise. When asked to describe the defendant's accent, instead of talking about the defendant's phonetics or speech pattern, they mentioned the emotions they believed were conveyed in his tone of voice. It is unclear where this misunderstanding arose from, as many of the GAE jury groups asked me to clarify this question before they answered it. Additionally, this was not something any Qualtrics participant mentioned. In both datasets it was clear that participants had brought language attitudes with them to this study. However, there was no evidence to suggest that participants were significantly likely to act upon these language attitudes and actively discriminate against the defendant due to his accent.

6.5. Discussion

Similar to the U.K. version, this study sought to answer RQ I "Will a defendant be perceived differently by a jury based on whether he speaks in a standard or regional accent?" and RQ II "If a defendant is perceived differently, will this go as far as affecting the verdict?" Additionally this chapter examined RQ III "Does collecting individual responses lead to reliable results on how a jury would function or does the group condition make a large enough difference to invalidate research results that used only individual participation?" These were all addressed in the context of the United States. The analysis addressed these questions through the specific questions laid out in section 6.3.8. The deliberation transcripts from the mock juries were looked at qualitatively for any specific themes or references.

In all three results sections the answer to RQ I was *yes*, due to the presence of language attitudes. However, the answer to RQ II was *no*, as those same language attitudes did not cause the defendant to be treated differently in any way. As discussed in chapter 5, this does not imply that accent discrimination is a myth and

never happens. Rather, it simply adds nuance by suggesting that it is not an automatic phenomenon, and a non-standard dialect alone is not enough to trigger it. Regarding RQ III, there were clear ways in which working as a group caused different behaviour and outcomes in comparison to working individually. This finding aligns with psychology research. Therefore, the answer to RQ III was *yes*.

6.5.1. Qualtrics

Within the Qualtrics dataset there were no significant differences in the findings. As in chapter 5, evidence of language attitudes was found to be present, as shown by the answers given to question 7. However, as there were no significant differences in the results dependent on the accent condition, it is clear that these language attitudes did not have an effect on participant responses. This is further evidence that the presence of linguistic stereotypes and participants' language attitudes does not automatically lead to discrimination. Within many of the analysis questions, there was an apparent trend of the standard (GAE) being preferred. Although these trends were not significant in any of the tests, the fact that they continued to reappear across analysis questions could offer further evidence of (potentially weak) language attitudes being present. Overall, the U.S. Qualtrics results align with both the findings of the U.K. Qualtrics results and the findings from earlier research, namely Seggie (1983), Dixon et al. (2002), and Dixon and Mahoney (2004).

The same potential methodological missteps noted in chapter 5 also apply here. Stronger effects may be observed if a witness is used as the stimulus rather than a defendant. Additionally, the speed limit evidence may have been too strong, and its irrelevance (due to being five miles previous to the incident) may not have been clear enough. Like YE, Southern American English speakers are not stigmatised as being reckless or likely to commit a crime. Because the crime scenario did not require participants to ascribe a high level of intelligence to the defendant, there is no evidence to suggest that any of the negative perceptions of Southern American English were triggered. However, as there was variation in the verdicts, there is equally no evidence that any of these potential issues adversely skewed the results of the study.

6.5.2. Mock Juries

The mock jury dataset had significant results for questions 11 and 12, though trends were observed in some of the data that did not have statistically significant results. In Figure 27 it is clear that the prosecutor was preferred over the defendant. This could be due to the circumstances, as the defendant was standing trial for a potential crime whereas the prosecutor had done nothing perceivably wrong. As all but one jury found the defendant guilty, it is also understandable why the defendant would be perceived negatively. Yet despite this apparent negative perception of the defendant, participants were more likely to recommend punishments such as community service rather than a harsher punishment such as prison and an additional penalty. This can be viewed in Table 40. These trends were all regardless of the accent guise.

For question 11 participants reported having a significantly more difficult time coming to a consensus as a group in the GAE guise than in the Southern American English guise. While 17 out of 18 mock juries returned a guilty verdict, this result suggests that participants struggled more with giving that verdict after listening to the defendant testify in GAE than in Southern American English. The end result was the same, but the conditions did affect participants and their deliberations, even if unconsciously and in a relatively minor way. It is possible that this effect would be amplified if the non-standard accent had been more heavily stigmatized than Southern American English. In the future it would be of note to compare GAE with accents used by socially stigmatized groups, such as African Americans and AAE. Additionally, as with YE, it is possible that an accent guise is not strong enough to trigger discrimination and instead dialect guises ought to be used.

The significant result of question 12 indicated that participants' confidence in the verdict increased after working as a group and coming to a mutual decision, as compared to how confident mock jurors initially felt in their pre-deliberation verdict. This is evidence of group polarization, a well-known and well-researched psychological group dynamic phenomenon (Forsyth, 2017; Myers & Lamm, 1976). It has previously been found to occur in mock jury research and is unsurprising that

mock jury participants feel more confident about a verdict they know others agree with (Bornstein & Greene, 2011). This is a clear example of how working as a group can cause significant differences to working as an individual.

As with both the U.S. Qualtrics dataset and the U.K. Qualtrics dataset, it was evident through question 7 that participants did hold and manifest language attitudes towards both accents used. These language attitudes were similar to that discussed in section 6.2 (Lippi-Green, 2012; Preston, 2002b; Tamasi & Antieau, 2015; Wolfram & Schilling-Estes, 2016). GAE was perceived as "normal" or "educated" while participants commented on the defendant's poor grammar or rural roots when given the Southern American English guise. However, as only question 11 was affected by accent condition, it is once again possible to say that the presence of these language ideologies did not significantly affect participant judgments in any discernible way.

The responses to question 13 reveal how firmly these language ideologies are held. The majority of GAE groups did not understand what I meant by "accent" while the majority of Southern American English groups did not need to be given a definition. This suggests that the non-accent myth and standardisation ideals were so strongly held, that GAE did not even register with participants as being an "accent." Even after I explained this question to participants, many of them still appear to have misunderstood, as they often wrote about the defendant's emotions in response to question 7, rather than his pronunciation. In contrast to that, the Southern American English guise was noticeable enough, as something "other," that it was generally understood to be an accent, and therefore no question was needed and question 7 was not misunderstood. It appears that the standard accent falls below participants' level of consciousness, whereas Southern American English, and perhaps all non-standard accents, is above their level of consciousness.

One side note to this is that sociolinguists have begun to research how accents might carry indexical meaning with regards to emotions, just as much as they are indexed by other factors such as social class, gender, ethnicity, etc (Eckert, 2018). It is possible that GAE is indexed more closely with emotions than with specific stereotypes of those who speak it, which could explain why participants were more likely to note down emotional responses rather than pronunciation or stereotypical ones. However, because I directly explained to participants that I was

looking for a description of how they spoke, relating to their accent and pronunciation, I would still maintain that this is a misunderstanding on the part of the participants with regards to what I said I was looking for. That being said, it could be of interest to future researchers to also consider how emotions might be indexed and what impact that may have on how a testimony is viewed.

The deliberation transcripts did not reveal any recurrent patterns in regards to which accent guise mock juries were given. However, there were a few aspects I observed regarding how they worked as a jury. Many mock jurors mentioned the defendant's emotions and their own perception of his attitude in the trial, yet there was no evidence in any of the deliberations that this factored into their final decision on the verdict. This is not always true with juries. Research has noted the potential for jurors to empathise with either the defendant or victim, therefore adding an emotional aspect into their decision making process. For instance, U.S. juries who heard opening and closing statements urging them to empathise with the defendant's situation were found to be less likely to convict than juries who did not hear such statements (Haegerich and Bottoms, 2000). Equally, juries who are presented with highly graphic photographs and witness testimonies were likely to act more punitively than juries who did not receive such emotionally charged evidence (Salerno & Bottoms, 2009). There is an ongoing debate regarding this in legal and psychological fields, as it is agreed that some emotion is necessary in order to ensure a humane verdict, but too much emotion can cause juries to disregard evidence and ideas of justice and instead the verdict becomes purely reactionary (Linder, 1996; Salerno & Bottoms, 2009). While this was not an issue in my results, it is something for any future research to account for in order to make sure that the presence, or lack of, emotions do not adversely skew the findings.

Across all 18 groups, participants cited incorrect facts about the mock trial less than 1% of the overall time spent deliberation. Of the times that a mock juror did get something wrong, in each case another mock juror would correct them, and, in all but one instance, would give the accurate information. Additionally, even in the groups that deliberated for as long as 13 minutes, not one mock jury ever went off topic. Rather, they kept each other focused, working to find an answer they could all agree upon. Finally, even when juries unanimously agreed upon the verdict from the start, every mock jury discussed and justified their decisions to one another. While

some groups did this for longer than others, feeling the need to go into more detail, their very need to justify their decisions is evidence that each jury took the study seriously and honestly sought a just verdict. These are some examples of strengths in the jury system and have been found to be true in other research into juries and group dynamics (Bornstein & Greene, 2011; Devine et al., 2001; Myers and Lamm, 1976).

The deliberations equally revealed weaknesses in the jury system. In each deliberation that did start off unanimous, despite discussing the case and which verdict would be appropriate, the deliberations did not cause anyone to change their minds. Additionally, with each deliberation that did not begin unanimous, it was always the majority verdict that was decided upon at the end, with the minority either giving in to, or being swayed by, the majority. This finding is in line with previous research into juries. Bornstein and Greene (2011) noted that whenever pre-deliberation verdicts were taken from individual jurors, the actual verdict that was decided upon after deliberations was generally that of the majority. Devine et al. (2001) noted that this was not just a general trend, but actually occurred in 90% of the juries studied. If the majority gain confidence in their verdicts due to knowing others agree with them, they may be less open to having their minds changed through the deliberation process. This leaves the minority to either be outvoted or, in cases that require unanimity, forced to change their minds rather than be the cause of a hung jury. This would suggest that while working as a group does assist in getting key details right, staying on topic, and making sure their decision is well justified, the actual deliberation may not consistently make a significant difference to what the outcome will be.

In addition to the potential issues noted in chapter 5, it is also possible to see weaknesses in the methodology of this study through the deliberations. Many of the mock jurors missed the fact that five miles is a long way off, even after being directly told again in the text summary of the case. This had an impact on many juries' decisions, as they used the evidence of the defendant's speed as proof that he was negligent and therefore at fault for the incident. In a real court case, this evidence would have been explained as irrelevant by the defence attorney (if the evidence was even deemed admissible to court at all) therefore giving the jury a better understanding of this skewed evidence than my study gave them. In the future a more balanced mock trial should be put before participants, offering an

examination by both the defence attorney and the prosecution, thereby allowing multiple perspectives on the one topic. Additionally, many participants complained about the video being too short and not offering enough information. This would be another reason to offer a longer and more balanced video with examinations from both the defence attorney and prosecutor.

Another issue is in how many mock jurors misunderstood their role. It was a stated fact that the incident occurred. This was never in debate. What mock jurors needed to decide was whether the defendant was *at fault* for the incident. Instead, many jurors talked about how the defendant openly admitted to hitting the child and therefore he must be guilty. Again, this is something that likely would not have been an issue in a real trial. Jurors are given a great deal of instruction by the judge in how to properly weigh evidence and seek justice, certainly much more than I would have given them. While jurors do not always understand the instructions (Bornstein & Greene, 2011), it would be emphasized that jurors were to question whether the defendant was at fault due to proven negligence, not whether he did or did not hit the child. In the future it would be worth equally emphasizing what aspect it was that juries were to question and checking in with them before leaving them to deliberate to make sure they still understood what that question was. Additionally, a simple flow chart could be made, explaining the routes to a guilty verdict or a not guilty verdict, a helpful tool that is utilized in actual court cases (The Secret Barrister, 2018). An example of one such flow chart can be viewed in Figure 29. Taking steps such as these may help a mock jury better understand what "guilty beyond reasonable doubt" entails.

ROUTE TO VERDICT

AFFRAY
(section 3 of the Public Order Act 1986)

Question 1
Did BAS / RAA use, or threaten, violence towards another person?   (See **Note 1**, below)

Yes, I'm sure he did

No, or, he may not have done

Question 2(a)
Did he genuinely believe that it was necessary to use or threaten that violence so as to defend himself and/or another?   (**Note 2**)

No, I'm sure he didn't

Yes, or, he may have done

Question 2(b)
Was the force he used or threatened 'reasonable' in the circumstances as he perceived them to be?   (**Note 3**)

No, I'm sure it wasn't

Yes, or, it may have been

Question 3
Was the conduct of all of them, taken together, such as would cause a person of reasonable firmness present at the scene to fear for his personal safety?

Yes, I'm sure it was

No, or, it may not have

G U I L T Y

N O T   G U I L T Y

NOTE 1
You must consider each Defendant one-by-one; examining the evidence for and against each of them separately.

NOTE 2
It is not 'unlawful' to use reasonable force to defend yourself and/or another person. The first step is to consider whether the defendant genuinely may have believed he needed to use or threaten violence so as to defend himself or another. If he genuinely did believe that, or he may have done, then go to Question 2(b).
If, on the other hand, you are sure that he did not genuinely believe he needed to use or threaten violence at all, then the prosecution will have proved that he acted 'unlawfully' and you must go on to Question 3.

NOTE 3
Only the use of 'reasonable' force can be lawful; a person may not use more force than is reasonably necessary so as to defend himself or another against the threat which he genuinely anticipated. Obviously, a person who genuinely thinks he or another is about to be attacked may react on the spur of the moment. He cannot realistically be expected to weigh-up precisely how much force he needs to use to defend himself or that other person. If he has only done what he honestly and instinctively thought was necessary, then that would be strong evidence that it was 'reasonable'.
On the other hand if he 'goes over the top' - using force out of all proportion to what he genuinely anticipated might happen to him or another, then that would be 'unreasonable'.

Figure 29. Jury Flowchart (The Secret Barrister, 2018)

Another limitation of note is the large majority of female participants over male participants. With only 23% of the participants being male, the results cannot be considered a balanced or accurate representation of a jury demographic. Some of the mock juries held only female jurors, and of those mock juries which did have males, in most cases they were the minority in each group. As I was limited in my participant pool to those who willingly volunteered, this was an issue I was aware of as the study was being run but was equally unable to correct. While it is not unusual for the majority of participants to be female (as observed in the studies outlined in chapter 3), it must still be noted that this has the potential to skew results and lessens the ecological validity.

Past research into jury decision making has found that male jurors do respond differently to female jurors based on a number of factors, including the gender of the defendant and/or victim, the type of crime, or how reliable a witness is perceived to be (Corder & Whiteside, 1988; Deitz et al., 1982; McCoy & Gray, 2007; Pozzulo et al., 2009; Wuensch & Moore, 2004). For instance, male jurors are more sympathetic or more likely to act leniently toward male defendants than female defendants; and this is equally true of female jurors in regards to being more sympathetic toward female defendants as compared to male defendants (Wuensch et al., 2002). Therefore it is not only possible but highly likely that 77% of the participants being female had an impact on the results. Based on this research, it is possible that the mock juries were more likely to find the defendant guilty due to his gender and the incident involving a child. However, as this was not something that was explicitly mentioned by participants in any of their responses, I have no way of measuring this. It would explain why there was less variation in verdicts across all of the mock juries than in the Qualtrics datasets.

