Reclaiming Colourful Language: The Impact of Linguistic Reclamation on the Cognitive Processing of LGBTQ+ Slurs

Daniel Edmondson

Thesis submitted for the degree of Doctor of Philosophy

School of English

University of Nottingham

May 2022

Abstract

This interdisciplinary psycholinguistic and sociolinguistic thesis explores the cognitive processing of slurs, in particular those targeting the LGBTQ+ community. It also seeks to determine whether the phenomenon of linguistic reclamation has any effects on the processing of slurs.
Existing research on the processing of taboo language more broadly has identified that taboo words are accessed from the mental lexicon more slowly than non-taboo words, yet are better remembered than non-taboo language after they are encountered. This body of research has also suggested that among other factors, degree of emotional arousal and perceptions of a word's tabooness are key modulators of both of these effects. However, until now, research into taboo language processing has not at all considered whether these effects are produced for slur words as a specific category of taboo language, nor whether linguistic reclamation might alter arousal responses and tabooness judgements, to the extent that expected taboo language processing effects are no longer produced.

The research consisted of five separate studies. The first two studies were preliminary surveys used to identify the most recognisable LGBTQ+ slurs in contemporary British English, gain insights into their use and reclamation, and produce a database of normed emotional, non-emotional, and reclamation behaviour ratings for these words. Two lab experiments followed, in which the lexical accessibility and recall likelihood of LGBTQ+ slurs was compared both to non-taboo negative words, and other taboo words which were not slurs. Comparisons were also made based on participant demographic data, and the normed ratings were used to model explanations of the results. The final study was a pupillometry study which sought to directly measure an impact of linguistic reclamation on physiological arousal responses. Data collection for this final study had to be postponed due to the COVID-19 pandemic, but the design and potential outcomes of the study are comprehensively detailed within this thesis.

I found no overall differences between LGBTQ+ slurs and either the taboo or non-taboo word categories in terms of their lexical accessibility. For recallability, LGBTQ+ slurs – like other types of taboo language – were significantly likelier to be recalled than non-taboo words, but were not significantly likelier to be recalled than other examples of taboo language. However, I did find a difference in recall likelihood between LGBTQ+ slurs and taboo words which were not slurs, exclusively among non-heterosexual participants. I also found that participant gender identity and sexual identity were the best overall predictors of lexical accessibility in my experiments, while word length in characters, familiarity, negativity and tabooness were the best predictors of recall likelihood, depending on the specific comparisons being made. Most importantly, I found that in all of my experiments, at least one measure of linguistic reclamation significantly predicted both lexical accessibility and recall likelihood, with increased reclamation associated with better performance in both kinds of processing task.

My findings make significant original contributions to both the psycholinguistic and sociolinguistic study of taboo language, slurs, and linguistic reclamation. The research offers comprehensive normed data on LGBTQ+ slurs, quantifiable measures of their reclamation, as well as the first dedicated exploration of the cognitive processing of slurs and the impact of a sociolinguistic phenomenon on the same. To sociolinguistics, the research provides a new and original evidence base for contemporary theory and perspectives on slurs and their reclamation, as well as an indication that participation in a fundamentally political language project can transform both conscious *and* unconscious responses to language embedded within issues of systemic power, social marginalisation, and hegemony.

Acknowledgements

As thanks, I would like to make the following acknowledgements:

To Gerald Hayes, the cleverest person I've ever known, who I wish so badly could have read this.
To Dr Dominic Thompson and Dr Lucy Jones, I couldn't have asked for a better supervisory team.
To David Edmondson and Christine Hayes, for their constant love and support.
To my friends in Nottingham, Manchester, Cornwall, and Reading, for always keeping me going.
To my fellow PGR students at the University of Nottingham, for all the coffees and mutual support.

I couldn't have done this without all of you.

Table of Contents

List of Tables

List of Figures

Abbreviations Glossary

ANOVA – Analysis of Variance

AoA – Age of Acquisition

BNC – British National Corpus

CoP – Community of Practice

LGBTQ+ – Lesbian, Gay, Bisexual, Transgender, Queer, and other non-cisgender/non-heterosexual identities.

ROI – Region of Interest

RSVP – Rapid Serial Visual Presentation

SCR – Skin-Conductivity Responses

Chapter 1 – Introduction

Many current debates surrounding identity and marginalisation revolve around the use of language, particularly that which is considered offensive and derogatory. Such questions include what kinds of language should be considered *slurs* against particular groups; how slurs do and should make individuals feel; and if, when, and by whom it is acceptable to use them. Since the 1980s, a body of psycholinguistic research has been developing which focuses on the cognitive processing of taboo stimuli, particularly of taboo language, with slurs included in this category. More recently, sociolinguistic research has taken an interest in the positive reappropriation of taboo slurs, a process often referred to as linguistic reclamation. However, these two lines of enquiry have not yet been combined. Accordingly, the impact of linguistic reclamation on the cognitive language processing of slurs is the subject of my research, as set out in this thesis. In this introductory chapter, I begin by defining key terms and outlining the three main research questions which my studies seek to answer. I then explain why I believe these are important interdisciplinary questions to ask, before outlining some of the theoretical and methodological challenges of researching this topic. Finally, I provide an overview of the structure of this thesis, as a piece of combined psycholinguistic and sociolinguistic work.

1.1. Defining Key Terms

Before I outline my research questions, I begin by providing some concise definitions of the terms they will contain, which I will use throughout my thesis. More comprehensive discussions of these concepts – including how I have arrived at the definitions given in this section – can be found in my review of the existing literature (Chapter 2).

I take *taboo language* to refer to words or propositions with meanings that are recognised by native speakers to be socially prohibited "for a specific community…at a specifiable time in specifiable contexts" (Allan & Burridge, 2006, p. 11), typically because they are "perceived to be a potential cause of discomfort, harm, or injury" (Allan, 2019, p. 3). I discuss the problems with and approaches to defining taboo language throughout section 2.1. of my literature review.

Slurs are a particular class of taboo language, used by more hegemonically powerful (Gramsci 1971; 1985) social groups to derogate (Chen, 1998; Croom, 2011) members of less hegemonically powerful social groups (Herbert, 2015). I discuss contemporary understandings of slurs in section 2.2.1. of my literature review.

Linguistic reclamation describes a sociolinguistic phenomenon particular to slurs. It refers to positive and non-derogatory uses, by members of marginalised groups, of the slurs used by members of hegemonically powerful groups to derogate them (Bianchi, 2014; Herbert, 2015). The reclamation of any individual slur can be regarded as a political language project (Herbert, 2015), although reclamation projects do not tend to be entered into as coordinated political efforts, they are not equally participated in by all members of the group(s) able to do so, and there is no consensus among those groups as to their appropriateness or effectiveness (Brontsema, 2004). I discuss linguistic reclamation in sections 2.2.2. to 2.2.4. of my literature review.

Regarding both slurs and linguistic reclamation, I have chosen to focus my own research on slurs targeting the *LGBTQ+* (Lesbian, Gay, Bisexual, Transgender, Queer, and other non-cisgender/non-heterosexual identities) community, as it has a rich history of navigating oppression and marginalisation through its use of language, which includes subjecting a number of LGBTQ+ slurs to processes of reclamation (Baker, 2008; Coleman-Fountain, 2014; Russell, Clarke, & Clary, 2009;

White, 1980). It is also a community to which I belong, and accordingly possess a personal experience of and epistemological standpoint on the reclamation of its slurs (for a discussion of standpoint theory see section 2.2.4.; Bar-On, 1993; Hankinson-Nelson, 1993).

Finally, wherever I mention *cognitive language processing*, I am referring to the "mental processes…involved in producing and understanding language…operating largely independently of other cognitive systems" (Traxler, 2012, p. 27).

1.2. Research Questions

My research asks three primary questions:

1. Is the cognitive language processing of LGBTQ+ slurs similar to that of other kinds of taboo language?
2. Does the sociolinguistic phenomenon of linguistic reclamation impact the cognitive language processing of LGBTQ+ slurs?
3. What are the sociolinguistic implications of any effects of reclamation on the cognitive language processing of LGBTQ+ slurs?

While these are the main questions that my research seeks to answer, these are not the only insights that it provides. Throughout my thesis, I also discuss what my findings reveal about LGBTQ+ slurs themselves, about linguistic reclamation as a sociolinguistic phenomenon, and about the broader social contexts underpinning both LGBTQ+ slurs and their reclamation. I will now discuss why I consider all of these to be important and valuable questions to ask.

1.3. Motivations for the Research

My first reason for asking these questions is that – as a consequence of investigating the general effects of tabooness on language processing – psycholinguistic studies have not typically focused on any one type of taboo language. However, being informed by sociolinguistic perspectives on the role language plays in replicating and reinforcing existing social power dynamics, I argue there is compelling reason for taboo language processing studies to focus specifically on slurs. As I alluded to in section 1.1. and discuss in depth in section 2.2.1. of my literature review, slurs exist specifically to cause harm, operating along axes of social marginalisation, exclusion, and oppression. They are more capable of this than any other kind of taboo language, to the extent that some recent research has identified the utterance of slurs as a type of speech act which actively denies rights to an individual or group (Herbert, 2015; McGowan, 2012). If – as I am convinced – slurs are an especially harmful form of taboo language, then understanding the unconscious and affective responses they generate is a much-needed application of taboo language research.

As a consequence of having not focused specifically on slurs, no taboo language processing research has accounted for linguistic reclamation as a phenomenon particular to slurs. Throughout section 2.3. of my literature review, I emphasise the stress that the literature places on the cognitive effects of taboo language being a consequence of its high degree of negativity, shock value, and perceived social unacceptability. As my definition of linguistic reclamation in section 1.1. suggests, it is reasonable to hypothesise that linguistic reclamation might subvert all three of these perceptions of slurs, and by extension slurs may not exhibit the processing effects associated with taboo language when reclaimed. If this is true, then existing studies have not accounted for a factor which

could significantly affect responses to a number of their stimuli, which raises important potential validity concerns for some previous research.

However, I argue that one of the most important motivations for exploring linguistic reclamation in relation to taboo language processing is not which findings from previous studies can be further verified by examining reclamation, but what an entirely new language processing study could tell us about reclamation. To date, studies of linguistic reclamation have been situated firmly within sociolinguistics, and have been the subject of predominantly theoretical and political discussions. Where sociolinguistic reclamation studies have included a data analysis element, these have been largely discursive analyses focusing on conscious attitudes towards participation in reclamation projects. This analysis is extremely important to understanding the driving forces behind reclamation as a sociolinguistic phenomenon, but I am not convinced it is the sole motivator of engagement in this process, and the effects of reclamation on unconscious, affective responses to slurs themselves remains unexplored. An entire area of psycholinguistic study is devoted to the identification of emotional responses to language, uniquely suited to exploring this issue.

I therefore argue that interdisciplinary psycholinguistic and sociolinguistic research such as my own – which combines investigation of taboo language processing with the study of linguistic reclamation – will be able to reveal entirely new effects of linguistic reclamation which are critical to understanding the nature and impact of the phenomenon. As I suggested at the very start of this chapter, this in turn can inform public debates about how oppressive language affects us emotionally, and how it may be possible for those targeted by slurs to effect positive changes in this regard.

Having now discussed the importance of this interdisciplinarity, I now discuss some of the challenges that arose from the outset in attempting to combine psycholinguistic theory and methods with sociolinguistic theory, and how I address these challenges throughout my thesis.

1.4. Challenges Combining Psycholinguistic and Sociolinguistic Research

There is a dearth of research combining the fields of psycholinguistics and sociolinguistics, and in some respects the two might be considered difficult to align. Psycholinguistic experimentation often requires the control or outright elimination of variability arising from highly personal or contextual uses of language (Dörnyei, 2007) while much contemporary sociolinguistic research – especially regarding language and identities – explicitly resists and criticises the kind of strict categorisations and generalisations often involved in quantitative, experimental studies (for which I outline reasons later in this section). Because the topic of my research centres both psycholinguistic and sociolinguistic phenomena, my research exposed this interdisciplinary conflict, and it was essential for me to find a means of resolving it.

In particular, I considered it important to conduct my psycholinguistic studies in a way which allowed for rigorous experimental control and the identification of general patterns in language processing within particular groups, but avoided the kind of *essentialism* that has been heavily criticised in sociolinguistic work on identity. Essentialism refers to the theoretical position that particular sociocultural and/or biological characteristics can be used to categorise individuals into certain social groups, and that such categorisations can be used to explain attributes and behaviours of all members of those groups (Bucholtz, 2003).