Many of the mock jury results align with social psychology theories on group dynamics and jury decision making. The results also support linguistic principles on standardization and how non-linguists often view the standard as a "neutral" variety. Furthermore, due to the two Qualtrics datasets, it is clear that the lack of gender diversity did not influence the outcomes of RQs I and II, as both of the Qualtrics datasets, which had a more equal split between genders, did not find evidence of accent discrimination either. Therefore, while it is highly likely that the majority female participants affected the results in some way, there is no evidence that this

193

affect was strong enough to invalidate the main findings and any conclusions based on them.

One final consideration is the effect my own accent could have had on this study. According to research by Hay, Drager, and Warren (2010) it is possible that participants will be influenced by any other accents they encounter, and this includes the speech of the person running the experiment. While this can become a confounding variable in perception studies, it is highly unlikely that my accent did have an effect. It cannot have affected the UK or the US Qualtrics data, as I had no verbal interaction with the participants. As for the mock juries, I had limited contact with them. I only spoke at the start to thank them for coming and to give instructions for the study. The mock jurors mainly interacted with one another throughout the study, as they spent the majority of time deliberating on a verdict together. Additionally, there is no evidence in either their recorded deliberations or their questionnaires that I adversely primed them towards a specific response due to my GAE accent. Therefore, although it is possible that an experimenter's accent can have an effect, there is no reason to believe that my accent affected or skewed the results in any way.

Overall 79 out of 82 mock jurors found the defendant guilty. While this high number of convictions may have been due to the aforementioned methodological issues, it does suggest that the evidence was deemed too strong by the mock jury participants for any extra-legal factors to make a difference. As there was more variation in verdicts in both the U.S. and U.K. Qualtrics datasets, it is more likely this was due to group polarization than any methodological limitations. Regardless, because participants did overwhelmingly find the evidence strong enough to convict, I was unable to test the liberation hypothesis, as it would only be observed when the evidence is weak.

6.5.3. Research Question III

When cross-comparing the results from the U.S. Qualtrics and the mock juries datasets only questions 1 and 4 revealed statistically significant differences, though questions 2 and 3 had near significant results. Additionally, there was some variation in the qualitative answers as well. Question 4 initially suggested that online

individuals were likely to recommend different punishments than groups who had deliberated together. However, after sorting the recommended punishments into either a *lenient* or *harsh* category, and accounting for the difference in overall participant numbers, no significance was found.

Question 1 regarding participants' verdicts was significant, as the Qualtrics dataset had more variation in verdicts than the mock juries dataset, which was almost entirely made up of guilty verdicts. This is evidence that working as a group does affect verdict outcome in a significantly different way from working as an individual. This could be due to the way group dynamics often caused the minority to be swayed, or forced, to join the majority decision. Furthermore, psychology research has long found that groups will often make *riskier* decisions than individuals (Bougheas, Nieboer & Sefton, 2013; Stoner, 1968). This could be due to how working as a group tends to increase individual's confidence in a decision, allowing riskier decisions to seem more achievable or justifiable (Forsyth, 2017). It may be that a guilty verdict was perceived as riskier, due to the serious consequences for the defendant if the participants decided he was at fault. This risk is amplified by the knowledge that the defendant will still face these consequences even if they were wrong and he was not actually at fault. This then would add further nuance to why individual participants were more likely to vary in their verdicts, whereas group participants predominantly chose to render a guilty verdict.

The trends in questions 2 and 3 may have also been caused by group dynamics, in particular group polarization. Mock jury participants tended to be more confident in their decisions and this is reflected in the ratings they gave in Figure 28, where they were more likely to give the defendant lower than neutral ratings. These trends are unsurprising, as group polarization theory suggests that groups will go more to the extremes, whereas individuals are more likely to stay more central or neutral (Bornstein & Greene, 2011; Forsyth, 2017; Myers & Lamm, 1976). Overall, this is a clear indication that group dynamics can and do affect the outcome of trials, and therefore any research into trials and jury decision making ought to include the use of mock juries rather than individuals.

With regards to the qualitative analysis, the mock jury participants were much more uniform in their reasons for changing their answers after reading the text

summary than the Qualtrics participants. As noted in the results, this uniformity could be evidence of how working as a group tended to align their opinions. In both datasets language attitudes were obvious in how participants spoke about the two accents; however, there was little to no evidence that these language attitudes were acted upon. Again, this suggests that the presence of language attitudes has little effect on whether accent discrimination occurs. Rather it may be due to the *type* of language attitude present, regarding what it indexes and whether that variety is associated with a stigmatised community.

While there were not statistically significant differences for every analysis question, between the results listed here for RQ III and the findings from the deliberation transcripts, there is clear evidence to suggest that working as a group can lead to different results than working as an individual. Due to this, where possible studies ought to employ mock juries in order to achieve greater ecological validity, as it is questionable whether individual responses really would reflect how a jury would respond. Therefore, the results for the previous studies listed in chapter 3 that did not use groups, or who used groups of such variable sizes, ought to be handled cautiously and with scepticism.

6.6. Conclusion

This chapter examined the perceptions of GAE and Southern American English and then sought to discover whether those perceptions would be acted upon, causing discrimination. After analysing the data, except for question 11 in the mock juries dataset, it was clear that accent had no overall effect, suggesting that the perceptions found in previous research are simply that: perceptions. There is no evidence in this study that these perceptions regularly play an active part in people's decision making and judgments within the United States. This answers both RQ I and II: while the defendant was perceived differently based on his accent, it in no way effected the verdict. This concurs with the results found in chapter 5, further strengthening the conclusion that accent discrimination is not an automatic occurrence when language attitudes are present. Moreover, many of the findings align with previous linguistic and social psychology literature.

RQ III was also considered. Significant differences between working as an individual versus working in a group were found, shown in the deliberation transcriptions and in the results listed in 6.4.3 and discussed in 6.5.3. This aligns with psychology literature on the effect of group dynamics and group polarisation theory. Juries keep one another on task, correct each other's mistakes, and encourage greater confidence and unanimity in any verdict given. Therefore if research is undertaken to better understand trials and verdicts, mock jury focus groups and not individual mock jurors ought to be used.

Chapter 7. Research Question IV: Comparison of U.K. and U.S. Studies

This chapter seeks to answer RQ IV: "Are the results of the first two research questions similar enough across western nations, such as the United Kingdom and the United States, as to be valid outside of the country the study was conducted in?" In order to address this, the U.K. Qualtrics dataset results (outlined in chapter 5.4) and the U.S. Qualtrics dataset results (outlined in chapter 6.4.1.) were compared. Chapter 6 found significant differences between working as a group versus working as an individual, and therefore suggested that individual results when researching juries and trials may be invalidated. It was recommended that groups be used to achieve greater ecological validity. However, despite this it is still worthwhile to examine the U.K. and U.S. individual Qualtrics responses. If there are any significant differences between the two datasets, it will serve as a reminder to exercise caution before taking the results of a study conducted in another country, or even region, and assuming the same would be true elsewhere.

This chapter compares the results of the U.K. participants to the results of the U.S. participants. However, this is not a direct comparison, as that is not possible, due to the differences in culture. While I sought dialects that had similar language attitudes associated with them, as discussed in chapter 4, I have no way of knowing (and therefore no way of showing) YE is to SSBE what Southern American English is to GAE for each member of my respective national group of participants. In fact, it is more probable that they are not direct equivalent. YE is one regional dialect among a multitude of them within the relatively small geographical span of the United Kingdom. SSBE is known to be the standard variety and is also clearly associated with a specific region (acknowledged in the very name of that variety)

(Hughes, Trudgill, & Watt, 2012; Mugglestone, 2003). In comparison, Southern American English covers a larger geographical area and is a marked regional variety in a nation that is significantly larger than the U.K. but arguably has fewer distinct linguistic varieties, thus making those that are marked stand out even more. GAE is not associated with any one region, and because of its lack of a definition, it is easier for more speakers to claim to use it (Allbritten, 2011; Clopper, Pisoni, & de Jong, 2005; Kretzschmar, 2008). These U.K. and U.S. varieties may be similarly indexed in some ways, but they are by no means identical.

However, I believe this comparison is still valid. The goal of this chapter was never to compare language varieties with identical indexical meanings, but rather to learn to what degree researchers can, and should, rely on studies conducted in foreign countries. Dixon et al. (2002) cited Seggie (1983) in their article. They replicated Seggie's methodology and interpreted their results in the light of what he had found. Yet Dixon et al. was conducted in Britain 20 years after Seggie conducted his study in Australia. Frumkin (2007) discussed Sobral Fernández and Prieto Ederra's (1994) work and also discussed her findings in light of what they had found, despite the fact that Sobral Fernández and Prieto Ederra not only conducted their study in a different nation 13 years earlier, but also had a different language variety entirely. In each of these cases the justice system will also have differed and thus affected the results in a way that may not be replicable or relevant in another country. While my comparison is not an exact like with like match, it will definitely share more commonalities than Dixon et al (2002) and Seggie (1983) or Frumkin (2007) and Sobral Fernández and Prieto Ederra (1994) simply because it was run at the same point in time, with the exact same methodology, by the same researcher.

To my knowledge no one has previously run the same study researching juries and extra-legal factors in multiple countries and then compared those results. However, other studies have compared the jury systems between the U.S. and western European countries in regards to how well their legal systems work. The U.S. and U.K. have different procedures for selecting their juries, how many jurors must agree in order for a verdict to be given, and what level of involvement juries have in trial proceedings and sentencing (Leib, 2008; Martin, Kaplan & Alamo, 2003). A potentially relevant finding to this research is that U.S. courts generally

give harsher sentences than U.K. courts (Tonry, 2001). Due to this, it is likely that U.S. participants will recommend harsher punishments than their U.K. counterparts.

Although mock juries should be used when researching trial outcomes involving juries, this comparison is viable because I have not used mock juries. This removes any potentially confounding variables regarding differences in the jury system between the two countries and instead focuses on whether participants are likely to act in a similar manner when the environment is near identical. I did not expect to find distinct differences in every category between the two countries. However, if only the verdict and recommended punishment significantly differed, this would imply that trial outcomes as a whole significantly differ between these countries.

7.1. Analysis

As with chapters 5 and 6, when analysing this data I used a mixed methods design in order to investigate the effects of the specific dataset on characteristics of the defendant in comparison to the prosecutor, verdict of the mock trial, and any recommended punishments. There were 54 participants in the U.K. study and 56 participants in the U.S. Qualtrics study. This near even number of participants makes the comparison more reliable. Because no significant effects regarding accent guise were found in either dataset, they were analysed as a whole (U.K. Qualtrics' results compared against U.S. Qualtrics' results) rather than analysing them according to accent guise (SSBE v. GAE and YE v. Southern American English).

RQs I and II were analysed using the following list of questions. These were the same for each dataset in chapters 5 and 6, with one exception. Because I changed the way the question was formatted in the U.S. version (using a Likert scale from 1-6) from how it was formatted in the U.K. version (using three categories "unsure," "somewhat," and "certain"), I was unable to cross compare how confident participants were in their verdicts.

Did the particular dataset (U.K. Qualtrics or U.S. Qualtrics) lead to a difference in:

Q1. Participants' verdicts?

Q2. Whether participants were more likely to rate the defendant lower in characteristics?

Q3. What level of punishment was suggested?

Q4. How long participants took to finish the questionnaire?

Q5. The quality rating participants gave the videos?

Q6a. Whether participants changed their verdict after reading the text summary.

Q6b. If so, why?

Q7. Whether participants were more likely to offer negative opinions on the defendant?

Q8. How participants spoke about the accent of the defendant?

Questions 1-6a were analysed quantitatively while questions 6b-8 were analysed qualitatively. What now follows are the results of the cross-comparison analysis.

7.2. Results

Question 1: Did participants' verdicts differ significantly depending on the dataset?

The U.K. had an even split between verdicts while the majority of U.S. participants rendered a *guilty* verdict. This can be viewed in Table 62.

	verdict		
	Guilty	Not Guilty	*total*
U.K.	27	27	54
U.S.	38	18	56
total	65	45	110

Table 61. RQ IV Verdict

A Chi-Square test for association was conducted in order to determine if the dataset type had an effect on individual verdicts. The results of the test narrowly failed to reach significance at the 5% level (p = .057), shown in Table 63. This suggests a

potential trend. Additional data may have shown that the U.S. is more likely to render a guilty verdict than the U.K.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**3.626ᵃ**	**1**	**.057**
Continuity Correction	2.925	1	.087
Likelihood Ratio	3.646	1	.056
N of Valid Cases	110		

Table 62. RQ IV Verdict – Chi-square test

Question 2: Were participants more likely to rate the defendant lower in characteristics?

Overall, the prosecutor was perceived much more favourably by the U.S. participants than by U.K. participants, as U.K. participants gave generally neutral ratings (3-4) while the U.S. participants gave fairly positive ratings (4-5). The defendant, however, was perceived slightly more favourably by the U.K. participants, though all the ratings were in the neutral 3-4 range. This variation can be viewed in Figure 30, with further details in Appendix V.
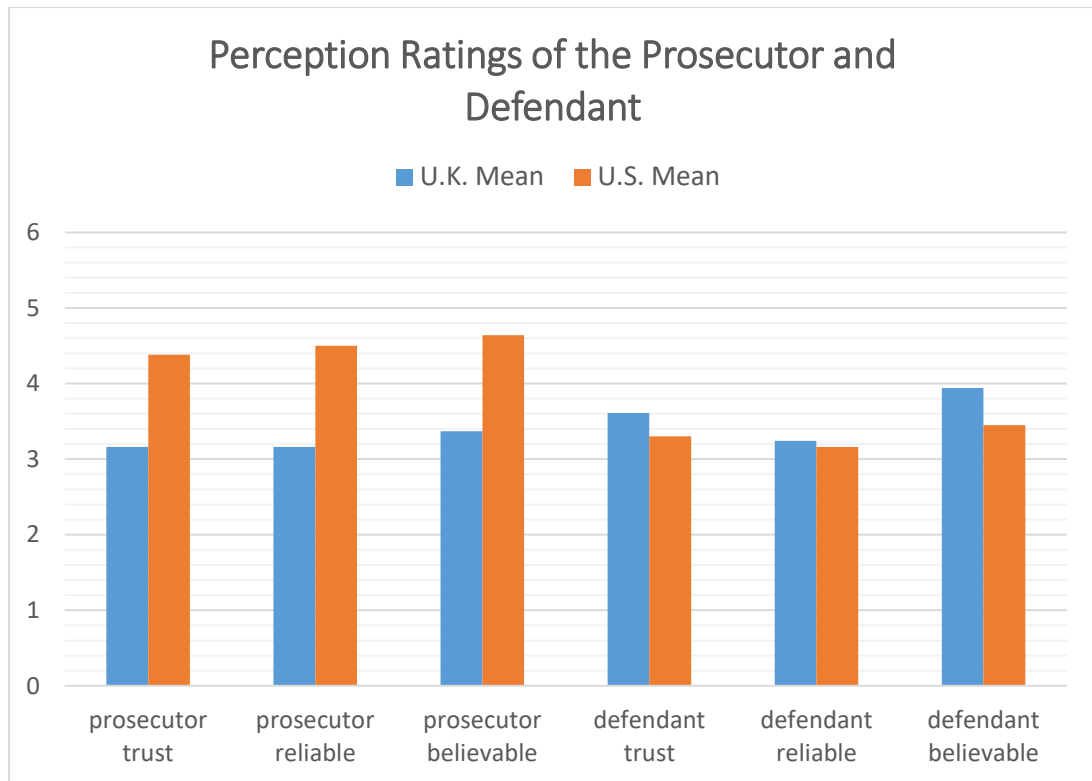
Figure 30. RQ IV: Question 2 - Perception Ratings

I conducted Independent t tests in order to determine if there were any significant differences in perceived characteristics of defendant and prosecutor between U.S and U.K groups. There was a significant result for all of the prosecutor's characteristics (p < .001), but none of the defendant's characteristics, though there was a near significant result at the 5% level for the *believability* of the defendant (p = .062), shown below in Table 64. As the other two ratings for the defendant were not close to significance, I have no reason to believe that the near significant finding is meaningful. The difference in prosecutor ratings could suggest that prosecutors in general are viewed differently in these two societies.