There is some temptation to take an essentialist approach when conducting experimental research, most often implicitly through the way that participants are grouped and data is analysed. This is because such a perspective supports viewing identities as clearly delineated independent variables, exhibiting consistent behaviours and explanations for those behaviours (Bucholtz, 2003). This assumption is convenient to measuring the impact of such variables, and in turn for isolating

any measurable impact of the dependent variable(s) under investigation (Dörnyei, 2007; Johnson & Christensen, 2004).

However, an entirely essentialist approach faces three significant criticisms. The first is theoretical: essentialism clashes with Butler's (1990) position – commonly accepted in contemporary language and identity research – that identity is not something which people *have* but something which people *do*, and thereby essentialism "reduces the diversity of humanity to a small set of attributes and behaviours recognised by the theory" (Bucholtz, 2003, p. 400). The second criticism is methodological: it may cause a researcher to falsely impose their own categories and/or assumptions onto participants. While this may suit the purposes of their research, it can have a detrimental effect on the credibility or representativeness of the study, and of the conclusions ultimately drawn from it (Cameron, 1990; Romaine, 1984). The final critique is political, in that strict essentialist categorisation has serious potential for disempowerment. For one, the behaviours identified as uniformly belonging to certain groups can be based on harmful or reductive assumptions about those groups. Furthermore, essentialist categorisation can exclude individuals from groups with which they may identify, on the basis that they do not meet potentially false prerequisites imposed by the researcher(s) (Bucholtz, 2003).

To be clear, I do not intend to misrepresent psycholinguistic, experimental, or quantitative research by presenting essentialism as an inevitable consequence of using such methods. However, it is harder to avoid these criticisms when categorising a large sample of participants based on identity is a necessary part of the methodology, as can be the case in psycholinguistic experimentation. This is sometimes ill-considered in psycholinguistic research, leading to questionable hypotheses and conclusions in otherwise credible and informative research, for example concerning men's and women's collective responses to words with sexual versus romantic meaning (Geer & Bellard, 1996). On the other hand, much sociolinguistic research has dealt with the criticisms it has identified by avoiding generalisation as much as possible, focusing instead on individual speakers and discourses. It is from these issues – and the lack of clear attempts to resolve them on both sides – that I consider an interdisciplinary tension to arise.

To find a solution, I looked to *variationist sociolinguistics*, a branch of sociolinguistics concerned with the identification and social evaluation of variations in language use (Chambers, 2004), and typically employs quantitative methods to do so, including controlled experiments (Mendoza-Denton, 2004). I identified two particular perspectives from this subdiscipline which I considered applicable to my work. First, that the problem could not be fixed – as some sociolinguistic perspectives attest – by a simple change in methodological approach, as changing methodology also changes the object of investigation (Mendoza-Denton, 2004). I could no longer claim to be investigating effects of linguistic reclamation on unconscious language processing if I did so using non-experimental methods, as I would then be investigating conscious responses and attitudes to reclamation and its effects instead. Therefore, I needed to focus on how psycholinguistic experiments could be suitably adapted to avoid the problems of essentialism as much as possible. The second observation I made suggested a means of achieving this: Potter (2000) and Schegloff (1997) recommend that researchers refer to participants' explicit orientations to their own identities when forming groups for quantitative sociolinguistic analysis. I therefore resolved to shift away from presuming the presence of particular identity groups in my data and the association of particular linguistic behaviours with those groups, and move towards allowing participants to identify themselves and their own linguistic behaviours and formulate groups for analysis accordingly. Across all of the studies I describe in my thesis, I indicate wherever I have taken steps reflecting this interdisciplinary approach. With regard to these studies, and to the rest of the thesis in general, I now move onto describing its overall structure.

1.5. Structure of the Thesis

As a piece of interdisciplinary psycholinguistic and sociolinguistic research, the structure of my thesis combines the typical structures of theses from both disciplines. It begins with a review of the existing literature (Chapter 2), having two primary objectives. First, to discuss and expand upon the concepts defined in section 1.1., all of which are essential to understanding and conducting my research. Second, to identify what previous studies have revealed about the cognitive processing of taboo language, focusing on two aspects in particular: word recallability and lexical accessibility. Because no taboo language processing studies have previously explored any potential effects of linguistic reclamation, I also draw on findings from across the literature to establish why such effects might exist, and what they may be. This chapter is intended to provide an in-depth discussion of the issues most relevant to the research as a whole, not to foreground any individual study described in later chapters. Instead, in the style of a psycholinguistic thesis, each study chapter includes a research background section, discussing the literature which directly informed its hypotheses and design.

There is no dedicated methodology chapter following the literature review, as might be expected from a purely sociolinguistic thesis. Instead, Chapters 3 to 7 of my thesis follow the common structure of a psycholinguistic thesis, separately reporting the respective aims, methods, and findings of the five studies I carried out as part of my research. Though independent of each other, each study follows a logical progression from the one before it, and all integrate lines of sociolinguistic discussion wherever they relate to my findings.

In Chapter 3, I describe the first study I conducted as part of my research: a survey to identify LGBTQ+ slurs in contemporary British English, their respective frequencies, and more detailed information about their use. As mentioned in section 1.3., LGBTQ+ slurs are underrepresented in psycholinguistic studies of taboo language, so I needed to conduct my own research to identify slurs for use in my later studies.

In Chapter 4, I describe the second survey study that I conducted, which I designed to produce a database of normed word property and reclamation behaviour ratings, for the most frequently identified slurs to emerge from the study described in Chapter 3. This was so that I could use this data to investigate the potential effects of these properties and reclamation behaviours in my slur processing experiments.

Chapters 5 and 6 describe a pair of related lab experiments I conducted to investigate the cognitive processing of LGBTQ+ slurs. Both of these experiments combined a lexical decision task (to gather data on lexical access) with a free recall task (to gather data on word recall). The difference was in the stimuli being compared: the experiment I describe in Chapter 5 compared LGBTQ+ slurs to two categories of non-taboo language, while the experiment I describe in Chapter 6 compared the same slurs to other examples of taboo language which were not slurs.

In Chapter 7, I describe a pupillometry study I designed to explore the impact of linguistic reclamation on emotional arousal responses, using pupil dilation as a physiological rather than self-reported measure of arousal, as I had used in Chapters 5 and 6. Data collection for this study was interrupted due to the Covid-19 pandemic. However, in this chapter I provide a thorough justification for using this type of experimental paradigm to explore the impact of linguistic reclamation on arousal; a detailed account of my experiment design and procedure; an overview of potential outcomes that may arise from the study once completed; and suggestions as to how this study could help confirm and clarify the results of my studies described in Chapters 4 to 6.

In Chapter 8, I provide a general discussion, bringing together all of the findings from Chapters 2 to 7 in order to answer the research questions introduced in section 1.2. Here, I particularly focus on findings which manifested across all of my studies, including (but not limited to) a relationship between slur processing and identity groups relating to those slurs; other language properties affecting slur processing; and relationships between particular language properties, specific kinds of reclamation behaviour, and the processing of slurs. I also discuss the significance of these largely psycholinguistic findings to the field of sociolinguistics, as well as the limitations of my research and directions for future study.

In Chapter 9, I conclude my thesis by summarising the answers to my three research questions provided by my general discussion, emphasising the significance and original contributions of my research to both psycholinguistics and sociolinguistics, discussing the potential real-world applications and impacts of my findings, and ending with my final reflections on conducting a piece of interdisciplinary psycholinguistic and sociolinguistic research.

Chapter 2 – Literature Review

In this chapter, I provide a review of the existing literature which considers the four key concepts defined in my Introduction in more depth: taboo language, slurs, linguistic reclamation, and the cognitive processing of taboo language. The challenge for my research as a piece of interdisciplinary work is ensuring that these various psycholinguistic and sociolinguistic issues are foregrounded in sufficient detail, as all are of equal importance to every study that I describe in the chapters that follow. For this reason, as I stated in section 1.5. of my introduction (Chapter 1), this literature review is primarily focused on establishing comprehensive definitions and understandings of the issues most central to my research as a whole. Within each of the study chapters that follow, I provide shorter overviews of the literature most relevant to each study in dedicated research background sections.

I begin by setting out in detail my definitions of taboo language and of slurs, and my reasoning for defining them in the way that I have. Having a thorough understanding of these terms is essential for the rest of my research, but as I will explain, the existing literature has tended to offer either incomplete or contested definitions of both terms. Once I have shown how I resolved these issues and provided clearer definitions, I define and discuss linguistic reclamation as a phenomenon particular to slurs, and explain the particular focus of my research on LGBTQ+ slurs in more detail. Finally, I discuss the body of psycholinguistic literature on taboo language processing informing all of my work. Within this section, I use the definitions I establish earlier in my review to discuss the likely processing of slurs as a type of taboo language, and consider how the phenomenon of linguistic reclamation might affect slur processing.

2.1. Defining Taboo Language

2.1.1. Challenges in Defining Taboo Language

I begin by discussing how to define the term *taboo language*. Most people can identify particular words and phrases that they would consider to be taboo, but providing a definition which can encompass all of these is surprisingly difficult. This difficulty is evidenced in the literature, which has often featured a somewhat unreliable approach to defining taboo language.

A number of existing studies have used what I term 'list-based' definitions of taboo language. Jay, Caldwell-Harris, and King (2008, p. 83) define taboo language as "a class of emotionally arousing references with respect to body products, body parts, sexual acts, ethnic or racial insults, profanity, vulgarity, slang, and scatology". Other phrases which appear in similar definitions include 'bodily functions' (e.g. Jay, 1992) and terms like 'obscenity', 'blasphemy', 'name-calling' and 'verbal aggression' (e.g. Jay, 2000). While this approach to defining taboo language is useful for highlighting different types of expression that might be considered taboo, it is problematic for a number of reasons. First, it does not give us a sense of *why* these particular references are taboo, because it does address what makes something taboo to begin with. Second, these subcategories of taboo language are not always clear and discrete; it is rarely specified what distinguishes 'profanity' from 'blasphemy' or 'obscenity', for example. Third, these lists are not (and perhaps cannot be) exhaustive; the inclusion of certain types of reference in list-based definitions are always, to some extent, selective and subjective on the part of the researcher. Taking the example of Jay et al.'s (2008) definition, it is surely not only ethnic and racial insults that are taboo, but also those that are misogynistic, homophobic, transphobic, ableist and classist. Even this

expanded list that I have provided likely has omissions or exceptions. In my view, these problems make list-based definitions generally unhelpful to social scientific study, be it linguistic, psychological, sociological, or a combination of all three. Instead, a definition that can be concise, consistent and widely applicable is necessary, to properly enable comparisons of findings across studies and research contexts. I therefore take a different approach to defining taboo language: one which instead examines the origins of the term *taboo,* and the more contemporary understandings of taboos and taboo language that have developed from this early research.

Brandes (2019) notes that 'taboo' first entered the English language following Captain Cook's 1777 visit to Tonga. He explains that the word is derived from the terms 'kapu', 'tabu', and 'tapu', which occur across various Polynesian languages to refer to things which are sacred or holy, and are therefore "of prohibited or restricted use within society" (Brandes, 2019, p. 374). Within anthropology, taboo has come to refer to "prohibitions that are applied in specific social circumstances or to particular segments of society", with violations of these rules being met by some form of social sanction (Brandes, 2019, p. 374). Allan (2019, p. 3) expands on this, noting that taboos are particularly applied to behaviours which are "perceived to be a potential cause of discomfort, harm, or injury". Though my research focuses on taboo language, taboos can be applied to a wide range of behaviours, many (if not most) of which are non-linguistic. For example, one of Cook's first recorded uses of the term was in reference to two of his dinner guests, who were forbidden from touching food on account of both having recently washed the bodies of the dead (Cook, 1967). This taboo on handling food after preparing bodies for burial not only served as a mark of respect, but also as a means of avoiding harms – both real (i.e. disease) and superstitious – that breaking this rule might bring upon the living. The whole range of taboo behaviours comes with an equally wide range of potential punishments, from the mild disapproval aimed at those eating with their hands at a sit-down restaurant in the United Kingdom, to the death penalty still given to those found guilty of extramarital affairs in several countries (Allan, 2019). Appreciating the sheer range of behaviours that can be tabooed is important to studying taboo language, because this reminds us that linguistic taboos are actually a relatively new and underexplored subject in the sociological and anthropological study of taboo behaviours.