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
prosecutor - trust	5.792	103	**.000**	1.212	.209
prosecutor - reliable	6.592	103	**.000**	1.337	.203
prosecutor - believable	5.799	103	**.000**	1.276	.220
defendant - trust	-1.202	103	**.232**	-.309	.257
defendant - reliable	-.298	103	**.766**	-.084	.282
defendant - believable	-1.887	103	**.062**	-.533	.283

Table 63. RQ IV Perception Ratings – Independent T Test

Question 3: What level of punishment was suggested?

As laid out in Table 65, only a few U.K. participants recommended any type of prison sentence. Instead the majority of them either suggested a lesser penalty or did not recommend a punishment at all. U.S. participants were more evenly split between a harsh or lenient recommendation, in that roughly half recommended something "other" (e.g. community service, etc.) while the other half believed the defendant should receive a prison sentence.

	punishment				
	None	Other	Prison	Prison+	*total*
U.K.	5	17	4	1	27
U.S.	0	20	18	0	38
total	5	37	22	1	65

Table 64. RQ IV Recommended Punishment

I performed a Fisher's Exact test in order to determine if recommended punishment was associated with dataset type. The test returned significance, at p < 0.05, shown in Table 66. This result confirms that participants were likely to give a different recommended punishment depending on whether they were from the U.K. or from the U.S. Specifically, that U.S. participants were more likely to recommend a stronger punishment than U.K. participants. This confirms the hypothesis made regarding this and is line with earlier research.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	13.683[a]	3	.003	.001
Likelihood Ratio	16.327	3	.001	.001
Fisher's Exact Test	**13.454**			**.001**
N of Valid Cases	65			

Table 65. RQ IV Recommended Punishment – Fisher's Exact

Question 4: Did the duration of how long participants took to finish the questionnaire differ significantly depending on the dataset?

Descriptive statistics were assessed to determine the mean duration of time. As before, the length of time was measured in seconds. There was an average difference of nearly 300 seconds between the datasets, equating 4.5 minutes. U.K. participants on average spent nearly 20 minutes (1152.45 seconds) on the questionnaire while U.S. participants spent an average of 14.5 minutes (881.13) completing the questionnaire. I used an independent t test to determine if the length of time to complete the survey differed based on dataset type. As seen below in Table 67, this was statistically significant ($p < 0.05$). Therefore, on average U.K. participants took significantly longer than U.S. participants when completing the questionnaire.

		t	df	**Sig. (2-tailed)**	Mean Diff.	Std. Error Diff.
Duration	Equal variances assumed	-2.041	107	**.044**	-271.328	132.908
	Equal variances not assumed	-2.018	87.507	**.047**	-271.328	134.474

Table 66. RQ IV Duration – Independent T Test

Question 5: Did the quality rating participants gave the videos differ significantly depending on dataset?

The two datasets had almost exactly the same ratings in both quality categories, shown in Table 67.

	quality		
	High	Low	*total*
U.K.	40	14	54
U.S.	41	15	56
total	81	29	110

Table 67. RQ IV Video Quality Ratings

A Chi-square test of association was performed in order to determine if the video quality ratings differed based on the dataset type. Unsurprisingly, these differences were not significant ($p = .919$). It is clear that participants reacted in a similar manner to the quality ratings regardless of which dataset they belonged to.

What now follows are qualitative analyses of the responses participants gave in regards to questions 6-8.

Question 6: Did participants change their verdict after reading the text summary? If so, what reason did they give?

Within the U.K. dataset nine participants changed their verdict from *guilty* to *not guilty*. Six said it was due to the speed limit evidence being five miles prior to the incident, making it irrelevant. Two participants said it was due to realizing the children had more culpability than they had previously noted. The final participant did not specify what change their mind. An additional participant changed their mind from *not guilty* to *guilty* due to the speed limit evidence. This was analysed in Tables 13 and 14 in section 5.4. The five U.S. dataset participants who changed their verdict all did so for separate reasons. Additionally, two of the participants changed from *not guilty* to *guilty*, while the other three changed their verdicts from *guilty* to *not guilty*. There was no true pattern in the U.S. dataset here. This was outlined in Table 33 in section 6.4.1. Due to that, it was not possible to find any patterns between the

U.S. and U.K. datasets. This does suggest, however, that participants missed pertinent information in the video and having a text summary was helpful.

A chi square test was run in order to test whether participants were more likely to change their verdicts depending on which nation they were from. However, the test was not significant ($p = .170$). Therefore, the text summary was not found to affect one group of participants more than another, simply because of their nationality.

Question 7: Were participants more likely to offer negative opinions on the defendant in the non-standard guise?

After analysing both datasets Tables 15-17 and 34-36 in chapters 5 and 6 it was found that accent did not appear to make a significant difference in whether participants gave them positive or negative characteristics. Even the specific category types that answers were sorted into (e.g. "actions," "emotions," etc.) were similar between the two datasets.

Question 8: How did participants speak about the accent of the defendant?

Again, the responses between the two datasets, as outlined in chapters 5 and 6, were fairly similar. Both datasets were more likely to mark the standard guise as more "normal" or "neutral" than the non-standard guise. Participants often used emotional terms, such as "he sounded nervous" or "he seemed insincere." U.S. participants were more likely to give a specific location they believed the defendant, and his accent, were from than the U.K. participants. In both datasets it was clear that participants had brought language attitudes with them to this study. However, there was no evidence to suggest that participants were significantly likely to act upon these language attitudes and actively discriminate against the defendant due to his accent.

7.3. Discussion

This study sought to answer RQ IV regarding whether datasets from the U.K. and the U.S. would be similar or significantly different. The analysis addressed this question through the specific questions discussed in section 7.1.

Questions 2, 3, and 4 yielded significant differences, while question 1 neared statistical significance. Question 4 found that U.K. participants were significantly more likely to spend longer completing the questionnaire than U.S. participants. This could be evidence that U.K. participants took the study more seriously and therefore gave more consideration to their answers than the U.S. participants. This would not be surprising, as many of the U.S. participants were recruited through MTurk and paid a small fee to complete the questionnaire. Because the MTurk participants were paid a single amount for the questionnaire as a whole, rather than how long it took them to complete the questionnaire, it follows that they would be more motivated than the U.K. participants to finish the questionnaire quickly, in order that they might move on to their next paying study. Therefore, it is possible that this significant result is due more to how I recruited my participants rather than which country they were from.

Question 2 found that U.S. participants were significantly more likely to give the prosecutor higher ratings than the U.K participants for all of his perceived characteristics. The U.K. participants generally gave neutral ratings (ranging 3-4) and this could be due to them perceiving the prosecutor as simply doing his job. This would result in the U.K. participants having no strong feelings, and explaining their overall neutral ratings. Within the U.S., research suggests that people are likely to view lawyers fairly positively, unless they have had personal experience with them or are university-educated with an above average salary (Hengstler, 1993). While comparative statistics for the U.K. were not found, it is possible that this explains the U.S. participants' more favourable ratings. I have no reason to believe any of them make a significant amount of money based on their reported professions or have had negative personal experiences with lawyers based on their responses. Equally, because there is a great deal of schooling required to be a prosecutor, more so in the U.S. than in the U.K., it is also possible that the U.S. participants rated the prosecutor more favourably due to his perceived level of education.

As hypothesised, question 3 results suggest that U.S. participants were more likely to recommend a harsh punishment than U.K. participants. Because the U.S. still has capital punishment in many states, while the U.K. had completely abolished it by the 1970s (Zimring, 2003), the U.S. is often stereotyped as dealing out harsher punishments than the U.K. (Dieter, 2003; Hongju Koh & Pickering, 2003). Research into conviction rates in the United States and other western nations, including the United Kingdom, would suggest that this perception is accurate, as the U.S. convicts more defendants and gives harsher sentences than its European counterparts (Tonry, 2001). Therefore, the significant finding of question 3 aligns with previous research and public perception, in that Americans are more likely to recommend harsher punishments than Europeans, including those from the U.K.

While there was a significant difference in recommended punishment, no significant differences were found in regards to the actual verdict given, though it neared significant levels. This may offer (weak) evidence that trial outcomes between these two nations are likely to differ. Even if this is not with regards to the verdict, the future of a convicted defendant does look significantly different depending on which country they are in. This is confirmed by previous research (Tonry, 2001). As this is regardless of any language variable, it can be concluded that any research into trials and legal procedures ought to report and rely on information that is relevant to the country they are studying. Findings from another country should not be used as evidence for an argument without a fair deal of caution and scepticism.

Once again similar language attitudes were noted in both datasets, but as mentioned in the previous two chapters, there is no evidence to suggest that these language attitudes played any significant part in how participants responded. As before, the importance of this finding cannot be overstated. There is now evidence that accent discrimination is not an automatic occurrence in two western nations, suggesting that when accent discrimination is triggered, there must be a number of factors at play, only one of which will be language attitudes.

The limitations in the methodology brought to light in chapters 5 and 6 may equally apply to the results in this chapter. However, as both of these datasets had variation in the verdicts given, it is clear that overall participants did not feel there

was only one "correct" and "clear" verdict. Therefore, any limitations in this research were not severe enough to adversely affect participant responses.

7.4. Conclusion

This chapter examined RQ IV by cross-comparing the U.K. Qualtrics dataset and the U.S. Qualtrics dataset. After analysing the data, there were three significant results among eight analysis questions. The findings for analysis questions 2 and 3 can be understood in light of psychology literature and general public perceptions. The significant result for question 4 was more likely to be due to how I recruited participants in the two nations, rather than any cultural differences. Language attitudes were observed in both datasets and yet neither had evidence of accent discrimination. As this result has now been found in two western countries with their own language attitudes and SLI, it strengthens the conclusion that accent discrimination is not an automatic occurrence when language attitudes are present. Overall, it can be said that while most of the results from one nation could be applied across the Atlantic to the other nation, caution must still be exercised when doing so as there may be cultural differences as shown by the significant result of analysis questions 2 and 3, and the trend apparent in question 1. Furthermore, with regards to trial and jury research, because the outcome may significantly differ depending on the country, these countries ought to be studied separately.

Chapter 8. Conclusion

There is evidence in multiple contexts that accent discrimination has affected numerous people's lives, in both minor and major ways. Whether it is costing a person more in time and resources in order to find housing, convicting a defendant based on unsubstantiated evidence, or even just regular emails deriding an aspect of the speaker's identity, accent discrimination can have serious consequences (BBC News, 2017; Massey & Lundy, 2001; Wiehl, 2002). Therefore, there is a clear real-world impact that research into this topic can offer.

The main question this dissertation sought to address was whether the presence of language attitudes lead directly to accent discrimination. What was discovered through the literature review chapters and the results of the studies conducted shows that the answer is, as expected, more nuanced than a straight *yes* or *no*. Yes, there is evidence language that attitudes lead to accent discrimination. Witness Rachel Jeantel was ignored and ridiculed due to her language use (Rickford & King, 2016; Sullivan, 2016; Thompson, 2015). Charles Clifford and Larry Hyppolite were convicted due to general misunderstandings of linguistic principles (Rodman, 2002; Wiehl, 2002). MP Angela Rayner has received numerous emails recommending she "improve" her speech and stop speaking in her non-standard dialect (BBC News, 2017; Gill, 2016; Phillips, 2017). However, this discrimination does not occur across the board, and therefore must be considered on a case by case basis rather than offer a broad generalization. There are a number of areas where further research can be conducted to continue the investigation into the possible variables that trigger the occurrence of accent discrimination

8.1. Summary

Through the news articles and social media posts examined in chapter 2, it is clear that people do experience discrimination due to prejudices held towards language varieties. This is further backed up by the research done looking at housing, employment, and education arenas. Based on the demographics of those who self-reported facing discrimination, as well as evidence within academic studies, it appears women may experience this to a greater degree than men.

This discrimination does not only occur when people make snap judgments, but also in the courtroom, where decisions are supposed to be carefully weighed and thought through. This is seen across the court cases described in chapter 3. While it certainly cannot be claimed that accent discrimination affects every trial, there are enough affected to suggest that this may be a systematic problem that is little understood, or even noticed, by the courts themselves. The first amendment of the U.S. constitution and Title VII Discrimination Act both provide potential solutions when fighting against accent discrimination. The first amendment gives Americans freedom of speech. Commonly interpreted as giving people the right to say *what* they wish, this could be interpreted as equally giving people the right to say what they want *how* they want to say it. In other words, it could set a precedent that every language variety is acceptable, even if it is not the standard. Additionally, if the Title VII act redefined *accent* to be understood as a linguist would use it (as defined in chapter 1), this could add further legal protection to people who face discrimination due to their speech.

There has not been a great deal of research conducted over the last 45 years studying accent discrimination within courtrooms, and what has been done is scattered across multiple Western nations. Their results (laid out in chapter 3) would also suggest that accent discrimination does occur within courts. However, they too found more nuanced results than simply automatic, widespread discrimination. Some found that accents and the stereotypes associated with them could be linked to type of crime (blue collar v. white collar) (Seggie, 1983); a stigmatized ethnicity increased the likelihood of discrimination (Dixon, Mahoney & Cocks, 2002); some researchers found that greater familiarity with an accent increased its positive ratings (Frumkin, 2007; Sobral Fernández & Prieto Ederra, 1994) while other researchers found that familiarity had no effect and rather it was perceptions of standardness that

had the greater effect (Dixon & Mahoney, 2004). Some of the differences in these findings could be due to the methodology used and others could simply be a result of language attitudes differing by culture and period of time.

Through these studies, three main research gaps were identified: individual participants versus groups of participants, gender of speaker guise, and audio versus visual stimuli. Most of the researchers did not consider how jurors working as groups could differ significantly from individual jurors offering their own opinion. Although collecting responses from individuals is less time consuming and requires fewer resources and participants, the lack of a group dynamic severely limits the ecological validity of any results when studying how a *jury* would judge in a trial. This is backed up by a multitude of psychology literature on group dynamics, including group polarization and risk taking (Bornstein & Greene, 2011; Bougheas, Nieboer & Sefton, 2013; Forsyth, 2017; Myers & Lamm, 1976; Stoner, 1968). Only one earlier linguistic study used a female guise, every other study used only male actors. As chapter 2 noted, it is possible that women receive more overt discrimination than men, and this appears to be largely unstudied. Finally, audio recordings were more commonly used than video, again lessening the potential for ecological validity as most juries would be able to see who was testifying and therefore have more variables to use in their consideration of a speaker than just voice.

Chapter 4 considered these gaps and discussed the two pilot studies run in order to test the second and third gap. The first pilot study tested whether the speaker would be perceived differently due to their gender and found no evidence to suggest this was occurring. Therefore I chose to use only a male guise so that my results would be more comparable to the results of previous research. The second pilot study tested the use of an audio versus a video stimulus and again found no significant differences in ratings between the two. However, I still chose to use video stimuli as I felt this improved the ecological validity of my studies. I also decided to collect data from both individual participants and groups of mock juries for my full studies in order to learn whether the group dynamics truly do have a significant effect. Additionally I chose to run my study in both the United Kingdom and United States to test how differing cultures may affect language attitudes and study outcomes. The four research questions were as follows:

Research Question I: Will a defendant be perceived differently by a jury based on whether he speaks in a standard or regional accent?

Research Question II: If a defendant is perceived differently, will this go as far as affecting the verdict?

Research Question III: Does collecting individual responses lead to reliable results on how a jury would function or does the group condition make a large enough difference to invalidate research results that used only individual participation?

Research Question IV: Are the results to the first two research questions similar enough across western nations, such as the United Kingdom and the United States, as to be valid outside of the country the study was conducted in?

I hypothesised that the defendant would be perceived differently by a jury based on the accent guise. However, I was open-minded as to whether that would go far enough to affect the verdict. I predicted that results from the mock juries would differ significant from individual participant responses and I had no general predictions about what would be found with respect to RQ IV.

Across all the studies listed in chapters 5, 6, and 7 language attitudes were clearly present. However, there was no evidence that this had an effect on either individual jurors' or mock juries' verdicts and other responses. Therefore, while the answer to RQ I was *yes*, the defendant was perceived differently as evidenced by the language attitudes present, this did not affect the verdict, or other areas of judgment, leaving the answer to RQ II a resounding *no*. This is an important finding, as it shows that language attitudes do not automatically lead to discrimination. While language attitudes have the *potential* to affect behaviour and cause discrimination, as seen in chapters 2 and 3, this is by no means a constant and regular rule to be generalised.

The results for RQ III can be seen in chapter 6. The answer to this question is also more nuanced than a straight *yes* or *no*. It was noted that there were few statistically significant quantitative differences in the responses when acting as an individual juror versus working in a group as a mock jury, although trends were

observed. However, there was evidence that the group dynamics had an effect on how confident participants were in their verdicts and that it caused them to be much more uniform in their responses across the board than the individual jurors were. Most notably, the mock jury groups differed significantly in their verdicts from the individual jurors. This suggests that the use of individual jurors in any type of trial research significantly inhibits ecological validity, as it is clear that group dynamics do have an effect on trial outcomes. Moreover, working as a group required participants to discuss and justify their decisions, something individuals did not have to do. Thus, whether it was a right or wrong decision, it is highly probably that there was more thought put into the decision the mock juries made than the decisions the individuals made, further raising the ecological validity of the mock jury dataset. For these reasons, it is strongly recommended that any future research into trials and jury decision making use groups of participants and consult psychology literature on group dynamics. There is a great deal that can be better understood by working across disciplines and using the expertise developed in other fields to better inform linguistic research.

Finally, RQ IV was analysed and discussed in chapter 7. A few significant results and trends were observed, most notably that U.S. participants were significantly more likely to recommend a harsh punishment for the defendant than the U.K. participants. Furthermore, there was a trend of U.S. participants having more variation in their responses (e.g. U.K. participants gave neutral ratings to both the prosecutor and defendant, U.S. participants had significant differences; U.K. participants were evenly split between verdicts, U.S. participants had a greater amount of not guilty verdicts). However, the rest of the analysis questions did not find any significant results or even trends. This suggests that, while sentencing outcomes may significantly differ depending on whether one is tried in the U.S. or the U.K., neither country is likely to automatically discriminate against a speaker due to the use of a non-standard accent. Thus, the answer to RQ IV with regards to psychology or legal research into trials and juries is *no*, the results are not similar enough and need to be considered on a country by country basis. However, with regards to language attitudes, the answer to RQ IV is *yes*, the results are the same, as neither found evidence of accent discrimination. With further research, it may be

found that the various factors which trigger accent discrimination differ between the two countries due to differences in the cultures.

Overall, this dissertation found that while accent discrimination occurs, it is not automatically triggered and does not universally affect non-standard accent speakers. It is possible that not all language attitudes have a strong enough effect, or a consistent effect across groups of people, to be a cause of concern. Rather, it could be language attitudes associated with stigmatised communities that activate accent discrimination. Therefore, further research is needed in order to discover when and where accent discrimination is most likely to arise, and the impact it might have in those contexts. Discovering this would allow for more relevant recommendations to be put to the courts, and elsewhere, in order to reduce the possibility of accent discrimination and to give speakers a recourse to pursue when it does happen.

8.2. Further Possibilities

All three datasets analysed in chapters 5, 6, and 7 revealed that language attitudes had been brought to the study by the participants. However, in not one dataset did those language attitudes cause any detectable accent discrimination. It is possible that language is the vehicle that allows discrimination, but it is less likely to be discriminated against for its own sake. In the future it would be worth questioning how much more likely accent is to be used as something to discriminate against when another protected trait, like ethnicity, is an active variable. Research findings suggest that, in the U.S., one of the largest extra-legal influencers in a trial is ethnicity or race (Esqueda, Espinoza & Culhane, 2008). Therefore, it is likely that this will equally affect other extra-legal factors, such as language. Even if a linguistic variety such as AAE is only discriminated against due to the ethnicity of those who speak it, that is still discrimination and is a factor that ought to be considered in more depth.

Equally, while language attitudes do not automatically trigger biased behaviour, it could be that it is more likely to occur in the extremes, such as regarding a community of speakers who are heavily stigmatized or who carry a large amount of prestige. The non-standard varieties I used in this study, Yorkshire English and Southern American English, had both positive and negative stereotypes,

and it could be that the negative stereotypes either were not strong enough or did not apply in a court setting. It may be that accent discrimination is more likely to be found where the stereotypes are mainly negative, or preferential treatment noted where the stereotypes are mainly positive. Again, this may reveal more about what people think of the community who speak that variety, rather than the variety itself. However, it adds further nuance to what is currently understood about accent discrimination.

Research into language attitudes, while not an automatic predictor of human behaviour, adds to the information available in order to allow for a better educated guess regarding how someone may act. For instance, a person who thinks speakers of Southern American English sound stupid may be more likely to act discriminatorily than a person who thinks speakers of Southern American English sound friendly. Future research could look into whether people are more likely to act according to the *main* characteristic they ascribe a language variety or speech community.

Researching language attitudes with regards to *how* a speaker is saying something is only half of the equation; *what* the speaker is saying must also be considered, giving context to the environment. Montgomery and Moore (2018) found that evaluating how people perceive linguistic features within the framework of what was said lead to varying perceptions of the speaker's identity. It is likely that this effect would be observed in a court context as well, as juries need to evaluate the reliability of a testimony (*what* is said) and it is well-documented that they use extra-legal factors to assist with this (Devine et al., 2001; Esqueda, Espinoza & Culhane, 2008). Therefore, in future research it would be worthwhile to also include conversation analysis techniques in order to study the way in which *what* was said may intersect with *how* it was said (i.e. what accent the speaker has) with regards to when accent discrimination may be triggered.

Another consideration for future research is testing whether participants' level of awareness of any language condition has an effect. It is possible that accent discrimination is more likely to be uncovered in situations where language is foregrounded, but not as readily when language is not perceived to the focus. This could be the case with many of the previous studies on this topic, as there was no

217

evidence they masked their interest in the accent variable. However, in the same vein, many of the trials discussed in chapter 3 also had language foregrounded in such a way that the judges or juries themselves could have been primed. This can be seen in every trial listed, where either a lawyer, judge, or the defendant directly mentioned language or linguistic principles and used it as a part of their argument. Therefore it would be worth studying whether foregrounding the language variable in this way is more likely to trigger accent discrimination than a situation where the language variable is present but never mentioned.

The above are all research questions brought to light by the results of this dissertation. Additionally, there are a few issues left by the methodology of my studies that could be filled by future research. Due to time and resource constraints, I prioritized other variables such as individual responses versus group responses over a gender variable. I did not have the capacity to properly examine gender. It would not have been enough to simply use a female defendant, as that would not have given any information about how female testimonies may be treated differently from male testimonies. Rather, in the future it would be worthwhile to create a full video stimuli with four conditions: standard accent v. non-standard accent and male v. female speaker. This could answer the question raised in chapter 2, regarding whether it is true that women are more likely to be discriminated against than men. If there are no differences in discrimination strength, it could be that women are more aware of the discrimination they face, and therefore more likely to raise it as an issue.

Another area to be looked into in the future is the effect of the lawyer's accent, both for the defence and the prosecution. In a trial, it is the lawyers who are heard from the most, and therefore it is possible that their accent also has an effect on the outcome of a trial. Again, it would be worthwhile to not only add accent as a variable, but also gender, and question whether it makes a difference if the lawyer is male or female.

Accent discrimination is not an automatic outcome when language attitudes are present. Yet there are clear cases of accent discrimination in the news, in research, and in trial transcripts. There is still a great deal to be learned about accent discrimination, and the subtleties regarding how it works and when it will be

triggered. This dissertation is only one piece of the puzzle, but it adds further knowledge to the ongoing conversation. With time, as more is discovered, it is possible that enough could be learned to accurately predict when and where this discrimination will occur, potentially impacting future legal policies.

Appendices

Appendix I: Gender Effect Pilot Study

<u>"Researcher" Instructions for Participants Script</u>

> *"Thank you for participating in this study. The task you are being given deals with short-term memory and the amount of information it is able to store. There are two parts to this study. In the initial part you will be given a list of 16 words. You may study them for as long as you like. When you are ready, proceed to the following page and write down as many words as you are able.*

> *"In the second part, instead of reading a set of words yourself, I will read them out to you. Once I have read through the list, proceed to the following page and again write down as many as you are able to remember."*

<u>Further Results</u>

I condensed the four accent conditions into two, GAE and Southern American English, and tested these as the independent variable against the neutral/negative ratings. The baseline intercept was automatically set to GAE. This test returned significant as Southern American English and GAE differed from one another by $p < 0.001$, as illustrated in Figure 31 and Table 68.
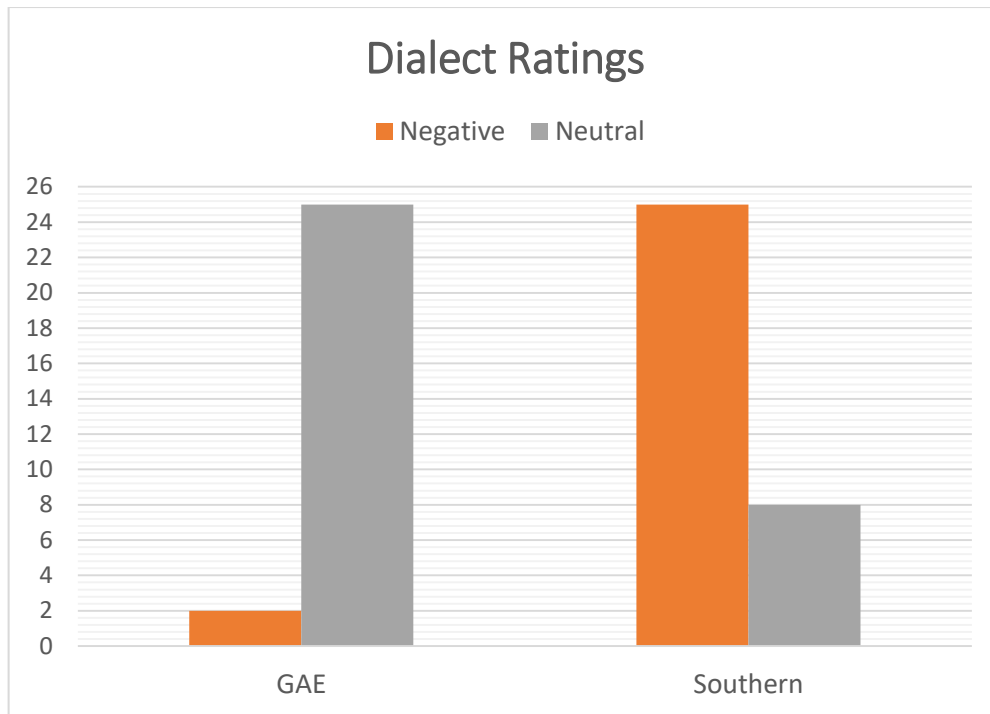
Figure 31. Gender Effect Pilot Study - Accent Ratings

	Estimate	Std. Error	z value	Pr (>|z|)	Sig.
(Intercept)	-6.073	1.52	-3.995	< 0.001	***
Southern	3.547	0.8313	4.267	< 0.001	***

Table 68. Gender Effect Pilot Study - Accent Ratings

(Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)

Similar to Table 3, the coefficients in Table 94 show that GAE was more likely to be seen as having no effect on participants' ability to complete the study. In comparison, the probability that Southern American English would be rated as "hindering" the participants' ability increased. Neither accent was viewed as actively "helping" the participant complete the task.

     I also condensed the four conditions into just Male and Female, and ran these as the independent variable against the neutral/hinder dependent variable. However, as suggested by the results of Table 3, the gender of the speaker was not significant. Not only were they not statistically significant, but as seen in Figure 32 it is clear that the two gender results are near identical.
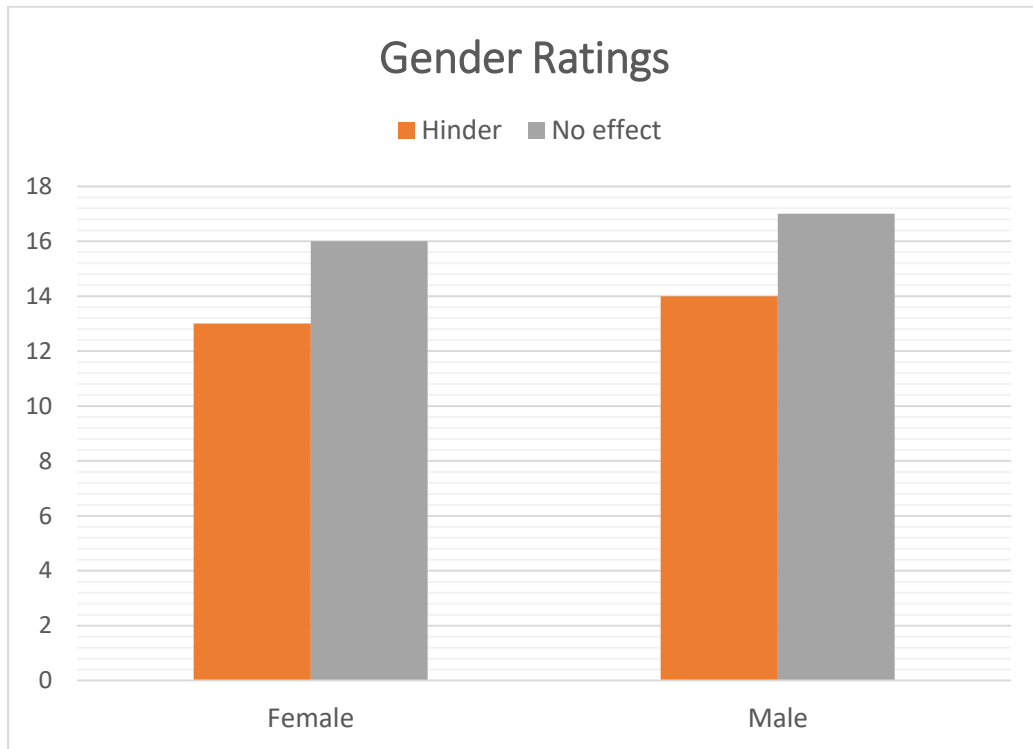
Figure 32. Gender Effect Pilot Study - Gender Ratings

Appendix II: Participant Information and Consent

University of York

Heslington, York, YO10 5DD, UK

Researcher: Grace Wood

Email: glw525@york.ac.uk

You are invited to take part in a research study. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully. If there is anything you do not understand, or if you want more information, please contact the researcher using the email address listed above.