Taboos of all types can also be extremely localised, and can seem perplexing even to members of the same broader society if they are not part of the specific group in which a taboo is enforced. Henningsen (1957) observed tabooing of all spoken references to boats and fishing among residents of a single Scandinavian fishing village, for fear of invoking failure or accident when out on the water. Similarly, theatre-goers in both the United Kingdom and the United States might be familiar with the practice of avoiding wishing good luck to an actor before they go on stage, instead appearing to wish harm, as in the case of the phrase "break a leg". This arises from the superstition that wishing good luck actually brings bad luck upon a performance, and vice versa (Dundes, 1994; Huggett, 1975; Spears, 2000). The use of the phrase has therefore served to avoid breaking a very specific taboo. While the expression itself is now widely encountered as a good luck wish outside of the theatre, the taboo it protects against is not, and would not be recognised in almost any other English-speaking context. Indeed, this particular taboo may now be considered excessively superstitious even in the context in which it has traditionally been observed.

This discussion points to a definition of taboos (and by extension taboo language) which centres on community membership and context, an approach which is becoming increasingly common in contemporary research. Allan and Burridge (2006, p. 11), for example, define a taboo as "a proscription of behaviour for a specific community of one or more persons at a specifiable time in

specifiable contexts". I add to this Allan's (2019) more recent observation that such proscriptions apply specifically to behaviours which are felt to cause some kind of personal or social harm. Initially, such a definition may seem too broad or vague. However, I argue that the three factors it highlights – community, time and context – profoundly affect the perception of behaviours as being taboo, and that this is especially true for linguistic taboos.

On this basis, I define taboo language as words and propositions with meanings that are recognised by native speakers to be prohibited for specific communities, in specifiable times and contexts, because they are felt to cause personal or social harm. The rest of this section of my review focuses on unpacking the individual aspects of this definition.

2.1.2. Relevance of Specific Communities to Defining Taboo Language

To interrogate Allan and Burridge's (2006) use of the term *community*, I refer to the *social identity theory* developed by social psychologists Tajfel and Turner (1986). This introduced the concepts of *ingroups* and *outgroups*, explaining that when individuals perceive themselves or others as belonging to a particular group, be it as a member of a sports team (e.g. Lalonde, 1992) or a broader socially-defined group like the LGBTQ+ community (Walters, Paterson, McDonnell, & Brown, 2019), they ascribe to themselves or others an ingroup identity. This identity is positioned against members of a contrasting outgroup, i.e. anyone perceived not to be a part of that group. As these examples make clear, both ingroups and outgroups can vary greatly in size, but it is generally agreed that they comprise more than one person.

Tajfel and Turner (1986) also observed that the distinctions made between ingroups and outgroups necessarily imply the existence of *intergroup boundaries* – parameters which determine who belongs to the ingroup and who does not. Giles and Giles (2013, p. 146) describe these as being "symbolically equivalent to geographical borders, yet…reflected in more psychological and communicative dimensions". So as not to simply reify existing (and perhaps harmful) social divisions, it is important to recognise that intergroup boundaries are just as subjective as the perception of ingroup and outgroup membership itself (Giles & Giles, 2013). Furthermore, it should be acknowledged that intergroup boundaries are not fixed, and can be permeated or destabilised. Regarding in/outgroups and linguistic nationalism, for example, Kim (2001) notes that a bicultural native speaker of one language may be a nativelike speaker of another group's language. Such an individual could disrupt nationalistic ingroup and outgroup distinctions, insofar as the language(s) a person speaks acts as an intergroup boundary between both. Despite the existence of such cases, Giles and Giles (2013) point out that some ingroup members may feel that the characteristics and customs defining their ingroup are unknowable to anyone outside of it. This is a problematic sentiment, in that it can lead to perceptions of others having inauthentic or invalid community membership, not necessarily based on any ingroup consensus or genuinely shared experience.

There are two ways that this understanding of communities and group identities can immediately be related to the formation and function of taboos. First, as in the cases of Henningsen's (1957) fishing village or the superstitions of the theatre to which I referred in section 2.1.1., recognition of and adherence to shared taboos can be a particularly salient indicator of ingroup membership (Allan, 2019). Second, by virtue of being able to signal ingroup membership, taboos can also be used to police its boundaries. Because a key component of taboos is that breaking them comes with social repercussions (Allan, 2019; Brandes, 2019), they can also be considered a very effective means of maintaining distinctions between who is labelled 'in' and who

is labelled 'out'. Swift consignment to the perceived outgroup therefore becomes another social consequence of breaking the ingroup taboo.

It is in this regard that social identity theory also illuminates why certain behaviours and associated identities become tabooed, and how this relates to wider social power structures. While subjective perceptions of in/outgroup membership and intergroup boundaries are difficult to generalise, both are closely related to another important concept within social identity theory: the *intergroup status hierarchy*. Giles and Giles (2013, p. 147) explain this as the social privileging of certain groups over others, noting that "there is often a multicultural consensus about the rank ordering of social groups in a particular society". Critically, they describe an inverse correlation between a group's position in the intergroup status hierarchy and the degree to which behaviours associated with that group are tabooed: the most highly tabooed behaviours are associated with those lowest in the hierarchy. Sex workers, for example, experience an intense degree of social stigmatisation and exclusion, and the practice of sex work is highly tabooed. The direction of causality in such relationships – whether the behaviour is taboo because of those it is associated with, or vice versa – can be difficult to determine and can vary between taboos. However, it is critical that such relationships exist, centring issues of community exclusion so prominently.

This is particularly important to the study of linguistic taboos, because relationships between particular taboos and issues of social power are often evidenced in language, and indeed are the source of many examples of taboo language. To return to the example of sex work(ers), references to the same serve cross-linguistically as taboo phrases and insults: English 'whore', French 'putain', Spanish 'puta', Russian 'шлюха' etc. A number of sociolinguistic studies point to this as evidence of cross-national disapproval of sex work(ers) and (given the typical gendering of such terms) of the sexuality and sexual autonomy of women (Cameron, 1992; Schulz, 1975; Stanley, 1977). Indeed, the strength of this taboo is such that these words are still taboo when used beyond their denotative meaning; 'whore' can function as a highly taboo insult regardless of whether its target is a sex worker, because the comparison is so negatively perceived.

With regard to these examples, I note that many of the examples given by researchers of taboo language targeting some aspect of a person's identity have not typically focused on LGBTQ+ identity. This will become clear from many of the examples I reference throughout this review, which are often racial, ethnic, or misogynistic expressions. These examples are still highly valuable for illustrating issues relating to taboo language generally, which very much apply to taboo language referencing the LGBTQ+ community. However, they also highlight the relative lack of attention that has been given to taboo language referencing LGBTQ+ identities in the literature, leaving fewer examples to review. However, in section 2.2.4., I explicitly identify some examples of LGBTQ+ slurs that have been researched in the literature, and discuss their history and use.

It should now be clear the connection I am highlighting between taboos and communities that are particularly subjected to social exclusion and marginalisation, as this is central to the topics of my research. Accordingly, the concept of the intergroup status hierarchy is one to which I frequently return, particularly in section 2.2. where I discuss the nature of slurs, LGBTQ+ slurs specifically, and how the taboo language arising from this kind of structural oppression can be subverted by its targets.

2.1.3. Relevance of Specifiable Times and Contexts to Defining Taboo Language

Compared to more abstract concepts like community, explaining the relevance of social change over time to what is considered taboo is relatively straightforward. The part of this section in which I

discuss specifiable times is therefore quite brief compared to my discussion of factors affecting interactional context, but is nevertheless important to understanding how perceptions of taboos can change. Understanding temporal changes is also important for understanding slurs and linguistic reclamation, as I discuss in section 2.2.

2.1.3.1. Specifiable Times: Temporal Shifts in Perceptions of Taboos

In much the same way that communities and societies do not stay the same even decade to decade, the taboos belonging to these groups are subject to similar changes over time (Allan, 2019). A particularly interesting case is that of 'cunt' in English, described by Hughes (2006, p. 110) as "the most seriously taboo word in English for centuries, remaining so for the vast majority of users". In a detailed case study exploring the comprehensive history of the word, Burridge and Benczes (2019) track several shifts in the use of 'cunt' over the centuries, from occurring frequently and unremarkably in Middle English medical texts and placenames, to being strikingly absent from Samuel Johnson's *Dictionary of the English Language* (1755), as well as censured and vaguely defined in Frances Grose's *Dictionary of the Vulgar Tongue* (1785). Interestingly, Hughes (2006) suggests that the increasing tabooness of 'cunt' over time may have been so pronounced as to have contributed to the decline of unrelated but similar-sounding words, most notably the replacement of 'count' for the French borrowing 'earl' in contemporary English, with its equivalent for women ('countess') being distinct enough to survive.

2.1.3.2. Specifiable Contexts: Intended Meaning, Received Meaning, and Speech Acts

By contrast, the issue of how taboos are judged within the context of their use is more complicated. Culpeper (2019, p. 29) emphasises the importance of context to linguistic taboos when he suggests that "language is taboo when it conflicts with what people expect in a particular context, or what they desire or think should be the case". However, he acknowledges that this statement does not go far enough in distinguishing language which is taboo from language which is simply impolite. Expressing distaste when given a gift is unlikely to be expected or desired, but this does not mean that the language used to do this is taboo.

Indeed, Culpeper describes as a relative absence of taboo language discussion from linguistic studies of politeness. Even in Brown and Levinson (1987) – a seminal publication in language and politeness research – only passing reference is made to taboo language, as an example of language which threatens *positive face*: "the positive identity you claim for yourself [which] is instantiated by how others treat you" (Culpeper, 2019, p. 32). Brown and Levinson argued that taboo language specifically threatens positive face by indicating that a particular speaker doesn't value or fear the same things as a particular hearer. With regard to Allan's (2019) argument that shared taboos are strong indicators of group membership, the use of taboo language to threaten positive face is therefore connected to the use of taboo language as an indicator of ingroup/outgroup membership.

This points to a contextual factor that is critical to considering language in use: speaker and listener perspectives, or meaning as it can be both intended and received in communication. This is something that more contemporary writing on taboo language has also stressed: "who is speaking and what they are talking about makes a big difference. The nature of the speaker-listener relationship, the social and physical setting, the topic of discussion, and the intended meaning of the message must be considered before we comprehend what taboo words mean" (Jay, 2019b, p. 87).

In common debates surrounding taboo and offensive language, a speaker's intended meaning is often positioned as the most important contextual factor determining whether a particular use should be accepted or not. There are certainly cases where this might be true. Politeness research also tells us that face-threatening acts like taboo language use (see Brown & Levinson, 1987) can be used to demonstrate the strength of a relationship – i.e. that the interlocutors are sufficiently close for such an act not to harm a speaker's face. When speaker-listener relationship is strong enough, then, taboo language can be used for entirely positive acts like showing affection, being humorous, and signalling a positive relationship (Jay, 2019b; Wald, 2019). Additionally, some research has demonstrated that taboo language use can signal (and provide relief from) emotions like anger and frustration (Averill, 1983; Jay, 1992; Jay, 2000), with some studies even evidencing that the act of swearing may help individuals cope with physical pain (e.g. Stephens, Atkins, & Kingston, 2009; Stephens & Umland, 2011). Even though the feelings being expressed are negative, the breaking of the linguistic taboo is meant to attract empathy and gain comfort, not to cause harm. However important it is, though, speaker intent does not determine listener reception to taboo language use. This is an equally important factor, and is equally determined by context both within and beyond a specific interaction in which taboo language is used.

To help us understand this, *speech act theory* (Austin, 1962) explains that what an utterance means and achieves in discourse operates on three different levels: *locutionary*, *illocutionary* and *perlocutionary*. The locutionary act concerns the phonological and syntactic arrangement of an utterance, as well as the denotative meanings of the words it contains. Intent becomes relevant to meaning and effect at the level of the illocutionary act: how an utterance is both intended and expected to be received in the context in which it is being used (Holtgraves, 2002). For example, when someone says "I am sorry", it is more often interpreted that they are signalling an apology, rather than declaring that they are experiencing ("I am") a feeling of regret ("sorry"). The perlocutionary act – which importantly the speaker does not control – concerns the effect(s) that an utterance has on its listener (Holtgraves, 2002). Using the "I am sorry" example, the listener may recognise that the speaker is sincerely apologising and express relief or forgiveness. Alternatively, they may not recognise the utterance as a sincere apology, and it may fail to achieve its intended outcome.