Mock Jury Perceptions of Cross-Examinations
Researcher: Grace Wood

What is the research about?

This study looks into the effectiveness of a prosecutor's examination style and which variables need to be taken into consideration. It is also interested in how juries deliberate and make their decisions.

Who can participate?

Anyone from the United States or who has lived in the United States for two or more years.

What does the study involve?

This survey has two parts. The first will ask a series of questions about you (while maintaining anonymity) in order to learn if answers differ by demographics. The second part will involve watching a short clip of a trial during which the prosecutor cross-examines the defendant. You will be asked to deliberate and come to a verdict as a group. This deliberation will be filmed. You will then be asked a series of questions based on what you observed.

At the end you will have the opportunity to give any feedback or suggestions on how you feel the study could be improved.

This study generally takes 30 minutes.

What will happen to the data I provide?
The data you provide will be used alongside the data of other participants to gain insights into which variables prosecutors ought to be more conscious of during trials. Your data will be stored securely in the University of York, and the results will be written up and used as a part of Grace Wood's PhD dissertation.

Will I know the results?
Because your responses are completely anonymous, it will not be possible to provide you with your own results.

By signing, you agree that:

- *You have read and understood the information about this study.*
- *You understand information you provide will be held in confidentiality and that you have full anonymity.*
- *You understand you can withdraw from this study at any point, without giving a reason, and in such cases your data will be destroyed.*
- *You are giving consent for any information you provide to be used not only in this project, but also in future research.*

Name: _____

Date: _____

Appendix III: Demographic Questions

U.K. Demographic Questionnaire

1. What is your gender?

   o Male

   o Female

   o Prefer not to specify

2. What is your age?

   o Under 18

   o 18-24

   o 25-34

   o 35-64

   o 65+

3. What is your nationality? _____

4. Which region of the United Kingdom do you currently live in?

   _____

5. Have you ever lived anywhere else?

   o Yes – if yes, where else have you lived?

     _____

   o No

6. What is your ethnicity?

   o _____

   o Prefer not to specify

7. What is the highest degree or level of schooling you have completed? *If currently enrolled, highest degree received*

   o _____

8. What is your profession?

   o _____

9. Do you have any children?

   o Yes
   o No

10. In an election, who would you be most likely to vote for?

    o Conservatives
    o Liberal Democrats
    o Labour
    o Green
    o UKIP
    o Unsure
    o Prefer not to specify

11. How often do you drive a vehicle?

    o Daily
    o 3-4 times a month
    o Rarely
    o Never

12. Have you previously served as a member of a jury?

    o Yes
    o No

<u>Mock Juries Demographic Questionnaire</u>

13. What is your gender?

    o Male

    o Female

    o Prefer not to specify

14. What is your age?

    o Under 18

    o 18-24

    o 25-34

    o 35-64

    o 65+

15. What is your nationality? _____

16. Have you ever lived in another state or country?

    o Yes – if yes, where else have you lived?

       _____

    o No

17. What is your ethnicity? *If you prefer not to specify, leave blank*

    o _____

18. What is the highest degree or level of schooling you have completed? *If currently enrolled, highest degree received*

    o _____

19. What is your profession?

    o _____

20. Do you have any children?

   o   Yes

   o   No

21. Do you consider yourself to be:

   o   Conservative

   o   Liberal

   o   Centrist

   o   Unsure

   o   Prefer not to specify

22. How often do you drive a vehicle?

   o   Daily

   o   3-4 times a month

   o   Rarely

   o   Never

23. Have you previously served as a member of a jury?

   o   Yes

   o   No

Appendix IV: Qualtrics Questionnaire Images

After some consideration, how would you, as a juror, find the defendant based on the clip above?

○ Not guilty

○ Guilty

>>

Figure 33. Verdict

What caused you to reach that verdict?

>>

Figure 34. Reasons for verdict

How confident are you in the verdict you have given?

- ○ Unsure
- ○ Somewhat certain
- ○ Certain

>>

Figure 35. Verdict confidence

What did you find most convincing about the prosecutor's argument?

What did you find least convincing about the prosecutor's argument?

>>

Figure 36. Prosecutor

What did you find most convincing about the defendant's testimony?

What did you find least convincing about the defendant's testimony?

>>

Figure 37. Defendant

Can you tell what crime the defendant has been accused of?

○ Yes
○ No

>>

Figure 38. Crime

Do you associate this type of alleged crime with a specific demographic? *(gender, class, race, age, etc.)*

● Yes

○ No

>>

Figure 39. Crime demographic

Who do you associate this type of crime with?

[_____]

>>

Figure 40. Specific crime association

After additional time for consideration, would you change the verdict you gave?

● Yes

○ No

>>
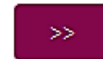
Figure 41. Potential change of verdict

What caused you to change your verdict?

>>

Figure 42. Reasons for change of verdict

Appendix V: Further Descriptive Results

This appendix contains the tables of numbers used to generate the graphs included within the results in chapters 5-7.

Chapter 5. U.K. Question 2: Perception Ratings

This table shows a comparison between the prosecutor ratings and the defendant ratings given in both video guises. The mean numbers were used to generate the readings in Figure 17 in chapter 5.4.

	Video	Mean	Std. Deviation	Std. Error Mean
prosecutor trust	SSBE	3.28	1.137	.227
	Yorkshire	3.04	1.160	.237
prosecutor reliable	SSBE	3.44	1.003	.201
	Yorkshire	2.88	1.076	.220
prosecutor believable	SSBE	3.52	1.229	.246
	Yorkshire	3.21	1.103	.225
defendant trust	SSBE	3.72	1.137	.227
	Yorkshire	3.50	.933	.190
defendant reliable	SSBE	3.40	1.118	.224
	Yorkshire	3.08	1.018	.208
defendant believable	SSBE	4.08	1.288	.258
	Yorkshire	3.88	1.116	.228

Table 69. U.K. Question 2 - Further Stats

Chapter 6. U.S. Qualtrics: Question 2 – Perception Ratings

This table shows a comparison between the U.S. Qualtrics prosecution and defendant ratings given in both video guises. The mean numbers were used to generate the readings in Figure 26 in chapter 6.4.1.

	Video	Mean	Std. Deviation	Std. Error Mean
prosecutor trust	GAE	4.43	.945	.197
	Southern A. E.	4.33	1.051	.183
prosecutor reliable	GAE	4.61	.941	.196
	Southern A.E.	4.42	1.062	.185
prosecutor believable	GAE	4.74	.964	.201
	Southern A.E.	4.58	1.173	.204
defendant trust	GAE	3.52	1.592	.332
	Southern A. E.	3.15	1.460	.254
defendant reliable	GAE	3.22	1.999	.417
	Southern A.E.	3.12	1.495	.260
defendant believable	GAE	3.43	1.701	.355
	Southern A. E.	3.45	1.603	.279

Table 70. U.S. Qualtrics Question 2 - Further Stats

Chapter 6. U.S. Mock Juries: Question 2 – Perception Ratings

This table shows a comparison between the U.S. mock juries' prosecution and defendant ratings given in both video guises. The mean numbers were used to generate the readings in Figure 27 in chapter 6.4.2.

	Video	Mean	Std. Deviation	Std. Error Mean
prosecutor trust	GAE	4.50	1.281	.203
	Southern A. E.	4.26	1.211	.187
prosecutor reliable	GAE	4.65	1.099	.174
	Southern A.E.	4.69	1.070	.165
prosecutor believable	GAE	4.63	1.334	.211
	Southern A.E.	4.67	1.141	.176
defendant trust	GAE	2.83	1.279	.202
	Southern A. E.	2.95	1.306	.201
defendant reliable	GAE	2.65	1.252	.198
	Southern A.E.	2.78	1.351	.211
defendant believable	GAE	2.80	1.436	.227
	Southern A. E.	3.38	1.667	.257

Table 71. U.S. Mock Juries Question 2 - Further Stats

Chapter 6. Research Question III: Question 2 – Perception Ratings

This table shows a comparison between the U.S. Qualtrics dataset and the U.S. mock juries dataset in regards to the prosecution and defendant ratings given in both datasets. The mean numbers were used to generate the readings in Figure 28 in chapter 6.4.3.

	Dataset	Mean	Std. Deviation	Std. Error Mean
prosecutor trust	US Qualtrics	4.38	1.001	.134
	Mock Juries	4.38	1.244	.137
prosecutor reliable	US Qualtrics	4.50	1.009	.135
	Mock Juries	4.67	1.078	.119
prosecutor believable	US Qualtrics	4.64	1.086	.145
	Mock Juries	4.65	1.231	.136
defendant trust	US Qualtrics	3.30	1.513	.202
	Mock Juries	2.89	1.286	.142
defendant reliable	US Qualtrics	3.16	1.703	.228
	Mock Juries	2.72	1.288	.142
defendant believable	US Qualtrics	3.45	1.628	.218
	Mock Juries	3.10	1.576	.174

Table 72. Research Question III: Question 2 - Further Stats

Chapter 7. Research Question IV: Question 1 – Perception Ratings

This table shows a comparison between the U.K. Qualtrics dataset and the U.S. Qualtrics dataset in regards to the prosecution and defendant ratings given in both datasets. The mean numbers were used to generate the readings in Figure 30 in chapter 7.2.

	Dataset	Mean	Std. Deviation	Std. Error Mean
prosecutor trust	U.K.	3.16	1.143	.163
	U.S.	4.38	1.001	.134
prosecutor reliable	U.K.	3.16	1.067	.152
	U.S.	4.50	1.009	.135
prosecutor believable	U.K.	3.37	1.167	.167
	U.S.	4.64	1.086	.145
defendant trust	U.K.	3.61	1.037	.148
	U.S.	3.30	1.513	.202
defendant reliable	U.K.	3.24	1.071	.153
	U.S.	3.16	1.703	.228
defendant believable	U.K.	3.98	1.199	.171
	U.S.	3.45	1.628	.218

Table 73. Research Question IV: Question 1 - Further Stats

Appendix VI: Transcription Guidelines

These guidelines were compiled by me, based on a number of different online recommendations. I used the ones I felt would best serve the purposes of my research and adapted them to be specific to my deliberations (e.g. the speaker being marked by their juror number).

1. Transcribe literally; do not summarize or transcribe phonetically.

2. Informal contractions are to be retained. For example, "gonna", "isn't", "meds". Keep the sentence structure even if it is not grammatically correct.

3. Pauses are indicated by suspension marks in parenthesis (…).

4. Transcribe all verbal communication, including monosyllabic answers. Add interpretations where necessary, for example "Mhm (affirmative)" or "Mhm (negative)".

5. Words with an emphasis are CAPITALIZED.

6. Every contribution by a speaker receives its own paragraph. Add a blank space in between speakers.

7. Emotional non-verbal utterances are transcribed in brackets: [laughter] [sighs].

8. Incomprehensible words are indicated as (inc.). For unintelligible passages indicate the reason: (inc., cell phone ringing) or (inc., microphone rustling). If you assume a certain word but are not sure, put the word in parenthesis with a question mark: (driving?). Mark all inaudible or incomprehensible passages with a time stamp.

9. Discontinuations are marked by /. For example, "I was worri/ concerned."

10. Each speaker is marked by their juror number, e.g. "Juror #3:"

11. Speech overlaps are marked by //. At the start of an interjection, put // and again at the end of the interjection. The other person's interjection is still on a separate line, but also enclosed in //.
    Juror #3: I don't know //why we need to discuss this.//
    Juror #5: //So far I don't have// any concrete ideas.

12. Symbols such as *percent* are spelled out.

13. Once you have a completed transcription, listen to the audio and read through your transcription at the same time to catch any mistakes.

Appendix VII: Non-Significant Table Results

This appendix contains the tables with the full information for the tests that were not significant in chapters 5, 6, and 7.

Chapter 5

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**1.852[a]**	**1**	**.174**
Continuity Correction[b]	1.185	1	.276
Likelihood Ratio	1.863	1	.172
N of Valid Cases	54		

Table 74. UK Verdict – Chi-square test

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	1.840[a]	2	.399	.447
Likelihood Ratio	1.865	2	.394	.447
Fisher's Exact Test	**1.806**			**.447**
No. of Valid Cases	54			

Table 75. UK Confidence – Fisher's Exact

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	1.868[a]	3	.600	.722
Likelihood Ratio	2.235	3	.525	.722
Fisher's Exact Test	**1.855**			**.836**
No. of Valid Cases	27			

Table 76. UK Recommended Punishment – Fisher's Exact

	video	Mean	Std. Deviation	Std. Error Mean
Duration	SSBE	1083.81	845.487	165.814
	YE	1218.56	828.918	159.525

Table 77. UK Questionnaire Duration

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Duration	Equal variances assumed	-.586	51	**.561**	-134.748	230.005
	Equal variances not assumed	-.586	50.827	**.561**	-134.748	230.092

Table 78. UK Questionnaire Duration – Independent T Test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**1.543[a]**	**1**	**.214**
Continuity Correction[b]	.868	1	.352
Likelihood Ratio	1.560	1	.212
No. of Valid Cases	54		

Table 79. UK Video Quality Ratings – Chi-square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**2.299**[a]	**1**	**.129**
Continuity Correction[b]	1.502	1	.220
Likelihood Ratio	2.282	1	.131
N of Valid Cases	56		

Table 80. US Qualtrics Verdict – Chi-square test

	video	Mean	Std. Deviation	Std. Error Mean
confidence	GAE	4.61	1.118	.233
	Southern A.E.	5.06	.966	.168

Table 81. US Qualtrics Confidence

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
confidence	Equal variances assumed	1.614	54	**.112**	-.452	.280
	Equal variances not assumed	1.572	42.892	**.123**	-.452	.287

Table 82. US Qualtrics Confidence – Independent T Test

	t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.
prosecutor - trust	.370	54	**.713**	.101	.274
prosecutor - reliable	.670	54	**.506**	.184	.275
prosecutor – believable	.550	54	**.584**	.163	.297
defendant – trust	.900	54	**.372**	.370	.412
defendant – reliable	.206	54	**.838**	.096	.467
defendant - believable	-.044	54	**.965**	-.020	.446

Table 83. US Qualtrics Perception Ratings – Independent T Test

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	.629[a]	1	.428	.506
Likelihood Ratio	.203	1	.426	.506
Fisher's Exact Test	**.633**			**.506**
N of Valid Cases	38			

Table 84. US Qualtrics Recommended Punishment – Fisher's Exact Test

	video	Mean	Std. Deviation	Std. Error Mean
Duration	GAE	835.74	476.819	99.424
	Southern A. E.	912.76	570.310	99.278

Table 85. US Qualtrics Questionnaire Duration

		t	df	**Sig. (2-tailed)**	Mean Diff.	Std. Error Diff.
Duration	Equal variances assumed	.531	54	**.598**	-77.018	145.103
	Equal variances not assumed	.548	52.120	**.586**	-77.018	140.503

Table 86. US Qualtrics Questionnaire Duration – Independent T Test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**1.273[a]**	**1**	**.259**
Continuity Correction[b]	.675	1	.411
Likelihood Ratio	1.259	1	.262
N of Valid Cases	56		

Table 87. US Qualtrics Video Quality Ratings – Chi-square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**1.059[a]**	**1**	**.303**
Continuity Correction[b]	.000	1	1.000
Likelihood Ratio	4.123	1	.229
N of Valid Cases	18		

Table 88. US Mock Juries Verdict – Chi-square Test

	video	Mean	Std. Deviation	Std. Error Mean
confidence	GAE	5.08	1.163	.184
	Southern A.E.	5.29	.774	.119

Table 89. US Mock Juries Confidence

		t	df	**Sig. (2-tailed)**	Mean Diff.	Std. Error Diff.
confidence	Equal variances assumed	-.970	80	**.335**	-.211	.217
	Equal variances not assumed	-.961	67.427	**.340**	-.211	.219

Table 90. US Mock Juries Confidence – Independent T Test

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	.721[a]	2	.697	.700
Likelihood Ratio	.724	2	.696	.700
Fisher's Exact Test	**.748**			**.736**
N of Valid Cases	79			

Table 91. US Mock Juries Recommended Punishment – Fisher's Exact

	video	Mean	Std. Deviation	Std. Error Mean
Duration	GAE	257.1250	182.76329	28.89741
	Southern A.E.	234.5238	232.72304	35.90994

Table 92. US Mock Juries Deliberation Length

		t	df	Sig. (2-tailed)	Mean Diff	Std. Error Diff
Duration	Equal variances assumed	.487	80	**.627**	22.60119	46.36383
	Equal variances not assumed	.490	77.242	**.625**	22.60119	46.09322

Table 93. US Mock Juries Deliberation Length – Independent T Test

	Group	Mean	Std. Deviation	Std. Error Mean
confidence	US Qualtrics	4.88	1.046	.140
	Mock Juries	5.18	.983	.109

Table 94. RQ III Confidence

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
confidence	-1.761	136	.080	-.308	.175

Table 95. RQ III Confidence – Independent T Test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**.003ᵃ**	**1**	**.957**
Continuity Correctionᵇ	.000	1	1.000
Likelihood Ratio	.003	1	.957
N of Valid Cases	117		

Table 96. RQ III Recommended Punishment – Condensed Chi-square Test

	dataset	Mean	Std. Deviation	Std. Error Mean
duration	U.K.	1152.45	831.778	114.235
	U.S.	881.13	530.7	70.918

Table 97. RQ IV Duration

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**.010^a**	**1**	**.919**
Continuity Correction^b	.000	1	1.000
Likelihood Ratio	.010	1	.919
N of Valid Cases	110		

Table 98. RQ IV Video Quality Ratings – Chi-square test

References

Abad-Santo, A. (2013). My Star Witness Is Black: Rachel Jeantel's Testimony Makes Trayvon a Show Trial. *The Atlantic*. [online] Available at: https://www.theatlantic.com/national/archive/2013/06/rachel-jeantel-testimony-trayvon-martin-trial/313792/ [Accessed 25 Jun. 2018].

Allbritten, R. (2011). *Sounding Southern: Phonetic Features and Dialect Perceptions*. PhD Thesis. Georgetown University.

Bailey, G. (1997). Why Did Southern American English Begin. In: Schneider, E. (ed.) *Englishes Around the World*, 1st ed. Amsterdam: John Benjamins Publishing Co., pp.255-276.

Barry, E. (2005). *Learn English, Judge Tells Moms*. [online] LATimes. Available at: http://articles.latimes.com/2005/feb/14/nation/na-english14 [Accessed 10 Aug. 2018].

Bayard, D., Weatherall, A., Gallois, C. and Pittam, J. (2001). Pax Americana? Accent Attitudinal Evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*, 5(1), pp.22-49.

BBC News. (2017). *Labour MP Angela Rayner: 'I'm proud of my accent'*. [online] Available at: https://www.bbc.co.uk/news/uk-40554137 [Accessed 8 Aug. 2018].

Beal, J. (2006). *Language and Region*. London: Routledge.

Beal, J. (2008). English Dialects in the North of England: Phonology. In: B. Kortmann and C. Upton, (eds.) *The British Isles*. Berlin: Walter de Gruyter, pp.122-144.

Beal, J. and Cooper, P. (2015). The Enregisterment of Northern English. In: R. Hickey, (ed.) *Researching Northern English*. Amsterdam: John Benjamin Publishing Co., pp.27-49

Bell, A. (1984). Language style as audience design. *Language in Society*, 13(02), p.145.

Bernal, G., Bonilla, J. and Bellido, C. (1995). Ecological Validity and Cultural Sensitivity for Outcome Research: Issues for the Cultural Adaptation and Development of Psychosocial Treatments with Hispanics. *Journal of Abnormal Child Psychology*, 23(1), pp.67-82.

Beta.Gov.Wales. (2018). *Traffic Signs and Road Markings | beta.gov.wales*. [online] Available at: https://beta.gov.wales/traffic-signs-and-road-markings [Accessed 8 Aug. 2018].

Birdsong, D. (1999). Whys and Why Nots of the Critical Period Hypothesis for Second Language Acquisition. In: D. Birdsong, (ed.) *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah: Lawrence Erlbaum Associates Inc., pp.1-22.

Bornstein, B. (1999). The Ecological Validity of Jury Simulations: Is the Jury Still Out?. *Law and Human Behavior*, 23(1), pp.75-91.

Bornstein, B., Golding, J., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C. and Luecht, K. (2017). Mock Juror Sampling Issues in Jury Simulation Research: A meta-analysis. *Law and Human Behavior*, 41(1), pp.13-28.

Bornstein, B. and Greene, E. (2011). Jury Decision Making: Implications For and From Psychology. *Current Directions in Psychological Science*, 20(1), pp.63-67.

Bougheas, S., Nieboer, J. and Sefton, M. (2013). Risk-Taking in Social Settings: Group and Peer Effects. *Journal of Economic Behavior & Organization*, 92, pp.273-283.

Britain, D. (2007). *Language in the British Isles*. Cambridge: Cambridge University Press.

Brodsky, S. (2013). *Testifying in Court: Guidelines and Maxims for the Expert Witness*. 2nd ed. Washington: American Psychological Association.

Buhrmester, M., Kwang, T. and Gosling, S. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, 6(1), pp.3-5.

Buhrmester, M., Talaifar, S. and Gosling, S. (2018). An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use. *Perspectives on Psychological Science*, 13(2), pp.149-154.

Cameron, D. (2003). Gender and Language Ideologies. In: J. Holmes and M. Meyerhoff, (eds.) *The Handbook of Language and Gender*. Oxford: Blackwell Publishing Ltd., pp.447-467.

Cameron, D. (2012). *Verbal Hygiene*. Milton Park: Routledge.

Campbell-Kibler, K. (2006). *Listener Perceptions of Sociolinguistic Variables: The Case of (ING)*. Ph.D. Stanford University.

Campbell-Kibler, K. (2009). The Nature of Sociolinguistic Perception. *Language Variation and Change*, 21(1), pp.135-156.

Cargile, A. (2000). Evaluations of Employment Suitability: Does Accent Always Matter?. *Journal of Employment Counseling*, 37(3), pp.165-177.

Cargile, A., Giles, H., Ryan, E. and Bradac, J. (1994). Language Attitudes as a Social Process: A Conceptual Model and New Directions. *Language & Communication*, 14(3), pp.211-236.

Carlin, A. (2016). The Courtroom as White Space: Racial Performance as Noncredibility. *UCLA Law Review*, 63, pp.450-484.

Carodine, M. (2015). Contemporary Issues In Critical Race Theory: The Implications Of Race As Character Evidence In Recent High-Profile Cases. *University of Pittsburgh Law Review*, 75(4), pp.679-692.

Cesario, J. and Navarrete, C. (2013). Perceptual Bias in Threat Distance: The Critical Roles of In-Group Support and Target Evaluations in Defensive Threat Regulation. *Social Psychological and Personality Science*, 5(1), pp.12-17.

Cheshire, J. and Milroy, J. (2013). Syntactic Variation in Non-Standard Dialects: Background Issues. In: J. Milroy and L. Milroy, (eds.) *Real English: The Grammar of English Dialects in the British Isles*, 2nd ed. Milton Park: Routledge, pp.3-33.

Cheshire, J., Fox, S., Kerswill, P. and Torgersen, E. (2013). Language Contact and Language Change in the Multicultural Metropolis. *Revue Française de Linguistique Appliquée*, 18(2), 63-76.

Clifford v. Chandler [2003] No. 01-5926 (United States Court of Appeals, Sixth Circuit).

Clifford v. Commonwealth [1999] 97-SC-368-MR (Supreme Court of Kentucky).

Clopper, C. and Pisoni, D. (2004). Some Acoustic Cues for the Perceptual Categorization of American English Regional Dialects. *Journal of Phonetics*, 32(1), pp.111-140.

Clopper, C. and Pisoni, D. (2006). The Nationwide Speech Project: A New Corpus of American English Dialects. *Speech Communication*, 48(6), pp.633-644.

Clopper, C., Pisoni, D. and de Jong, K. (2005). Acoustic Characteristics of the Vowel Systems of Six Regional Varieties of American English. *The Journal of the Acoustical Society of America*, 118(3), pp.1661-1676.

Collins, K. A., and R. Clement. (2012). Language And Prejudice: Direct And Moderated Effects. *Journal of Language and Social Psychology*, 31(4), pp.376-396.

Corder, B. F., & Whiteside, R. (1988). A Survey of Jurors' Perception of Issues Related to Child Sexual Abuse. *American Journal of Forensic Psychology,* 6(3), pp.37-43.

Coupland, N. (1988). *Dialect in Use*. Cardiff: University of Wales Press.

Coupland, N. (2000). Sociolinguistic Prevarication About 'Standard English'. *Journal of Sociolinguistics*, 4(4), pp.622-634.

Coupland, N. and Aldridge, M. (2009). Introduction: A Critical Approach to the Revitalisation of Welsh. *International Journal of the Sociology of Language*, 195(1), pp.5-13.

Coupland, N. and Bishop, H. (2007). Ideologised Values for British Accents. *Journal of Sociolinguistics*, 11(1), pp.74-93.

Creber, C. and Giles, H. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences*, 5(2), pp.155-161.

Crowley, T. (1991). *Proper English*. Milton Park: Routledge.

Davies, W. and Ziegler, E. (2015). Language Planning and Microlinguistics: Introduction. In: W. Davies and E. Ziegler, (eds.) *Language Planning and Microlinguistics*. London: Palgrave Macmillan, pp.1-12.

Deitz, S., Blackwell, K., Daley, P. and Bentley, B. (1982). Measurement of Empathy Toward Rape Victims and Rapists. *Journal of Personality and Social Psychology*, 43(2), pp.372-384.

Del Valle, S. (2003). *Language Rights and the Law in the United States: Finding Our Voices*. Buffalo: Multilingual Matters Ltd.

Devine, D., Clayton, L., Dunford, B., Seying, R. and Pryce, J. (2001). Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups. *Psychology, Public Policy, and Law*, 7(3), pp.622-727.

Dhami, M. and Belton, I. (2016). Statistical Analyses of Court Decisions: An Example of Multilevel Models of Sentencing. *Law and Method*, 10, pp.247-266.

Dieter, R. (2003). International Influence on the Death Penalty in the U.S. *Foreign Service Journal*, 80(10), pp.31-38.

Dixon, J. and Mahoney, B. (2004). The Effect of Accent Evaluation and Evidence on a Suspect's Perceived Guilt and Criminality. *The Journal of Social Psychology*, 144(1), pp.63-73.

Dixon, J., Mahoney, B. and Cocks, R. (2002). Accents of Guilt? Effects of Regional Accent, Race, and Crime Type on Attributions of Guilt. *Journal of Language and Social Psychology*, 21(2), pp.162-168.

Eades, D. (2003). Participation of Second Language and Second Dialect Speakers in the Legal System. *Annual Review of Applied Linguistics*, 23, pp.113-133.

Eagly, A. (2007). Female Leadership Advantage and Disadvantage: Resolving the Contradictions. *Psychology of Women Quarterly*, 31(1), pp.1-12.

Eckert, P. (2000). *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), pp.453-476.

Eckert, P. (2018). *Meaning and linguistic variation*. Cambridge: Cambridge University Press.

Edwards, K. (2014). *Shut yer face! I'm fed up being ridiculed for my regional accent in academia*. [online] Telegraph.co.uk. Available at: https://www.telegraph.co.uk/women/womens-life/11270980/British-universities-Im-fed-up-of-being-ridiculed-for-my-regional-accent.html [Accessed 7 Aug. 2018].

Eligon, J. (2019). Speaking Black Dialect in Courtrooms Can Have Striking Consequences. [online] Nytimes.com. Available at: https://www.nytimes.com/2019/01/25/us/black-dialect-courtrooms.html?te=1&nl=race-related&emc=edit_rr_20190126 [Accessed 27 Mar. 2019].

Esqueda, C., Espinoza, R. and Culhane, S. (2008). The Effects of Ethnicity, SES, and Crime Status on Juror Decision Making: A Cross-Cultural Examination of European American and Mexican American Mock Jurors. *Hispanic Journal of Behavioral Sciences*, 30(2), pp.181-199.

Evans, B. and Iverson, P. (2004). Vowel Normalization for Accent: An Investigation of Best Exemplar Locations in Northern and Southern British English Sentences. *The Journal of the Acoustical Society of America*, 115(1), pp.352-361.

Fabricius, A. (2018). Social Change, Linguistic Change and Sociolinguistic Change in Received Pronunciation. In: N. Braber and S. Jansen, (eds.) *Sociolinguistics in England.* London: Palgrave Macmillan, pp. 35-66.

Feagin, C. (1996). Peaks and Glides in Southern States Short-a. In: G. Guy, C. Feagin, D. Schiffrin and J. Baugh, ed., *Towards a Social Science of Language: Papers in honor of William Labov*. Philadelphia: John Benjamins Publishing Co., pp.135-160.

Ferguson, R. (2002). *Why Judges Don't Explain Their Rulings*. [online] The West Law Firm. Available at: http://www.westfirm.com/judges_rulings.html [Accessed 25 Jun. 2018].

Finegan, E. and Rickford, J. (2004). *Language in the USA*. Cambridge: Cambridge University Press.

FLA. STAT. § 776.013 (2018).