Speech act theory offers a useful framework for explaining some of the complexities surrounding the violation of linguistic taboos. Some scholars, such as Leech (2014, p. 229), have argued that "swearwords and other taboo items are not specialised to particular illocutionary acts…", arguing instead that they serve as "general-purpose emotional aggravators". While I agree that taboo items are capable of performing multiple illocutionary acts, I disagree with the assessment that they should be thought of as "general-purpose". Indeed, some taboo language research (e.g. Jay & Danks, 1977; Jay 2019b) argues that the connotative meaning of taboo language (at the illocutionary level) dominates its denotative meaning (at the locutionary level), and therefore the illocutionary force of taboo language is actually essential to understanding its meaning. For example, when someone is called 'bastard', a statement is not usually being made about the perceived legitimacy of their birth, nor the marital status of their parents. Here, it is the illocutionary force of the taboo expression making the most significant contribution to its intended meaning. This example also highlights the use of taboo language for the specific illocutionary acts of labelling, insulting and/or othering, which I revisit in section 2.2.3.

The concept of the perlocutionary act is also key, because it tells us that while a speaker's illocutionary motivation for using taboo words is important, it does not necessarily determine how

such language is received by others. Furthermore, the understanding that speakers can be unsuccessful in producing the desired perlocutionary act(s) is particularly important in the case of taboo language, because taboos are by nature linked to social repercussions. Emphasising this, some taboo language scholars have suggested that while many taboos (including linguistic ones) can be observed and shared across societies, the proscription of particular words and phrases is often self-imposed by individuals (Crespo-Fernández, 2019). Burridge (2010, p. 3) summarises this neatly, arguing that people "act as self-appointed censors and take it upon themselves to condemn those words and constructions that they feel do not measure up to the standards they perceive they should say".

2.1.3.3. Specifiable Contexts: Metadiscursive Knowledge of Tabooness

Of the many factors that might influence a listener's judgement of taboo language use, there is one which is arguably as significant as perception of speaker intent, if not able to supersede it. Culpeper (2019) argues that the regular experience of certain expressions being used in markedly negative or prohibited contexts can 'tag' them as taboo. This is significant for three reasons. First, it helps to explain why speaker intent does not control how taboo language is received, in that the intent of a single speaker in a single interaction cannot override experiences of the same language having been used by many more speakers in many more interactions. Second, it helps to account for why responses to taboo language can differ interpersonally, because knowledge and experience of how such language is used will naturally differ from person to person. Finally, it can account for why some words are considered taboo while other words that might be considered similar are not. Here, Culpeper (2019) gives the example of 'paki' and 'brit' in British English: both are clipped forms of the demonyms 'Pakistani' and 'Briton', but the former is considered highly offensive while the latter is not. This, he argues, is due to 'paki' regularly occurring in the contexts of insults, threats and even hate crimes. I argue this is related to the positioning of British Pakistanis within the intergroup status hierarchy of ethnic groups in the United Kingdom, and by extension to systemic racism and marginalisation.

It is important to mention that these interactions between expressions and negative contexts do not need to be experienced directly to be marked as taboo by a listener, although such interactions may stand out more to listeners who do have personal experience of them. The concept of *metadiscourse* – where typical associations between certain expressions and certain contexts constitute part of broader knowledge of and discussion about linguistic use (Culpeper, 2011; 2019) – helps to explain this. Haugh and Culpeper (2018), for example, found that a majority of uses of 'nigger' in the British component of the Oxford English Corpus were instances of people talking about the slur, rather than their own uses or personal experiences of it. Simply knowing that certain language tends to be used in taboo contexts may therefore be enough to affect how that language is received by a listener, regardless of speaker intent, or even necessarily the listener's own experience.

Additionally, there is evidence to suggest that this knowledge of taboo metadiscourse develops quite early in life. Jay (1992) found that within children's lexicons, a rapid increase in the number of words which adults might identify as taboo takes place between the ages of one and four, and is co-emergent with a general increase in overall vocabulary. Linking this phenomenon to knowledge of taboo metadiscourse also developing during early childhood, Jay (2019a, pp. 97-8) notes that during the early phase of anger expression between the ages of two and four, "…children

have the opportunity to hear the name-calling, scatology, and epithets that parents use". By late childhood, Jay argues that conscious use of "symbolic emotional expressions in the form of offensive taboo words (slurs, curses, insults, name-calling)" (2019a, p. 97) is itself evidence of a new stage of development in children's use of emotional expressions. Critically, though, once awareness of taboo metadiscourse starts to develop, it keeps developing well into adulthood. Referring to findings from an earlier study (Jay & Jay, 2013), Jay suggests that younger and older children disagree about the inappropriateness of specific words, arguing in particular that "children at different ages have different views of taboo words' meanings because they have different knowledge about sexuality, gender identity, and social status" (Jay, 2019a, p. 104). However, while the Jay and Jay (2013) study makes a convincing argument that more "adult-like" taboo lexicons rely upon more "adult-like" knowledge, and demonstrates that children's evaluations of words as acceptable or unacceptable become more like those of adults as they grow older, it appears to conflate a child's ability to recognise a word as being prohibited with having a developed understanding of *why* this is the case. This is potentially a false equivalence; the accumulation of social feedback and experience might teach children as they age that using a taboo expression brings negative consequences, but it might take longer for them to appreciate the specific reasons that its use is considered harmful or insulting. This can be as true of adults faced with new taboo expressions as it is of children, but being able to recognise that a word is generally considered unacceptable is still a form of metadiscursive knowledge, clearly linked to the accumulation of social experience. I therefore suggest that as broader social awareness develops as one ages, so too do perceptions of language as being prohibited. However, I also acknowledge that developing these perceptions may not be indicative of similar development in a person's social and political knowledge of particular words and phrases.

Based on the body of work discussed in this section, I therefore argue that in much the same way that ingroup/outgroup distinctions affect the designation of certain language as taboo at the levels of communities and wider societies, so too do the illocutionary and perlocutionary forces of particular expressions significantly affect whether those expressions are perceived as taboo within specific, individual interactions.

2.1.3.4. Interactions between Community, Time, and Context

As a brief point, I note that the community, time, and interactional context can interact with each other in a number of ways, and can jointly affect taboo language judgements. For example, the status of certain words and phrases within particular ingroups, communities and societies can also be subject to changes over time. A useful case – to which I return in section 2.2.4. – is that of 'queer' in the LGBTQ+ community. Gamson (1998, p. 592) suggested that there exists a generational difference in attitudes to the word queer, in that younger members of the community had not been alive to experience what they described as "the sting, ostracism, police batons and baseball bats that accompanied it one generation earlier", and therefore displayed less aversion to its use. As I go on to explore, this is something of a problematic claim, not least because in the case of 'queer', the abusive history that the word carries can be viewed as a major motivator for its reclamation (Baker, 2008; Brontsema, 2004; Hall, 2014). However, the perception of a generational difference in the LGBTQ+ community's own use of 'queer' nevertheless illustrates that taboos within specific communities may also be subject to changes over time.

2.1.4. Relevance of Being a Native Speaker to Defining Taboo Language

There is one last factor which I argue is essential to making judgements of tabooness, though it has often been assumed or overlooked in existing taboo language definitions: that is native speakers who determine and are most likely to perceive linguistic taboos.

Dewaele (2019, p. 219) makes the point that "appropriate use of swearwords and taboo words requires proper semantic and conceptual representations as well as considerable sociocultural and sociopragmatic awareness". Referencing Thomas (1983), he argues that non-native speakers of a language – who often have incomplete semantic and conceptual representations of both taboo and non-taboo words – tend to be at risk of pragma-linguistic and socio-pragmatic failures, in which they are unable to accurately judge the force of the taboo word. He acknowledges that with sufficient development, non-native speakers can attain this knowledge for individual taboo words and phrases, but "might still misjudge the effect…on their interlocutors because of their status as outsiders" (Dewaele, 2019, p. 220). This last point is especially significant, because it highlights a relationship between native speaker status, ingroup/outgroup membership and taboo language judgements. First, native speaker and non-native speaker can themselves be considered ingroup/outgroup categories, which directly impact perceptions of taboo language because they are constructed around command of the language as a whole. Second, where ingroup/outgroup memberships are affected by linguistic behaviours like taboo language use, non-native speaker status may cause issues, in that any mistakes caused by incomplete knowledge of the language may instead be perceived as a failure to adhere to ingroup norms.

Furthermore, there is a body of psycholinguistic research which shows that native speaker status is key to making in-context judgements of tabooness. Based on findings from his own experimental research, Jay (1992; 2000) explicitly attributes the ability to judge the appropriateness of taboo words based on salient contextual information to being a native speaker. Similarly, Jay and Janschewitz (2008) found that non-native speakers are not as sensitive to contextual factors influencing the interpretation of taboo terms like insults. Pavlenko (2012, p. 409) attributes these differences to differences between native and non-native speakers in terms of their *affective processing* of the language in question, a term describing a type of response triggered by the emotional appraisal of a stimulus, in this case a verbal one. They explain that affective processing in a speaker's native language is more automatic, and is linked to heightened electrodermal reactivity to L1 emotion-laden words. By contrast, they note that non-native speakers experience decreased automaticity of affective processing, which "reduces interference effects and lowers electrodermal reactivity to negative emotional stimuli" (2012, p. 405). Pavlenko attributes this difference to the intense affective socialisation in their native language that native speakers receive from early childhood (see section 2.1.3.). Late bilinguals and language-learners do not receive this in their non-native language(s), which they consequently process intentionally and semantically, rather than automatically and affectively. A speaker's ability to perform automatic processing in a language is highly relevant to their ability to accurately identify taboo language in particular, because a number of psycholinguistic studies have evidenced that taboo words (particularly expletive swear words) are processed automatically as well (Damasio & Geschwind, 1984; Devinsky & Schachter, 1997; Finkelstein, 2019). It should come as little surprise, then, that research has also found that taboo words elicit a greater autonomic response from participants when written in their first language (Harris, Aycicegi, & Gleason, 2003).

On this basis, it is essential that any definition or explanation of taboo language takes into consideration the difference between native and non-native language use. I argue that this is especially the case for definitions of taboo language which stress the importance of context, as the linguistic competency that comes with native speaker status may directly impact appreciation of the contextual factors that inform judgements of language as being taboo.

2.1.5. Defining Taboo Language – Section Summary

In this section of my review, I have argued that while existing definitions of taboo language all provide some useful insights, they are often incomplete at best or unclear and selective at worst. In bringing all of these definitions together, I have provided a detailed argument for an understanding of taboo language that stresses the importance of community, time, interactional context, and native speaker status to defining taboo language. Accordingly, I have offered a definition of taboo language which aims to be as accurate as possible, and appropriate for both psycholinguistic and sociolinguistic study: words and propositions generally recognised by native speakers to be proscribed for a specific community, at a specifiable time, in specifiable communicative contexts, because they are felt to cause personal or social harm. This is the understanding of taboo language that I uphold for the rest of this review, and indeed the rest of my thesis.

2.2. Slurs and Linguistic Reclamation

Having defined taboo language, I now consider *slurs* as the specific type of taboo language with which my research is concerned. Defining slurs and outlining their function is necessary for explaining the concept of *linguistic reclamation*, which is also central to my thesis. Once I have explained these two concepts, I relate reclamation to my understanding of taboo language as set out in section 2.1., in order to explain how reclamation can affect whether particular slur uses are perceived as taboo. I then conclude this section with a focused discussion of LGBTQ+ slurs and their reclamation – the specific language that my research investigates.

2.2.1. Slurs as a Category of Taboo Language

Though various definitions of slurs have been offered, they generally agree on the importance of one factor: slurs are a type of taboo language specifically intended to derogate. Chen (1998, p. 130) succinctly identifies the slur as a linguistic example of a "derogatory sign or signifier". This idea is developed by Croom (2011, p. 343), who describes them as disparaging remarks or slights used to "derogate members of different classes", the utterance of which is itself an "explosively derogatory" act. These last points identify slurs as a class of taboo language, in that they are highly emotionally arousing (Jay, Caldwell-Harris, & King, 2008), a perceived cause of harm (Allan, 2019) and consequently are of restricted social use (Brandes, 2019).