Fletcher, T. (2012). 'All Yorkshiremen are from Yorkshire, but some are more "Yorkshire" than others': British Asians and the Myths of Yorkshire Cricket. *Sport in Society*, 15(2), pp.227-245.

Forsyth, D. (2017). *Group Dynamics*. 7th ed. Boston: Cengage.

Foulkes, P. (2006). Phonological Variation: A Global Perspective. In: B. Aarts and A. McMahon, (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing Ltd., pp.625-669.

Fraser, H. (2011). The Role of Linguists and Native Speakers in Language Analysis for the Determination of Speaker Origin. *International Journal of Speech, Language & the Law*, *18*(1), pp.121-130.

Freeman, R. (1998). *Bilingual Education and Social Change*. Clevedon [England]: Multilingual Matters.

Frumkin, L. (2007). Influences of Accent and Ethnic Background on Perceptions of Eyewitness Testimony. *Psychology, Crime & Law*, 13(3), pp.317-331.

Garrett, P. (2010). *Attitudes to Language*. Cambridge: Cambridge University Press.

Garrett, P. (2002). Language Attitudes and Sociolinguistics. *Journal of Sociolinguistics*, 5(4), pp.626-631.

Ghorshi, S., Vaseghi, S. and Yan, Q. (2008). Cross-entropic Comparison of Formants of British, Australian and American English Accents. *Speech Communication*, 50(7), pp.564-579.

Giles, H. (1970). Evaluative Reactions to Accents. *Educational Review*, 22(3), pp.211-227.

Giles, H. (1971). Patterns of Evaluation to R.P., South Welsh and Somerset Accented Speech. *British Journal of Social and Clinical Psychology*, 10(3), pp.280-281.

Giles, H. (1972). Evaluation of Personality Content from Accented Speech as a Function of Listeners' Social Attitudes. *Perceptual and Motor Skills*, 34(1), pp.168-170.

Giles, H. and Coupland, N. (1991). *Language: Contexts and Consequences*. Milton Keynes: Open University Press.

Giles, H. and Powesland, P. (1975). *Speech Style and Social Evaluation*. Oxford: Academic Press.

Giles, H., Henwood, K., Coupland, N., Harriman, J. and Coupland, J. (1992). Language Attitudes and Cognitive Mediation. *Human Communication Research*, 18(4), pp.500-527.

Gill, M. (2016). *Angela Rayner Called 'Thick As Mince' In Abusive Emails About Her Accent*. [online] HuffPost UK. Available at: https://www.huffingtonpost.co.uk/entry/angela-rayner-hits-back-at-abusive-emails-over-her-accent_uk_5811d990e4b04660a438156a?k1vmfpf73davpldi&guccounter=1&guce_referrer_us=aHR0cHM6Ly90LmNvL0g0NWpndElOY1E&guce_referrer_cs=2rt7SfMO03iFBhjQfeSKdA [Accessed 8 Aug. 2018].

Goodman, J., Cryder, C. and Cheema, A. (2012). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26(3), pp.213-224.

Gordon, M. (2008). The West and Midwest: Phonology. In: E. Schneider, ed., *The Americas and the Caribbean*. New York: Mouton dr Gruyter, pp.129-143.

Gov.UK. (n.d.). *Jury Service*. [online] Available at: https://www.gov.uk/jury-service/discussing-the-trial [Accessed 25 Jun. 2018].

Gray, J. (2010). *Does Listener Age Affect the Ratings of Guilt Attributed to a Suspect with a Standard of Non-standard Accent?* MSc Thesis. University of York.

Greene, E. and Bornstein, B. (2003). *Determining Damages*. Washington, DC: American Psychological Association.

Habick, T. (1980). *Sound Change in Farmer City: a Sociolinguistic Study Based on Acoustic Data*. PhD Thesis. University of Illinois.

Haegerich, T. and Bottoms, B. (2000). Empathy and Jurors' Decisions in Patricide Trials Involving Child Sexual Assault Allegations. *Law and Human Behavior*, 24(4), pp.421-448.

Haley v. Brown [2006] S.E.2d 62 (Court of Appeals of South Carolina).

Hall, J. (2013). *Teaching and Researching Language and Culture*. 2nd ed. Milton Park: Routledge.

Hall-Lew, L., Friskney, R. and Scobbie, J. (2017). Accommodation or Political Identity: Scottish Members of the UK Parliament. *Language Variation and Change*, 29(03), pp.341-363.

Hall-Lew, L., Starr, R. and Coppock, E. (2012). Style-Shifting in the U.S. Congress: The vowels of 'Iraq(i)'. In: J. Hernandez Campoy and J. Cutillas Espinosa, ed., *Style-Shifting in Public: New Perspectives on Stylistic Variation*. Amsterdam: John Benjamins Publishing Co, pp.45-63.

Hay, J., Drager, K. and Warren, P. (2010). Short-term Exposure to One Dialect Affects Processing of Another. *Language and Speech*, 53(4), pp.447-471.

Heaton, H. and Nygaard, L. (2011). Charm or Harm: Effect of Passage Content on Listener Attitudes Toward American English Accents. *Journal of Language and Social Psychology*, 30(2), pp.202-211.

Hengstler, G. (1993). Vox Populi - The Public Perception of Lawyers: ABA Poll. *American Bar Association*, 79(9), pp.60-65.

Hickey, R. (2015). The North of England and Northern English. In: R. Hickey, (ed.) *Researching Northern English*. Amsterdam: John Benjamin Publishing Co., pp.1-24

Hinton, P. (2000). *Stereotypes, Cognition, and Culture*. Hove, East Sussex: Psychology Press.

Hiraga, Y. (2005). British Attitudes Towards Six Varieties of English in the USA and Britain. *World Englishes*, 24(3), pp.289-308.

Hollien, H. (2002). *Forensic Voice Identification*. San Diego, Calif.: Academic Press.

Hongju Koh, H. and Pickering, T. (2003). American Diplomacy and the Death Penalty. *Foreign Service Journal*, 80(10), pp.19-25.

Howley, G. (2018). *Love Island: Audience reactions to northern accents show something deeply unsavoury about Britain*. [online] The Independent. Available at: https://www.independent.co.uk/voices/love-island-snobbery-class-northern-accents-audience-southern-a8415471.html [Accessed 7 Aug. 2018].

Huebner, T. (1999). Sociolinguistic Perspectives on Language Policy, Politics, and Praxis. In: Huebner, T. and Davis, K. (eds.) *Sociopolitical perspectives on language policy and planning in the USA*. Amsterdam: John Benjamins Publishing Co., pp.1-15.

Hughes, A., Trudgill, P. and Watt, D. (2012). *English Accents and Dialects*. 5th ed. Hodder Education.

Hurwitz, J. and Peffley, M. (1997). Public Perceptions of Race and Crime: The Role of Racial Stereotypes. *American Journal of Political Science*, [online] 41(2), p.375. Available at: http://www.jstor.org/stable/2111769 [Accessed 5 Jul. 2017].

Hyppolite v. State [2002] No. 48A04-0107-PC-00314 (Court of Appeals of Indiana).

Itv.com. (2018). *Does your accent affect your chances in life?*. [online] Available at: https://www.itv.com/goodmorningbritain/news/does-your-accent-affect-your-chances-in-life [Accessed 8 Aug. 2018].

Iyengar, S. and Westwood, S. (2014). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3), pp.690-707.

Jena, A. (2017). *'Null' research findings aren't empty of meaning. Let's publish them*. [online] STAT. Available at: https://www.statnews.com/2017/11/10/null-research-findings/ [Accessed 28 Mar. 2019].

Johnson, J. (2013). Love her or hate her, Rachel Jeantel is a star. *CNN*. [online] Available at: https://edition.cnn.com/2013/06/27/opinion/zimmerman-jeantel/index.html [Accessed 25 Jun. 2018].

Johnstone, B., Andrus, J. and Danielson, A. (2006). Mobility, Indexicality, and the Enregisterment of "Pittsburghese". *Journal of English Linguistics*, 34(2), pp.77-104.

Jost, J., Burgess, D. and Mosso, C. (2001). Conflicts of Legitimation among Self, Group, and System: The Integrative Potential of System Justification Theory. In: Jost, J and Major, B. (eds.) *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge: Cambridge University Press, pp. 363-388.

Keene, D. (2013). The "Why" and "How" of Focus Group Research. *The Jury Expert: The Art and Science of Litigation Advocacy*, 25(4).

Kerswill, P. (2007) Standard English, RP and the Standard—non-Standard Relationship. In: Britain, D. (ed.) *Language in the British Isles*. 2nd ed. Cambridge: Cambridge University Press, pp.34-51.

Kerswill, P. and Culpeper, J. (2018). Standard English and Standardization. In: J. Culpeper, P. Kerswill, R. Wodak, T. McEnery and F. Katamba, (eds.) *English Language: Description, Variation and Context*, 2nd ed. London: Palgrave Macmillan, pp.167-185.

Khan, S., Benda, T. and Stagnaro, M. (2012). Stereotyping From the Perspective of Perceivers and Targets. *Online Readings in Psychology and Culture*, 5(1), pp.3-10.

Kibbee, D. (1998). Legal and Linguistic Perspectives on Language Legislation. In: D. Kibbee, (ed.) *Language Legislation and Linguistic Rights*. Philadelphia: John Benjamins Publishing Co., pp.1-23.

Kirkham, S. (2013). *Ethnicity, social practice and phonetic variation in a Sheffield secondary school*. PhD Thesis. University of Sheffield.

Kohler, B. (2004). Racial Voice Identification: Judicially Condoning the Bogus Science of Hearing Color. *Temple Law Review*, 77(3), pp.757-780.

Kretzschmar, W. (2008). Standard American English Pronunciation. In: E. Schneider, ed., *The Americas and the Caribbean*. New York: Mouton dr Gruyter, pp.37-51.

Kristiansen, G. (2001). Social And Linguistic Stereotyping: A Cognitive Approach To Accents. *Estudios Ingleses de la Universidad Complutense*, 9, pp.129-145.

Ladegaard, H. (2000). Language Attitudes and Sociolinguistic Behaviour: Exploring Attitude-Behaviour Relations in Language. *Journal of Sociolinguistics*, 4(2), pp.214-233.

Lambert, W., Hodgson, R., Gardner, R. and Fillenbaum, S. (1960). Evaluational Reactions to Spoken Languages. *The Journal of Abnormal and Social Psychology*, 60(1), pp.44-51.

Langadinos v. Appalachian School of Law [2005] 1:05CV00039 (United States District Court for the Western District of Virginia Abingdon Division).

Leib, E. (2008). A Comparison of Criminal Jury Decision Rules in Democratic Countries. *Ohio State Journal of Criminal Law*, 5, pp.629-644.

Leith, D. (1997). *A Social History of English*. 2nd ed. London: Routledge.

Levinson, S. (2018). What Do Lawyers Know (and What Do They Do With Their Knowledge)? Comments on Schauer and Moore. *Southern California Law Review*, 58(44), pp.441-458.

Linder, D. (1996). Juror Empathy and Race. *Tennessee Law Review*, 63, pp.887-916.

Lippi-Green, R. (1994). Accent, Standard Language Ideology, and Discriminatory Pretext in the Courts. *Language in Society*, 23(2), pp.163-198.

Lippi-Green, R. (2012). *English with an Accent*. 2nd ed. New York: Routledge.

Llamas, C., Harrison, P., Donnelly, D., & Watt, D. (2008). Effects of Different Types of Face Coverings on Speech Acoustics and Intelligibility. *York Papers in Linguistics Series 2*(9), pp.80-104.

Local Government Act 1972, c.70. Available at: http://www.legislation.gov.uk/ukpga/1972/70 (Accessed: 2 August 2017).

Lodge, R. (1993). *French: From Dialect to Standard*. London: Routledge.

MacCoun, R. (1989). Experimental Research on Jury Decision-Making. *Science*, 244(4908), pp.1046-1050.

Maerowitz, M. (2014). "Perceptual Dialectology: Accents and Jury Verdicts". Undergraduate Thesis. Arizona State University, Tempe.

Mangat, L. (2018). A Jury of Your [Redacted]: The Rise and Implications of Anonymous Juries. *Cornell Law Library Prize for Exemplary Student Research Papers*, 15, pp.1-33.

Marshall, P. (2013). *28% of Britons feel discriminated against due to accent*. [online] ITV News. Available at: http://www.itv.com/news/2013-09-25/28-of-britons-feel-discriminated-against-due-to-accent/ [Accessed 8 Aug. 2018].

Martin, A., Kaplan, M. and Alamo, J. (2003). Discussion Content and Perception of Deliberation in Western European Versus American Juries. *Psychology, Crime & Law*, 9(3), pp.247-463.

Massey, D. and Lundy, G. (2001). Use of Black English and Racial Discrimination in Urban Housing Markets. *Urban Affairs Review*, 36(4), pp.452-469.

Matsuda, M. (1991). Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction. *The Yale Law Journal*, 100(5), pp.1329-1407.

McCoy, M. and Gray, J. (2007). The Impact of Defendant Gender and Relationship to Victim on Juror Decisions in a Child Sexual Abuse Case. *Journal of Applied Social Psychology*, 37(7), pp.1578-1593.

Milroy, J. and Milroy, L. (2012). *Authority in Language*. 4th ed. London: Routledge.

Milroy, L. (2000). Britain and the United States: Two Nations Divided by the Same Language (and Different Language Ideologies). *Journal of Linguistic Anthropology*, 10(1), pp.56-89.

Miner, S. (2013). Legal Implications of the Official English Declaration. In: Ricento, T. and Burnaby, B. (eds.) *Language and politics in the United States and Canada: myths and realities*, New York: Routledge, pp.171-184.

Mitchell, D. (2017). *Why is there still so much prejudice towards working-class accents?*. [online] the Guardian. Available at: https://www.theguardian.com/commentisfree/2017/jul/11/prejudice-working-class-accents-mp-thick-social-media [Accessed 7 Aug. 2018].

Montgomery, C. and Moore, E. (2018). Evaluating S(c)illy Voices: The Effects of Salience, Stereotypes,and Co-Present Language Variables on Real-Time Reactions to Regional Speech. *Language*, 94(3), pp.629-661.

Morley, J. (1996). Second Language Speech/Pronunciation: Acquisition, Instruction, Standards, Variation, and Accent. In: Alatis, J., Straehle, C., Ronkin, M. and Gallenberger, B. (eds.) *Georgetown University Round Table On Languages And Linguistics (GURT) 1996: Linguistics, Language Acquisition, And Language Variation: Current Trends And Future Prospects*. Washington, DC: Georgetown University Press, pp.140-160.

Mugglestone, L. (2003). *Talking Proper: The Rise of Accent as Social Symbol*. 2nd ed. Oxford: Oxford University Press.

Munafò, M. and Neill, J. (2016). Null is beautiful: On the importance of publishing null results. *Journal of Psychopharmacology*, 30(7), pp.585-585.

Myers, D. and Lamm, H. (1976). The Group Polarization Phenomenon. *Psychological Bulletin*, 83(4), pp.602-627.

Neill, M. (1994). Broken English and Broken Irish: Nation, Language, and the Optic of Power in Shakespeare's Histories. *Shakespeare Quarterly*, 45(1), pp.1-32.