The notion of class or group membership within definitions like Croom's draws a clear link to Tajfel and Turner's (1986) social identity theory, particularly to ingroups/outgroups and the intergroup status hierarchy, as discussed in section 2.1.2. Indeed, I argue that this is central to understanding how the derogatory function of slurs operates. This is a view supported by existing research: when discussing what he describes as "offensive slang terms", Jay (2019b, p. 83) notes both that they "are used to describe in- and outgroup members with the intent of insulting them" and that "these kinds of insulting slang terms for outsiders or non-conformists develop in all kinds of social communities". Slurs are able to derogate because they reflect, reproduce and reinforce

broader social power asymmetries, which exist between particular ingroups and outgroups at different levels of the intergroup status hierarchy.

Critically, this also suggests that to be considered a slur, a word does not derogate based on group differences alone, but also on differences in social power and prestige. In lay terms, slurs 'punch down', they do not 'punch up'. This is the most contemporary understanding of slurs in linguistic research. Tirrell (2012) and Herbert (2015) argue that it is these power dynamics which distinguish slurs from simple negative labels for particular social groups; Herbert (p. 133) compares the racial slur 'nigger' with the pejorative term 'whitey', noting that while both words derogate based on race, the latter does not reflect "long standing practice of systematic dehumanisation, marginalisation, and exclusion from social spaces" as the former does. For an example in relation to sexual identity, compare the slur 'faggot' (targeting non-heterosexual men) to the pejorative term 'breeder' (targeting heterosexual people). Both are generally recognised as negative labels attached to particular identities, but the systemic marginalisation of non-heterosexual identities (and not of heterosexual identity) distinguishes the former as being a slur, while the latter is not. Although she does not explicitly make this distinction, Popa-Wyatt (2016) notes that many of the most prohibited (and therefore the most taboo) slurs are those which target the most marginalised and oppressed social groups. Importantly, she also suggests that it is more socially acceptable (and therefore less taboo) to direct such language at members of more powerful groups, precisely because of the power such group membership affords, which is felt to lessen the impact of the derogation.

It is on this basis that alongside other scholars, I argue that any definition of slurs must regard them as a linguistic means of exercising hegemonic power. Baker (2008, p. 123) explains the Gramscian (1971; 1985) concept of *hegemony* as involving "the exercise of power, whereby everyone acquiesces in one way or another to a dominant person or social group". It is in this regard that hegemony and social identity theory can be understood as related, in that both consider the positioning of groups within a larger social power hierarchy. He argues that according to Gramsci's theory, this power structure is maintained not just through physical and economic power, but through ideology as well. This in turn allows individuals to participate in or acquiesce to the power structure without being fully aware of their status within it, or even of its existence.

For the purposes of this research, I consider slurs to be a class of taboo language widely used by members of more powerful social groups to derogate members of less powerful social groups. However – as discussed throughout section 2.1. – community, time and interactional context can still affect how taboo a particular use of a slur is judged to be. Jones (2012, p. 11) notes that even the powerful ideological messages that underpin hegemonic power structures "*influence* rather than *determine* the way that we view the world and, in this sense, may also be challenged" [italics in original]. With this in mind, I now turn to discussing one of the key means by which the slur can be subverted: the process of linguistic reclamation.

2.2.2. Defining and Understanding Linguistic Reclamation

Bianchi (2014, p. 35) defines linguistic reclamation as "uses by targeted groups of their own slurs for non-derogatory purposes, in order to demarcate the group, and show a sense of intimacy and solidarity". Given the nature of slurs as discussed in section 2.2.1., it may seem counterintuitive to suggest that they could be used with positive feeling, but Wald (2019, p. 349) powerfully observes that the use of taboo language which feels "…hostile to outsiders but warm to those playing with shared fire is broadly, deeply human".

Bianchi's (2014) definition is useful for several reasons: it explicitly connects reclamation to the use of slurs; it reflects an understanding of slurs as derogatory and hegemonic; and it outlines several key purposes reclamation can serve. However, Bianchi's description of targeted groups reclaiming "their own slurs" is potentially misleading. As I discussed in section 2.2.1., slurs are part of the linguistic arsenal of dominant groups, not marginalised ones. Reclamation is the active subversion of this status quo – the means by which slurs are taken up by members of the groups they are intended to harm. This is noted by Herbert (2015) when describing the problems with framing reclamation as a 'taking-back' of language: doing so presumes that control of the term (insofar as this is possible to have) was ever possessed by the targeted group. Arguably, this is also perpetuated by the word 'reclamation' itself (<re-> + <claim> + <-ation>), but alternatives have not been suggested, and the term is now generally accepted.

Instead of viewing reclamation as taking back a term that was never possessed to begin with, Herbert argues that reclamation should instead be considered a form of social protest, and that "reclamation projects are centred on the implicit idea that the targeted group *ought* to have control over the term that has been used against them" (p. 131) [italics in original]. Because total control over the use of any language is impossible, I argue that the issue is instead one of agency: of individuals targeted by a slur redefining how it relates to them. This understanding of reclamation as an act of personal resistance is neatly captured by Sinfield (1994, p. 205), who described it representing a defiant "we don't care what they call us" attitude on the part of the socially and politically marginalised.

Accordingly, while I broadly adhere to Bianchi's definition of reclamation, I also support Herbert's criticism and her view of reclamation as a fundamentally political act. Therefore, I argue that any definition of reclamation must also clearly consider what language is being reclaimed *from*, and *by whom*. My own definition of linguistic reclamation is as follows: non-derogatory uses, by members of marginalised groups, of the slurs which members of dominant groups use to target them.

Such an understanding positions reclamation as an ideologically-motivated process, and for this reason it is important to acknowledge that even within targeted groups, not all members engage in reclamation projects equally, and the individual decision to engage or not can be made for a variety of reasons (Brontsema, 2004). This is an observation unfortunately absent from some of the existing literature discussing linguistic reclamation, which can adopt a polarised view of slurs as either being reclaimed, or not being reclaimed at all. This is problematic for several reasons. First, it suggests a clear dichotomy where one does not exist. The literature never establishes a threshold of non-derogatory use that a slur must surpass in order to be considered reclaimed – presumably because this is impossible to pinpoint – but nevertheless suggests that some slurs are reclaimed, and some are not.

Second, it characterises reclamation as a process with a definite end point, but to extend Herbert's (2015) point about the broad aim of reclamation projects, once members of targeted groups start to claim agency over the way particular slurs are used towards them, that agency has to be maintained and upheld. As social attitudes and taboos change over time (see sections 2.1.3.) it is entirely possible that the oppressive meanings of slurs may be lost, potentially even in favour of reclaimed ones, but this outcome is slow to reach and not an inevitability. In general, reclaimed uses of slurs often coexist and compete with oppressive ones, so even successful reclamation projects must work to preserve the non-derogatory meanings they establish. This is not well-reflected by a view of slurs as simply 'being reclaimed'.

Most troubling, however, is that the labelling of a slur as either fully reclaimed or not reclaimed suggests complete consensus regarding the reclamation of that slur among those it targets. Brontsema (2004) criticises the tendency in previous research to imply that the reclamation of particular slurs is either collectively supported or collectively opposed. Instead, she proposes a more sophisticated model of attitudes and approaches to reclamation, which operates on two axes: whether an individual believes a slur can be separated from its derogatory history, and whether reclamation is subsequently supported or opposed. Brontsema argues that this leads to three main competing attitudes to reclamation: (1) slurs cannot be separated from their derogatory use, and therefore should not be reclaimed; (2) slurs can be separated from their derogatory use, and therefore should be reclaimed; (3) slurs cannot be separated from their derogatory use, and therefore should be reclaimed. Brontsema notes that position (3) is most similar to the radical, social-protest motivation for reclamation suggested by scholars like Sinfield (1994) and Herbert (2015), in which the derogatory history of the word actually motivates the desire to claim positive use of it. Theoretically, a fourth option also is possible: that slurs can be separated from their historical use, and therefore should not be reclaimed, but Brontsema notes that this would be counterintuitive to combatting linguistic oppression.

Brontsema's model offers a more nuanced understanding of reclamation in several key regards. First, it directly relates the phenomenon to issues of social oppression, which I have argued are critical to understanding both slurs and their reclamation. Second, it allows for different attitudes to reclamation to be represented, rather than suggesting any unilateral view of particular reclamation projects. Most importantly, it demonstrates that an individual's attitude toward a slur's oppressive history can motivate them either to reclaim it, or reject it. Compare the competing attitudes that Brontsema (2004) presents in positions (1) and (3): both suggest that slurs cannot be separated from their oppressive history, but (1) views this as reason to avoid use of the expression altogether, while (3) uses this as an explicit reason to reclaim it. Similarly, compare positions (2) and (3): both support engaging in reclamation projects, but (2) does so with the intent of divorcing a slur from its oppressive history and capabilities, while (3) argues that this cannot (and perhaps should not) be achieved.

Importantly, Brontsema (2004) does not argue in favour of one particular position, noting that each represents a different strategy for resisting the language of hegemonic oppression, and that there are contradictory problems particular to all three. Although it does not offer a clear, singular position to take, I argue that her model provides a significantly more accurate representation of contemporary attitudes toward reclamation projects. Therefore, in addition to defining reclamation in terms of hegemonic oppression, I also acknowledge that reclamation is not universally considered an appropriate, desirable or even possible response to that same power struggle.

2.2.3. Reclamation and Perceptions of Tabooness

Having outlined an understanding of linguistic reclamation, I now consider the ways in which the phenomenon can affect perceptions of a slur use as taboo. In particular, I focus again on how reclamation may affect the three main factors which determine tabooness, as identified by Allan and Burridge (2006): community, time, and interactional context.

Issues of community membership – particularly those of ingroup/outgroup membership and intergroup status – are central both to slurs and to attempts to reclaim them. As such, I simply note that because the idea of community membership is key to participation in reclamation projects

(Bianchi, 2014; Herbert, 2015; Sinfield, 1994) and because these same issues also affect judgements of tabooness (Allan, 2019; Allan & Burridge, 2006; Culpeper, 2019; Jay, 2019b), it follows that the reclamation of slurs might affect perceptions of their tabooness.

Discussion of reclamation with regard to temporal shifts in words considered to be taboo can be kept similarly brief. I argue that as more and more individuals participate in reclamation projects, they can drive broader changes in a slur's meaning over time. I evidence this further in section 2.2.4., where I discuss specific examples of LGBTQ+ slur reclamation. I argue that if reclamation can be part of a gradual shift toward a more positive (or at least less strongly negative) general understanding of a slur, then it can also contribute to a gradual decline in the degree to which it is perceived as taboo. However, as noted in section 2.2.2., whether such temporal change can entirely separate a word from its historical use is a question which divides opinion. Additionally, participants in an intergenerational interaction may have different levels of experience with a particular slur and any changes in its use over time. If collective acts of reclamation contribute to gradual reductions in perceived tabooness over time, a speaker or listener with more experience of a slur before it underwent this process may perceive its intent much more negatively.

I devote most focus in this section to the ways reclamation can interact with the contextual, within-interaction factors that authors like Allan and Burridge (2006) and Culpeper (2019) describe. In section 2.1.3., I particularly highlighted the importance of speaker intent and listener reception to making judgements of taboo language use. To explain how this concerns reclamation, I return to Wald's (2019) metaphor of playing with shared fire, mentioned at the start of section 2.2.2. Within a given interaction, for use of a slur to be successfully perceived as a reclamation, it is vital that both speaker and listener recognise each other as sharing the same fire (i.e. both being a target of the slur); this helps clarify whether it is being played with (i.e. used non-derogatorily), or potentially being used to exercise hegemonic power. Knowledge (or perhaps presumption, in some cases) of both speaker and listener identity is an important contextual factor affecting whether a use is perceived as a reclamation, and by extension as taboo.

I also refer back to the point made by Culpeper (2019), that the regular experience of words in negative or prohibited contexts can "mark" them as being taboo regardless of speaker intent (see section 2.1.3.). Here, I suggest that reclamation provides a possible opposite case, in that it offers a means of experiencing the same words in positive and permitted contexts. If such a parallel can be drawn, then by the same argument it is possible that reclamation might begin to reduce the extent to which a word is marked as taboo. Arguably, awareness of reclamation might similarly affect an individual's metadiscursive knowledge (Culpeper, 2011; 2019) of a slur. Even if someone does not engage in reclamation of a particular slur themselves, having an experienced understanding that others do might change their general perception of its tabooness, or at least narrow the contexts in which they continue to regard it as taboo.

Finally, I also discussed in section 2.1.3. the way in which perceived intent relates to the concept of the speech act. Particularly, I observed that the extent to which language is judged to be taboo can depend as much on the intended and perceived goal of its use within a particular interaction, as it can on the tabooness of its denotative meaning. This is important, because some research has suggested that slurs in particular can be used within *exercitive* speech acts – those which grant or deny rights to a person or group through their utterance – in their capacity to reproduce and reinforce oppressive social norms (McGowan, 2012; Herbert, 2015). If this is how a slur is situationally intended or perceived to be intended, it is likely to also be considered a taboo use; consider again Allan's (2019, p. 3) argument that behaviours are considered to be taboo if they

are perceived to cause "discomfort, harm, or injury". However, I argue that reclamation necessarily undermines the ability of slurs to function in such speech acts. In the redeployment of a slur to name oneself or the group to which one belongs, that slur cannot be perceived to be conferring or denying rights to others, because others are not the target of its use. As long as this is recognised by all participants within a particular interaction, it is possible that the slur will neither be intended nor perceived as an exercitive act, and therefore may not be perceived to be taboo to the same degree.

Having now discussed understandings of how reclamation can be defined and how it operates, as well as relating this to how reclamation might affect judgements of particular slur uses as taboo, I will now consider some specific examples of slurs which continue to be the subjects of reclamation projects. With regard to the present research, I take these examples from the LGBTQ+ community, and begin by justifying my decision to focus on this group in particular in my work.

2.2.4. LGBTQ+ Slurs and their Reclamation

My reasoning for focusing on examples of reclamation from the LGBTQ+ community in my research is twofold. First, I am drawing from feminist epistemologies and considering my own 'standpoint' as a researcher. Hankinson-Nelson (1993) describes how feminist *standpoint theory* emerged from feminist scholars gaining increased access to the academy, noting in particular that they were 'situated' by their feminism. This allowed them to observe and criticise how curricula and pedagogy had been found lacking because of the widespread exclusion of women's voices from knowledge (Bar-On, 1993), by virtue of possessing such a voice themselves and being able to see how it was not being represented. I therefore consider it an advantage to centre my research on a community of which I am myself a member, and on derogatory language which I have direct knowledge and experience of, some of which I choose to reclaim myself. While I acknowledge that this risks an element of researcher bias, I consider my personal and epistemological standpoint particularly advantageous given the specific focus of my research on slurs and reclamation, both of which I have consistently argued are bound to issues of community identity, ingroup membership and systems of structural oppression. I believe that being a member of the group targeted by such language affords me personal insight into its use and impact that I do not have for language targeting other marginalised groups to which I do not belong. I also believe that knowing my community membership will be of benefit to those who participate in my research; it is less likely to be a concern that my work is participating in or contributing to the hegemonic power structures which underpin slur use, if it is being carried out by someone who could themselves be a target of such language.

Second, the LGBTQ+ community is one with a rich history of navigating oppression and marginalisation through the use of language. For example, Baker (2002; 2008; 2019) has written extensively on the history of Polari, what he describes as an *anti-language* – a re-lexicalised version of an existing language used primarily by people who are "somehow apart from mainstream society, either residing on the edges of it, perhaps frowned on in some way, or hidden away or even criminalised, with attempts from the mainstream to expel or contain them" (2019, p. 17). In the case of Polari, these people were mainly gay and bisexual men living before the decriminalisation of same-sex relations in the UK in 1967. Though the use of Polari has since declined, to this day a rich vocabulary of language particular to the LGBTQ+ community exists, included among which are a number of reclaimed slurs (Baker, 2008; Coleman-Fountain, 2014; Russell, Clarke, & Clary, 2009; White, 1980).

I now discuss some key findings on LGBTQ+ slurs from the existing literature, noting in particular what they demonstrate about processes of linguistic reclamation and their relationship to judgements of tabooness. I begin with 'dyke', a slur used to derogate lesbian and bisexual women (or those perceived as such), particularly those whose gender expression is regarded as masculine. Pascoe (2007) notes how homophobic slurs are often directed toward those who are felt to be performing their gender incorrectly or unconventionally, even if the target may be heterosexual, because homosexuality is itself associated with a failure to adhere to heteronormatively gendered behaviours (Kiesling, 2001). Jones (2012) conducted an ethnographic study of a particular lesbian *community of practice* (CoP), defined by Eckert and Wenger (2005) as a number of people who come together in a shared enterprise, develop their own shared ways of doing things (practices) and subsequently develop a shared orientation to their environment. Jones' study focused on a lesbian hiking group, and within their linguistic practices she observed a reclamation of 'dyke'. She related this to a broader lesbian reclamation of 'dyke' (see Zwicky, 1997), but also noticed the reclaimed slur fulfilling a specific function within the particular CoP she was observing. Jones noted that when the hikers described someone as a dyke, they were indexing characteristics shared with the 'butch' label in lesbian culture, what she describes as a "non-feminised style of womanhood" (2012, p. 64). In itself, this is perhaps unsurprising, given the term's tendency to predominantly target women seen to be masculine. However, Jones observed that in contrast to 'butch', 'dyke' was used to specifically index the concept of an authentic lesbian identity. I do not intend to argue that this is the aim of all reclamations of 'dyke', as Jones' observations are particular to the group which she was studying, but this finding is nevertheless very revealing. It is significant that use of a taboo word – even with positive and approving intent – could be linked to issues of authentic ingroup membership. Here, I refer back to my discussion in section 2.1.2., regarding the purpose taboos can serve in maintaining and policing ingroup/outgroup distinctions (Allan, 2019; Brandes, 2019). For this reason, this example illustrates that reclaimed slurs lend themselves particularly well to communicating issues of group membership, which I argue in turn reinforces the point made by Herbert (2015) that reclamation is not simply an effort to neutralise slurs or turn them positive.

Next, I consider a pair of related homophobic slurs: 'fag' and 'faggot'. These slurs arguably parallel 'dyke' in their targeting of gay and bisexual men, especially because both Kiesling (2001) and Pascoe (2007) note that both can be similarly used to derogate men seen to be exhibiting femininity, regardless of their sexual identity. Compared to 'dyke' though, there is much less research on 'fag' and 'faggot' which suggests that they have been a focus of reclamation projects, instead noting a continuing resistance to them and the oppression they are felt to represent (Pascoe, 2005; Pascoe, 2007; Rivers, 2001; Smith & Smith, 1998). Indeed, some research has suggested 'fag' and 'faggot' have been the most prominent words used to derogate anyone perceived to be homosexual for several decades now (Johansson, 1981; Brontsema, 2004). However, there is emergent research which suggests that reclamation projects for both 'fag' and 'faggot' may now be underway, if less far along in their progress than those of other LGBTQ+ slurs, particularly in online discourse. Gordon (2019) conducted a corpus-based study of uses of 'fag' and 'faggot' on Twitter, finding that over half of the total tokens recorded in the study were reclaimed uses of the slurs as identity labels. However, Gordon also notes that the data did not indicate a unified or consistent reclaimed meaning or message, and that attitudes to the term are very divided in a manner much resembling Brontsema's (2004) model.

Finally, I turn to the LGBTQ+ slur most recognised as having been the focus of reclamation efforts, and perhaps the slur most frequently recognised within reclamation research as a whole:

'queer'. Indeed, the case of 'queer' is so salient an example of reclamation that it was used as the main body of evidence upon which Brontsema's (2004) model of reclamation was formed. Originally denoting a more general sense of abnormality, the word started to be directed particularly toward homosexual men who were felt to retain some features of masculinity (Brontsema, 2004; Gaucher, Hunt, & Sinclair, 2015); if they did not, they were likely to be targeted with homophobic language specifically connoting femininity instead. 'Queer' started undergoing more widespread reclamation within the LGBTQ+ community in the 1990s (Baker, 2008), a process which has led to a diversity of reclaimed meanings today. When used for self-identification, it offers a means of re-emphasizing the diverse, counter-cultural and subversive potential of LGBTQ+ identity; a political disavowal of a more mainstream, assimilationist LGBTQ+ politics (Stephens, 2011). As an identity label, 'queer' can also be used to denote flexibility, fluidity or resistance to categorisation in gender and/or sexual identity, as in the case of the label 'genderqueer'. Similarly, queer can also be conceptualised as an umbrella term which stands for a range of marginalised sexualities; rather than an identity itself, 'queer' functions as a hypernym while 'gay', 'lesbian' etc. are its hyponyms (Livia & Hall, 1997).

It has also been suggested that the word can also be used to refer to anybody "stigmatised for failing to comply with heteronormative mandates" (Edelman, 2004, p. 17). This reclaimed use has even been deployed within academic discourse to name whole fields of study and their associated theoretical positions, particularly *queer studies* and *queer theory,* both devoted to the criticism and destabilisation of heteronormative ideologies. However, this understanding should be treated with some caution, as even members of groups which use the term to exercise hegemonic power can "fail to comply" with the often punitive and restrictive standards of heteronormativity. Unfortunately, these uses tend to presume the kind of 'completed' reclamation process for 'queer' criticised in my earlier discussion (e.g. Brontsema, 2004; Herbert, 2015). Furthermore, queer theory tends to explicitly criticise and distance itself from the idea of a definite "queer subject" (Baker, 2008, p. 187), but as I have discussed, the existence and recognition of such subjects is necessary for understanding the reclamation of slurs which target them (Bianchi, 2014; Brontsema, 2004; Herbert, 2015; Sinfield, 1994).

2.2.5. Slurs and Linguistic Reclamation – Section Summary

The most contemporary literature on slurs makes it very clear that the particular relationship of slurs to hegemonic, institutionalised power structures definitionally distinguishes them from derogatory insults and other kinds of taboo language. Accordingly, current research also teaches how processes of linguistic reclamation – a sociolinguistic phenomenon particular to slurs – can be viewed as acts of protest against such power structures. Where the literature has sometimes been less clear is how the actual process of reclamation operates, and has occasionally mischaracterised reclamation projects as completable and/or universally participated in. My review has identified an alternative – and in my view, more accurate – perspective on the different attitudes towards and participation in reclamation projects within marginalised communities. It has also revealed how reclamation might itself affect perception of a word as being taboo, insofar as it interacts with the issues of community, time, and context that I previously argued determine judgements of tabooness. I have also used existing research to justify the focus of my research on LGBTQ+ slurs and their reclamation, and have detailed what is already known about particular examples of LGBTQ+ slurs and the degree to which they are being reclaimed. These perspectives inform my approach to slurs (particularly LGBTQ+ slurs) and linguistic reclamation throughout the entirety of my research. Having addressed these points, I can now turn to the central questions of my thesis: how taboo linguistic stimuli –

particularly slurs – are cognitively processed, whether linguistic reclamation might affect this processing, and in what regard.

2.3. The Processing of Taboo Language

I begin this section of my review by returning to one of the definitions I identified in section 2.1.1.: Jay, Caldwell-Harris, and King's (2008, p. 83) description of taboo language as "a class of emotionally arousing references…". Before approaching the issue of taboo language processing, it is necessary to explore psycholinguistic research on emotional language more broadly, focusing particularly on emotional arousal and the effects it has on our language processing.

2.3.1. The Effects of Emotional Arousal on Language Processing

van Berkum (2019, p. 740) offers a comprehensive working definition of *emotion* for the purposes of psychological and neurological research. He defines it as follows:

> "A package of relatively reflex-like synchronized motivational, physiological, cognitive, and behavioural changes, triggered by the appraisal of an external or internal stimulus event as relevant to the interests (concerns, needs, values) of the organism, and aimed at generating a prioritized functional response to that stimulus event. The changes involved need not emerge in consciousness, but to the extent that they do, they give rise to feeling".

This is a complex definition of emotion, which van Berkum (2019) explores each aspect in detail in his work. For the purposes of this review, I highlight some key insights from his definition concerning *emotional language* and the processing of the same. First, that the types of stimuli that can become associated with emotions are seemingly limitless, and that crucially these can include linguistic and semiotic signs when they become paired with actual or imagined experience of a particular emotional/affective response. Second, that emotions are fundamentally referential and evaluative, and it is in this respect that language – the prime demonstration of the human brain's capacity to represent and communicate ideas – stands out as a type of emotional stimulus worthy of close study. Thirdly, that emotional responses – to all stimuli, including linguistic ones – are immediate, and are not necessarily consciously acknowledged as *feelings*. As evidenced in section 2.2. of this chapter, discussion of slurs and their reclamation has been almost entirely focused on conscious feelings and opinions, while unconscious emotional responses remain largely unexplored. Finally, van Berkum quotes Besnier (1990, p. 443) to argue that "affect permeates all utterances across all contexts because the voices of social beings, and hence their affect, can never be extinguished from the discourse", suggesting that all language is to some extent emotional language, and that language processing cannot be studied completely without attention being paid to emotionality.