Nguyen, B. (1993). Accent Discrimination and the Test of Spoken English: A Call for an Objective Assessment of the Comprehensibility of Nonnative Speakers. *California Law Review*, 81(5), pp.1325-1361.

Norton, M., Frost, J. and Ariely, D. (2007). Less is More: The Lure of Ambiguity, or Why Familiarity Breeds Contempt. *Journal of Personality and Social Psychology*, 92(1), pp.97-105.

Nugaraitė, I. (2017). Standard Language Ideology and Its Influence on Lithuanian Migrants. Samogitians' Attitudes Towards Their Language Variety. *Sustainable Multilingualism*, 11(1), pp.14-33.

Office of the Federal Register (1983). *Code of Federal Regulations*. Washington D.C.: U.S. Government Printing Office.

Olivieri, V. (2015). Sub-State Nationalism in Spain: Primers and Triggers of Identity Politics in Catalonia and the Basque Country. *Ethnic and Racial Studies*, 38(9), pp.1610-1626.

Omoniyi, T. and White, G. (2006). *The Sociolinguistics of Identity*. London: Continuum.

Owens, C. (2019). Are Philly court reporters accurate with black dialect? Study: Not really.. [online] https://www.philly.com. Available at: https://www.philly.com/news/court-reporter-stenographer-african-american-english-aave-philly-transcript-study-20190122.html [Accessed 27 Mar. 2019].

Oxford English Dictionary. (2018a). *discrimination*. [online] Available at: http://www.oed.com/view/Entry/54060?redirectedFrom=discrimination#eid [Accessed 25 Jun. 2018].

Oxford English Dictionary. (2018b). *prejudice*. [online] Available at: http://www.oed.com/view/Entry/150162?rskey=ImDxQe&result=1&isAdvanced=false#eid [Accessed 25 Jun. 2018].

Oxford English Dictionary. (2018c). *prosody*. [online] Available at: http://www.oed.com/view/Entry/152998#eid27915687 [Accessed 20 Aug. 2018].

Paolacci, G. and Chandler, J. (2014). Inside the Turk. *Current Directions in Psychological Science*, 23(3), pp.184-188.

Pelly, J. (2018). *Will a Southern Drawl Close Doors for My Son?*. [online] Nytimes.com. Available at: https://www.nytimes.com/2018/06/01/well/family/will-a-southern-drawl-close-doors-for-my-son.html [Accessed 8 Aug. 2018].

Petyt, K. (1985). *Dialect and Accent in Industrial West Yorkshire*. Amsterdam: John Benjamins Publishing Co.

Piller, I. (2015). Language Ideologies. In: Tracy, K., Ilie, C., and Sandel, T. (eds.) *The International Encyclopedia of Language and Social Interaction.* vol. 2, The Wiley Blackwell-ICA international encyclopedias of communication, Wiley-Blackwell, Wiley, West Sussex, pp.917-927.

Phillips, J. [joglasg]. (2017, January 25). One of the emails Angela Rayner was sent about her accent – can you imagine sending that with your details? Jeeeesus. A teacher sent that? [Tweet]. Retrieved from https://twitter.com/joglasg/status/824275874837135361.

Powell, G. (2018). *Women and Men in Management*. 5th ed. Thousand Oaks: Sage Publications.

Pozzulo, J., Dempsey, J., Maeder, E. and Allen, L. (2009). The Effects of Victim Gender, Defendant Gender, and Defendant Age on Juror Decision Making. *Criminal Justice and Behavior*, 37(1), pp.47-63.

Preston, D. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Providence: Foris Publications.

Preston, D. (1999). *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins Publishing Co.

Preston, D. (2002a). Language with an Attitude. In: J. Chambers, P. Trudgill and N. Schilling-Estes, (eds.) *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing Ltd., pp.40-66.

Preston, D. (2002b). Perceptual Dialectology: Aims, Methods, Findings. In: Berns, J. and Marle, J. (eds.) *Present-Day Dialectology*. Berlin: M. de Gruyter, pp.57-104.

Preston, D. and Robinson, G. (2005). Dialect Perception and Attitudes to Variation. In: M. Ball, (ed.) *Clinical Sociolinguistics*. Malden: Blackwell Publishing Ltd., pp.133-149.

Przedlacka, J. (2005). Models and Myth: Updating the (Non)standard Accents. In: Dziubalska-Kołaczyk, K. and Przedlacka, J. (eds.) *English Pronunciation Models*. Bern: Peter Lang, pp.17-36.

Purnell, T., Idsardi, W. and Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. *Journal of Language and Social Psychology*, 18(1), pp.10-30.

Quigley, R. (2013). 'That's retarded sir': Trayvon Martin trial's star witness cracks when her credibility is questioned and is forced to admit she could not read letter she allegedly wrote. *Daily Mail*. [online] Available at: http://www.dailymail.co.uk/news/article-2349794/George-Zimmerman-trial-Rachel-Jeantel-Trayvon-Martin-prosecutions-star-witness-dragged-coals-defense.html [Accessed 25 Jun. 2018].

Rakić, T., Steffens, M. and Mummendey, A. (2011). When It Matters How You Pronounce It: The Influence of Regional Accents on Job Interview Outcome. *British Journal of Psychology*, 102(4), pp.868-883.

Richards, H. (2008). *Mechanisms, Motivations and Outcomes of Change in Morley (Leeds) English.* PhD Thesis. University of York.

Rickford, J. and King, S. (2016). Language and Linguistics on Trial: Hearing Rachel Jeantel (and other vernacular speakers) in the Courtroom and Beyond. *Language*, 92(4), pp.948-988.

R.L.G. (2015). *The last acceptable prejudice*. [online] The Economist. Available at: https://www.economist.com/prospero/2015/01/29/the-last-acceptable-prejudice [Accessed 7 Aug. 2018].

Rodman, R. (2002). Linguistics and the Law: How Knowledge of, or Ignorance of, Elementary Linguistics May Affect the Dispensing of Justice. *Forensic Linguistics*, 9(1), pp.94-103.

Ryan, C. (2003). Stereotype Accuracy. *European Review of Social Psychology*, 13(1), pp.75-109.

Ryan, C., Park, B. and Judd, C. (1996). Assessing Stereotype Accuracy: Implications for Understanding the Stereotyping Process. In: Macrae, C. Stangor C. and Hewstone, M. (eds.) *Stereotypes and Stereotyping*. New York: The Guilford Press, pp.121-150.

Salerno, J. and Bottoms, B. (2009). Emotional Evidence and Jurors' Judgments: The Promise of Neuroscience for Informing Psychology and Law. *Behavioral Sciences & the Law*, 27(2), pp.273-296.

Sathiyanathan, L. and Xing, L. (2018). *In a city of accents, what's considered 'cute' versus 'foreign' reveals a hierarchy | CBC News*. [online] CBC. Available at: http://www.cbc.ca/news/canada/toronto/the-accent-effect-toronto-1-1.4412854 [Accessed 8 Aug. 2018].

Schkade, D., Sunstein, C. and Hastie, R. (2010). When Deliberation Produces Extremism. *Critical Review*, 22(2-3), pp.227-252.

Seggie, I. (1983). Attribution of Guilt as a Function of Ethnic Accent and Type of Crime. *Journal of Multilingual and Multicultural Development*, 4(2-3), pp.197-206.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3-4), pp.193-229.

Sobral Fernández, J. and Prieto Ederra, A. (1994). Presentación, Discurso y Persuasión en Testigos. *Revista de Psicología Social*, 9(1), pp.13-18.

Stead, Marcus. [MarcusStead]. (2018a, August 5). [Tweet]. Retrieved from https://twitter.com/search?q=marcus%20stead&src=typd.

Stead, Marcus. [MarcusStead]. (2018b, August 5). How to attract tourists to Wales - have all road signs in English only, drop the Welsh signs, just 20% of Welsh people speak it and virtually zero outside Wales do. Also be more welcoming - don't switch to Welsh. [Tweet]. Retrieved from https://twitter.com/MarcusStead/status/1026230816064643072.

Steblay, N., Besirevic, J., Fulero, S. and Jimenez-Lorente, B. (1999). The Effects of Pretrial Publicity on Juror Verdicts: A Meta-Analytic Review. *Law and Human Behavior*, 23(2), pp.219-235.

Stoner, J. (1968). Risky and Cautious Shifts in Group Decisions: The Influence of Widely Held Values. *Journal of Experimental Social Psychology*, 4(4), pp.442-459.

Strongman, K. and Woosley, J. (1967). Stereotyped Reactions to Regional Accents. *British Journal of Social and Clinical Psychology*, 6(3), pp.164-167.

Storey v. Burns International Security Services [2004] No. 03-2246 (United States Court of Appeals, Third Circuit).

Sullivan, G. (2016). In Your Own Words: Investigating voice, Intertextuality, and Credibility of Rachel Jeantel in the George Zimmerman Trial. *Linguistic Society of American Proceedings*, 1, pp.1-15.

Sullivan, G. (2017). *Problematizing Minority Voices: Intertextuality and Ideology in the Court Reporter's Representation of Rachel Jeantel's Voice in the State of Florida v. George Zimmerman Murder Trial*. PhD Thesis. Georgetown University.

Sullivan, J. (1988). *ETA and Basque Nationalism: The Fight for Euskadi, 1890-1986*. London: Routledge.

Tamasi, S. and Antieau, L. (2015). *Language and Linguistic Diversity in the US*. New York: Routledge.

The Florida Bar (2017). *Florida Rules of Criminal Procedure*. Florida Courts.

The National Center for State Courts (2004). *Does Jury Size Matter? A Review of the Literature*. San Francisco: Judicial Council of California/Administrative Office of the Courts.

The New York Times. (2018, June 4). In Facebook [News Page]. Retrieved 8 Aug. 2018, from https://www.facebook.com/nytimes/posts/10151601299644999.

The Secret Barrister (2018). The Ben Stokes trial: what went wrong?. [Blog] *The Secret Barrister*. Available at: https://thesecretbarrister.com/2018/08/16/the-ben-stokes-trial-what-went-wrong/ [Accessed 24 Aug. 2018].

Thomas, E. (2002). Sociophonetic Applications of Speech Perception Experiments. *American Speech*, 77(2), pp.115-147.

Thomas, E. (2008). Rural Southern White Accents. In: E. Schneider, ed., *The Americas and the Caribbean*. New York: Mouton dr Gruyter, pp.87-114.

Thomas v. East Orange Bd. Of Educ. [2014] 998 F.Supp.2d 338 (United States District Court, D. New Jersey).

Thompson, M. (2015). Blackness as Character Evidence. *Michigan Journal of Race and Law*, 20(2), pp.321-348.

Tonry, M. (2001). Punishment Policies and Patterns in Western Countries. In: M. Tonry and R. Frase, (eds.) *Sentencing and Sanctions in Western Countries*. New York: Oxford University Press, pp.3-28.

Trudgill, P. (1999). Standard English: What It Isn't. In: T. Bex and R. Watts, (eds.) *Standard English: The Widening Debate*. London: Routledge, pp.117-128.

U.S. Census Bureau, (2016). *Census Regions and Divisions of the United States*. [image] Available at: http://www.census.gov/econ/census/help/geography/regions_and_divisions.html [Accessed 28 Mar. 2016].

U.S. Equal Employment Opportunity Commission (2016). *Enforcement Guidance on National Origin Discrimination*. Washington D.C.: Office of Legal Counsel.

United States Courts. (2018). *Types of Juries*. [online] Available at: http://www.uscourts.gov/services-forms/jury-service/types-juries [Accessed 22 Jun. 2018].

United States District Courts (n.d.). *Handbook for Trial Jurors Serving in the United States District Courts*. Washington D.C.: Administrative Office of the United States Courts.

Vaissière, J. and Boula de Mareüil, P. (2004). Identifying a Language or an Accent: From Segments to Prosody. In: *MIDL*. [online] Paris: Hal. Available at: https://halshs.archives-ouvertes.fr/halshs-00185539 [Accessed 17 Aug. 2018].

Wadham, J., Robinson, A., Ruebain, D. and Uppal, S. (2010). *Blackstone's Guide to the Equality Act 2010*. Oxford: Oxford University Press.

Wales, K. (2006). *Northern English: A Social and Cultural History*. Cambridge: Cambridge University Press.

Watt, D. (2018). Applied Dialectology: Dialect, Coaching, Dialect Reduction, and Forensic Phonetics. In: C. Boberg, J. Nerbonne and D. Watt, (eds.) *The Handbook of Dialectology*. Oxford: John Wiley & Sons, Inc., pp.219-232.

Watt, D. and Tillotson, J. (2001). A Spectrographic Analysis of Vowel Fronting in Bradford English. *English World-Wide*, 22(2), pp.269-303.

Watts, R. (2000). Mythical Strands in the Ideology of Prescriptivism. In: Wright, L. *The Development Of Standard English, 1300-1800*. Cambridge: Cambridge University Press, pp.29-48.

Webber, C. (2014). Teesside TV presenter Steph McGovern sent £20 to tone down Teesside accent. *The Northern Echo*. [online] Available at: https://www.thenorthernecho.co.uk/news/local/teesvalley/middlesbrough/116 24872.Teesside_TV_presenter_Steph_McGovern_sent___20_to_tone_down _Teesside_accent/ [Accessed 14 Nov. 2018].

Weiss, D. (2008). *Sentence Not Lost in Translation: Judge Orders Four to Learn English or Go to Jail*. [online] Times Leader. Available at: https://www.timesleader.com/archive/270564/stories-sentence-not-lost-in-translation104476 [Accessed 10 Aug. 2018].

Werner, C., Strube, M., Cole, A. and Kagehiro, D. (1985). The Impact of Case Characteristics and Prior Jury Experience on Jury Verdicts1. *Journal of Applied Social Psychology*, 15(7), pp.409-427.

Wiehl, L. (2002). Sounding Black in the Courtroom: Court-Sanctioned Racial Stereotyping. *Harvard Blackletter Law Journal*, 18, pp.185-210.

Williams, D. and Escudero, P. (2014). A Cross-Dialectal Acoustic Comparison of Vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America*, 136(5), pp.2751-2761.

Williams v. Frank [1991] 757 F. Supp. 112. (United States District Court, D. Massachusetts).

Wolfram, W. (1998). Language Ideology and Dialect: Understand the Oakland Ebonics Controversy. *Journal of English Linguistics*, 26(2), pp.108-121.

Wolfram, W. and Schilling-Estes, N. (2016). *American English: Dialects and Variation*. 3rd ed. Malden, Mass.: John Wiley & Sons, Inc.

Woolard, K. (1992). Language Ideology. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 2(3), pp.235-249.

Wuensch, K. and Moore, C. (2004). Effects of Physical Attractiveness on Evaluations of a Male Employee's Allegation of Sexual Harassment by His Female Employer. *The Journal of Social Psychology*, 144(2), pp.207-217.

Wuensch, K., Campbell, M., Kesler, F. and Moore, C. (2002). Racial Bias in Decisions Made by Mock Jurors Evaluating a Case of Sexual Harassment. *The Journal of Social Psychology*, 142(5), pp.587-600.

Yardi, S. and Boyd, D. (2010). Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society*, 30(5), pp.316-327.

Zimmerman v. State [2013] No. 5D13-1233 (United States District Court, Florida).

Zimring, F. (2003). *The Contradictions of American Capital Punishment*. New York: Oxford University Press.