This definition of emotions incorporates several dimensions across which they may be measured, but many psychological and psycholinguistic studies of emotional stimuli (e.g. Kensinger & Corkin, 2003; Lang, Greenwald, Bradley, & Hamm, 1993; Russell, 1980) tend to assess emotionality on two in particular: *valence* (the degree to which a stimulus is positive or negative) and *arousal* (how exciting or shocking the stimulus is). Both dimensions are relevant to the cognitive processing of language, but the degree to which each dimension affects language processing is not necessarily equal; research has generally suggested that of the two, arousal often has the more significant impact.

One well-researched effect which demonstrates this difference is the impact emotional arousal has on word memory. Several decades of research have suggested that all kinds of emotionally arousing stimuli are more memorable; Brown and Kulik (1977) famously described emotionally-charged memories as *flashbulb memories*, so named because individuals recall them as clearly as if they had been photographed. While more contemporary research (e.g. Phelps & Sharot, 2008; Talarico & Rubin, 2003) has since disputed that the emotional charge of a particular memory affects the accuracy with which it is remembered – arguing instead that individuals demonstrate more *belief* in the accuracy of highly emotional memories – there remains a strong body of evidence that emotional arousal and memorability are closely linked. This seems to be just as true in the case of emotional language. For example, Kensinger and Corkin (2003) found that language which either had clearly positive or negative valence, or which elicited emotional arousal, was better remembered than language which had neutral valence and was not arousing. However, their study also found that the benefit to word memory was stronger in words which evoked arousal than in words which did not, regardless of valence.

Some research (e.g. Ochsner, 2000) has suggested this may be due to the capacity of all emotional stimuli to prompt certain neurological and physiological responses, themselves more attributable to emotional arousal than to valence. One such response is that emotional stimuli prompt increased activation of the *amygdaloid complex* – a brain structure often associated with emotional functioning (Anderson & Phelps, 2002) – which in turn seems to prompt increased memory for the emotional stimulus. Indeed, research suggests that activation of the amygdala is linked to the process of memory-formation itself (Bzdok, Laird, Zilles, Fox, & Eickhoff, 2013; Packard & Cahill, 2001; van Berkum, 2019). Neurological evidence has suggested that damage to the amygdala can impair emotional responses in a number of ways, and that this includes the impairment of arousal-enhanced memory (Anderson & Phelps, 2002; LaBar & Phelps, 1998).

Activation of the *adrenergic system*, which is connected to the production of the hormones epinephrine (adrenaline) and norepinephrine (noradrenaline), has also been linked to the impact of emotional arousal on memory (Sharot & Phelps, 2004). It is particularly interesting that these two hormones appear to be involved in this relationship, as they are a crucial part of the *fight-or-flight response*, itself a series of physiological reactions produced by the autonomic nervous system to a stimulus which is perceived to be frightening, dangerous or threatening (Jansen, Ngyuen, Karpitskiy, Mettenleiter, & Loewy, 1995). This link is supported by a study conducted by Cahill, Prins, Weber, and McGaugh (1994), in which the system responsible for the release of these hormones was pharmacologically impaired. In a similar manner to the effects of damage to the amygdala, when this system was impaired, the study found that so too were the effects of emotional arousal on memory.

Recalling words is not the only type of language processing affected by emotional arousal, however. There is a (comparatively smaller) body of research which suggests that emotionality also affects *lexical access* – the process through which we access our mental lexicon and retrieve the meaning of any word we encounter (Traxler, 2012). Many of these experiments have been conducted using the *lexical decision task* paradigm: a kind of response-time study in which participants are presented with written stimuli and must give a response to indicate whether what they see are real words or not. The speed with which participants are able to give this response corresponds to the speed of lexical access for the word they encountered. I use the lexical decision paradigm in my own research, as part of the experiments I describe in Chapters 5 and 6.

The body of research on the general accessibility of emotional language is now somewhat dated, but these earlier studies tended to find – at least in neurotypical individuals – that emotional

language prompted both faster (Williamson, Harpur, & Hare, 1991; Strauss, 1983) and more accurate (Graves, Landis, & Goodglass, 1981; Strauss, 1983) lexical decision responses than emotionally neutral language. Unfortunately, such studies did not typically recognise (and therefore did not control for) a distinction between emotional arousal and valence. However, they appear to have focused more on valence than arousal, given their tendency to divide verbal stimuli into positive, negative and neutral categories (e.g. Williamson, Harpur, & Hare, 1991). This also presents an issue for these earlier findings, given that more recent studies have highlighted the importance of arousal, and not of valence. Because of these issues, and because technological advancements since these studies were published have allowed for more precise data collection and analysis, these results should be considered with some caution.

Nevertheless, it is useful to end this section of my review on the subject of lexical access, as it has been part of the focus of a much broader and better-developed body of psycholinguistic research into taboo language processing. This research has sought to compare the processing of taboo and non-taboo emotional language, which is the issue to which I now turn.

2.3.2. The Effects of Tabooness on Language Processing

I have provided evidence that emotionally arousing language has measurable effects on cognition and language processing, but have not yet discussed the effects specifically associated with taboo language. There is a growing body of evidence that taboo language should be considered somewhat distinct from other kinds of emotional language. In addition to the neurological effects of emotionally arousing language discussed in section 2.3.1., neurolinguistic research has also identified that expletive swearing in particular is linked to activation of the *basal ganglia*: a network of subcortical nuclei which is closely connected to the thalamus and the cortex, and which has been implicated in motor actions; learning and habit formation; and – most importantly – emotions and cognition (Finkelstein, 2019). However, the majority of findings on taboo language processing have come from psycholinguistic studies.

Such studies have generally suggested that while taboo language reproduces a number of emotional language effects, these tend to be significantly exaggerated for taboo items when compared to non-taboo items, even if the latter are also emotional (Jay et al. 2008; Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Madan, Schafer, Chan, & Singhal, 2017). One of the clearest examples of this is the performance of taboo versus emotional language in word memory tasks. Jay et al. (2008) conducted an experiment which included a surprise recall task, finding that three times as many taboo words were successfully recalled on average, compared to non-taboo emotional words. Buchanan, Etzel, Adolphs, and Tranel (2006) also found that taboo words were recalled with more success than any of the other word categories in their study, including moderate-arousal emotional words. As I discuss later in this section, more recent studies (e.g. Madan et al. 2017) have suggested that this exaggerated arousal effect observed for taboo words is not attributable to factors like word frequency and familiarity. However, Buchanan et al. (2006) suggest that the semantic relatedness (see Tulving & Pearlstone, 1966) of taboo words might also enhance their recallability, compared to emotionally arousing words that are less semantically related. The precise nature of the semantic relation of taboo words is an important question, to which I will shortly return.

One possible explanation for findings like these is that taboo language might produce stronger effects because it is more arousing than non-taboo emotional language. If emotional arousal is responsible for many of the effects discussed in section 2.3.1. – including those concerning

word memory – then it would make sense that the more arousing a word is, the stronger these effects would be. Furthermore, there is evidence to support the idea that taboo words are more arousing than high-arousal, non-taboo words. Janschewitz (2008) produced a comprehensive database of normed word ratings for taboo, emotionally valenced and emotionally neutral words. The emotionally valenced words were further divided into positive high-arousal, positive low-arousal, negative high-arousal and negative low-arousal categories. She found that the taboo category yielded a significantly higher mean arousal rating than any other word category in her results, including both the positively- and negatively-valenced high-arousal categories.

Experimental studies have produced similar findings. To further explore the results of their first recall experiment – as discussed earlier in this section – Jay et al. (2008) conducted a second study which measured *skin-conductivity responses* (SCRs) to taboo, emotional and neutral words. SCRs represent electrodermal activity in response to a stimulus (in this case a verbal one), which is amplified when the stimulus is more arousing. They found that taboo items elicited a higher frequency of SCRs than the other word categories, and that the amplitude of SCRs in each of the word categories was positively correlated with recall scores. Therefore, taboo words constituted the most arousing category, they were the most frequently recalled, and these two findings were found to be correlated.

However, a growing body of evidence suggests that while stronger arousal responses contribute significantly to the processing of taboo language, they do not fully account for its effects by themselves. To explain this, I return to the issue of the semantic relatedness of taboo language. The list-based definitions of taboo language that I describe in section 2.1.1. demonstrate that words classed as taboo can have a range of semantic relations, from body parts to religious terms not approved for casual use. This raises the question of what quality of their meaning is shared. In addition to other word properties like word frequency, arousal and valence, a number of psycholinguistic studies have focused on a property termed *tabooness*: how inappropriate the stimulus is considered to a general population (Bertels, Kolinsky, & Morais, 2009; Madan et al. 2017).

There are two distinctions that should be made regarding tabooness at this stage. First, assessing a word's tabooness is not the same as judging whether a particular use is breaking a taboo. Tabooness is an assessment of how socially inappropriate a word might be considered in most contexts by most people, not a judgement of whether a particular instance of its use breaks a taboo. This is why definitions of tabooness as an emotional property of a word focus on general-population, context-free perceptions, even though current definitions of taboo language stress the contextual dependency of language as taboo. Second, tabooness needs to be particularly distinguished from *offensiveness*, which is quite often considered its own word property in taboo language processing studies. Ratings of offensiveness tend to measure how upsetting or insulting an individual personally considers a word to be, while ratings of tabooness measure individual judgements of how inappropriate the word might be felt to be across an entire population (Madan et al. 2017). As such, it can be argued that perceptions of tabooness relate more to the metadiscursive knowledge described by Culpeper (2011), which I outlined in section 2.1.3. The shared semantic relation of taboo language, then, is a general recognition of socially inappropriate meaning.

Studies have shown that tabooness may be key to explaining aspects of taboo language processing which cannot be explained by emotional arousal alone. With respect to the effects of taboo language on word memory, Madan, Caplan, Lau, and Fujiwara (2012) explicitly compared taboo and non-taboo emotional language for memorability, finding a boost in memory for taboo words beyond what they were able to explain by emotional arousal. The authors argued that this

may indicate a taboo-specific mechanism influencing performance in the memory task, but that this could not be concluded from their study alone. Convincing evidence in support of this came later, in one set of findings from Madan et al. (2017). Their study set out to identify how a variety of word properties impacted emotional and taboo language processing: word frequency, number of letters and syllables, familiarity, personal use, imageability, contextual diversity, arousal, valence, offensiveness and tabooness. Results from their free recall test replicated the finding that taboo words are more memorable than non-taboo words, regardless of emotionality. What is particularly interesting, though, is that they also found that a group of *emotional word properties* (arousal, offensiveness and tabooness) produced the model which best explained the results of the recall test. This is a finding in line with the evidence discussed in section 2.3.1. Critically, though, if the tabooness variable was removed from the model, it was no longer able to explain their results.

I now return to the topic of lexical access, because tabooness also helps to explain an unusual finding from the literature. In contrast to the faster and more accurate lexical decision responses for emotional language generally reported in early studies (see section 2.3.1.), significantly impaired performance has often been demonstrated for taboo words in lexical decision tasks. An early study by Williams and Evans (1980) found that taboo words were not responded to as quickly or as accurately as neutral words, although this study – like several from its time – made no distinction between taboo words and non-taboo, emotional words. Instead, taboo words were simply used as examples of emotional language, without consideration that they might have their own effects. In this regard, these results cannot themselves demonstrate a specific tabooness effect. However, the authors made the important suggestion that their findings might be attributable to a *perceptual defence mechanism*, in which response times are slower due to an urge to avoid the processing of troubling stimuli. In this regard, the perceptual defence mechanism can be related to the concept of *approach-avoidance motivation* in behavioural psychology (for review see Higgins, 1997; Elliot, 2006; Elliot, Gable, & Mapes, 2006; van Berkum, 2019), in which *approach motivation* describes "the energization or direction by, or the direction of behaviour toward, positive stimuli" and *avoidance motivation* describes "the energization or direction by, or the direction of behaviour away from, negative stimuli" (Elliot, 2006, p. 112). Because taboo language is closely related to direct threats, social sanctions and punishments in a way that emotional language more broadly is not (see section 2.1.), it may be associated with stronger avoidance motivation. Following this logic, research has suggested that tabooness alone is responsible for an increased avoidance of processing, subsequently impairing lexical access (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005).

Indeed, as research using the lexical decision paradigm has developed, the distinction between taboo and otherwise emotional language has been made clearer. Geer and Bellard (1996) compared lexical decision performance for words with sexually explicit (taboo), romantic (non-taboo, emotional) and neutral meanings. Their results indicated that lexical decisions were slowest in the sexually explicit category, and that this effect could not be wholly explained by their measures of word familiarity, emotionality, or acceptability. As such, they suggested this delay may be caused – at least in part – by the words' taboo semantic content. Thomas and LaBar (2005) explicitly categorised their stimuli into groups of taboo words, low-arousal negative words and neutral words, finding that participants took longer to correctly semantically categorise the taboo words than they did either of the other categories. Here, there was a clear distinction made between taboo and non-taboo emotional language, with taboo language similarly seeming to undergo slower lexical access, although it should be noted that the categories did not eliminate a possible effect of emotional

arousal. With this in mind, some of the most convincing evidence again comes from the study conducted by Madan et al. (2017). Their lexical decision results replicated the finding that taboo words were associated with the slowest responses when compared to positive emotional words and neutral words. In comparison to the model they used to explain free recall findings, in which emotional word properties and tabooness produced the best explanation, they found that their *non-emotional word properties* (familiarity, personal use, word frequency, contextual diversity, number of letters, number of syllables) and tabooness produced the best explanation of these results. As with their free recall findings, if tabooness was removed from their model so that only the non-emotional properties were considered, its ability to explain their results was diminished.

Overall, we can see that in comparison to non-taboo emotional language, taboo language is slower to undergo lexical access, but is more memorable afterwards. There is also evidence that both of these effects are partly attributable to a specific tabooness word property, in addition to the effects of emotional arousal and similarly relevant properties. I will now end this section of the literature review by considering these findings in relation to all of my discussion thus far, in order to address the central question of this thesis.

2.3.3. Considering the Impact of Linguistic Reclamation on the Processing of Slurs

There is no research specifically investigating the processing of slurs as a subclass of taboo language, nor how the phenomenon of linguistic reclamation might affect slur processing. Accordingly, this part of my discussion is relatively brief. However, it is necessary to explain why I think linguistic reclamation might have any effects on slur processing to begin with. To do so, I draw on all of my discussion up to this point.

The evidence I have examined throughout this section has suggested that the cognitive effects observed for taboo language arise primarily from emotional arousal responses and tabooness judgements. As I have also noted, both of these variables are frequently assessed in general-population terms, particularly the tabooness property, which is explicitly meant to represent a decontextualised measure of social unacceptability.

However, this is problematised by linguistic reclamation as a phenomenon, for a number of reasons. First, a general-population approach does not align very well with the contemporary understanding of taboos and taboo language that I outlined throughout section 2.1., particularly regarding the definition set out by Allan and Burridge (2006) that stresses the ways that taboos interface with community, time and context. This is especially true regarding slurs and their reclamation; I stressed throughout section 2.2. that both are heavily connected to issues of community membership and use in-context, which might not be represented at a general-population level. In section 2.2.3., I also noted a variety of ways that reclamation might directly affect general perceptions of particular slurs, in particular how it could lessen arousal responses and tabooness judgements.

I therefore argue that any investigation of the processing of slurs needs to account for their reclamation, as the phenomenon may cause deviations from what is currently understood about taboo language processing. This, ultimately, is the position from which I developed the research questions I set out in section 1.2. of my introduction (Chapter 1).

2.3.4. The Processing of Taboo Language – Section Summary

Within this section, I have described the effects of emotional arousal – an unconscious physiological and neurological response to an emotionally charged stimulus – on elements of cognitive language

processing, evidencing in particular that it has generally been linked to enhanced word recallability and faster lexical access. I then outlined the effects of taboo language as a particular type of emotionally arousing language, noting that while the tabooness property appears to enhance word recallability even further, it has been seen to slow lexical access rather than accelerate it. These insights particularly informed my design and analysis of the psycholinguistic lab experiments I describe in Chapters 5 to 7. However, I noted that there is a total absence of psycholinguistic research concerning the particular processing of slurs as a type of taboo language. By extension, there has been no consideration of how linguistic reclamation might interact with other factors shown to affect various aspects of taboo language processing, such as familiarity, personal use, emotional arousal, and tabooness. This gap in the research identified by my review served as motivation for my own studies, which are devoted to answering these questions, and filling this particular gap in the research.

2.4. Literature Review Conclusion

To conclude this chapter, I return to the three questions I outline in the chapter introduction. First, what makes language taboo, and what definition of taboo language do I adopt in my research? Second, what makes a taboo word or phrase a slur, and what is meant by the linguistic reclamation of slurs? Third, what is known about the manner in which taboo language is cognitively processed, and how might linguistic reclamation thereby affect the processing of slurs?

Regarding defining taboo language, I have argued that the approach of listing types of taboo references is insufficient, for four main reasons: it does not tell us why types of reference are taboo, the types of reference listed are not always clearly defined and distinguished themselves, these lists are rarely exhaustive, and the selection of items is often determined at the individual discretion of the researcher. I argued that this makes such definitions unsuitable for social scientific work, where a definition that is as consistent and widely-encompassing as possible is preferable, to properly enable comparisons of findings across studies and research contexts. Instead, I have outlined an approach to defining taboo language which stresses the social purpose of all taboos as prohibitions of behaviour; primarily intended to prevent discomfort, harm, or injury; carrying social consequences for breaking them. In terms of how and when these proscriptions are applied to language, my review identified community, time, and interactional context as three factors particularly affecting taboo language judgements. In particular, I have stressed the importance of ingroup/outgroup membership, temporal shifts in meaning, and specific speaker and hearer relationships to determining which words are tabooed and when. I then reviewed research suggesting that all of these factors as they apply to language use require a level of pragma-linguistic and socio-pragmatic understanding typically only afforded to native speakers. This leads me to defining taboo language as language generally recognised by native speakers to be proscribed for a specific community, at a specifiable time, in specifiable communicative contexts, typically because they are perceived to cause some form of social or personal harm.

Regarding slurs and linguistic reclamation, my review led me to conclude that slurs are a class of taboo language widely used by members of more powerful social groups to derogate members of less powerful social groups. In doing so, I supported the argument from contemporary research that slurs are more than just derogatory labels, instead being a linguistic means of reproducing and reinforcing hegemonic power structures, in order to oppress members of marginalised groups. I then defined the process of linguistic reclamation as non-derogatory uses, by members of marginalised groups, of the slurs which members of dominant groups use to target

them. I also explicitly criticised the idea that slurs are ever entirely reclaimed or unreclaimed, instead evidencing that reclamation is a continuous and long-term political language project. Furthermore, I argued that the decision to engage in reclamation projects is fundamentally a personal one, and that even within communities that engage in reclamation, there are divides in opinion regarding its purpose and effectiveness.

Finally, regarding the cognitive processing of taboo language, I began by outlining the body of evidence concerning the processing of emotionally arousing language. I described how emotionally arousing language is generally associated with better recallability and improved lexical accessibility. I then evidenced that taboo language in particular is associated with even stronger recallability, but impaired lexical accessibility. I argued that these effects are not because taboo language is a more arousing category of emotionally arousing language, but because of a specific impact of tabooness, in combination with other word properties specific to the type of processing effect being observed. For lexical access, these other word properties include non-emotional ones like word frequency, familiarity, and personal use. For recall, they instead include emotional word properties like arousal and offensiveness.

The literature has not previously explored the impact of linguistic reclamation on taboo language processing, but based on all of my discussion in this chapter, I make the following general hypotheses in response to my first two research questions (see section 1.2. of Chapter 1). First, that the cognitive processing of LGBTQ+ slurs might resemble that of taboo language more generally, but that there is reason to believe slurs might prompt particularly strong arousal and avoidance responses, especially among those they target. Second, that linguistic reclamation will affect responses to slurs, as there is good reason to believe it affects factors like familiarity, negativity, arousal, and tabooness, all of which contribute to taboo language processing. Regarding my final research question and the sociolinguistic significance of my findings, I need to establish what my findings are before I consider their significance to the field, but argue at this stage that my work will develop existing knowledge of LGBTQ+ slurs, lend new evidence to definitions and discussions of linguistic reclamation, and potentially reveal unconscious effects of a fundamentally sociolinguistic phenomenon on responses to very socially-situated language. The next five chapters of this thesis are devoted to the studies I conducted to identify LGBTQ+ slurs and the attitudes towards them and their reclamation, test the cognitive processing of LGBTQ+ slurs, and ultimately confirm the hypotheses I have just outlined.

Chapter 3 – LGBTQ+ Slur Identification Survey

3.1. Introduction

This chapter focuses on the first stage of data collection and analysis that I conducted: a survey study designed to identify frequently-used LGBTQ+ slurs in British English. This study was the initial step in exploring how linguistic reclamation might affect the processing of LGBTQ+ slurs, in that it enabled me to determine which LGBTQ+ slurs are currently in use. Collecting frequency data for LGBTQ+ slurs was essential groundwork for the experimental studies to follow; for any slurs to be used as stimuli in those studies, I needed to be confident that respondents would recognise them.

3.2. Research Background

Some taboo language processing studies have included LGBTQ+ slurs in their experimental stimuli, despite never focusing explicitly on such language. As one of the most comprehensive norming studies for emotionally arousing and taboo language, I began by exploring the stimuli included in Janschewitz (2008). This study included the words 'gay', 'homo', 'fag', 'queer', 'lesbo', 'cocksucker' and 'dyke' under its 'taboo' category, listed here from most to least familiar, based on the mean familiarity ratings identified in the results. This data provided an initial indication of which words might be recognised as LGBTQ+ slurs by the general public, even offering some information about their familiarity and frequency of use. However, it also raised two key questions regarding the study's stimuli selection and categorisation. First, why were only these 7 words pertaining to the LGBTQ+ community selected? Second, what motivated the decision to categorise 'gay' as taboo, given that the most familiar meaning of the word is arguably an entirely acceptable sexual identity label?

These questions were not answered by the report accompanying the database; Janschewitz (2008, p. 1067) explains that "the [taboo] words were selected from those that appeared in Jay (1992) as well as the discretion of the experimenter". Unfortunately, researcher discretion does not explain why stimuli were chosen, nor are all of Janschewitz's judgements regarding selection and categorisation of LGBTQ+ related words clear. For example, she identifies a category of slurs relating to 'sexuality', but includes 'homo' as part of an 'other insults' category. Furthermore, while she explicitly describes a category of racial slurs in her stimuli, she does not similarly recognise a category of LGBTQ+ slurs.

The Jay (1992) book cited as the other source of stimuli contains similar issues. 'Cocksucker', 'dyke' (spelled 'dike' in original), 'fag', 'homo' and 'queer' were present in Jay's studies, all of which were designed to identify the frequency and contexts of use of taboo words. The methodology explains that all taboo words were identified by 12 American college students (6 men and 6 women), whose job was to record instances of "dirty word utterances" (p. 120) as they spontaneously occurred both on- and off-campus. These recordings were taken in the form of field cards, which asked respondents to note the utterance along with the age and sex of the speaker, the target and any others present. They were also asked to record the manner in which the utterance was spoken, where it was spoken, and any additional comments about its use. As such, the LGBTQ+ related taboo language data in both Janschewitz (2008) and Jay (1992) was sourced from a very small sample of American English speakers, gathering data almost 30 years ago. It also means that the identification of taboo words in these studies was dependent on what these 12 individuals – for whom there was no demographic information provided other than their gender – themselves encountered in their own social circles and judged to be "dirty words" (p. 119).

'Gay' was not included in the words identified in these experiments, which casts further doubt on Janschewitz's (2008) categorisation of it as a taboo word. However, it did appear in a later chapter of Jay's (1992) book, in which he discusses offensiveness in regard to gender-specific insults, specifically words used to insult men. Here, he suggests that "one of the most offensive insults to a male is to suggest that he is sexually inadequate or a homosexual" (p. 178). This is a problematic suggestion, for a number of reasons. First, there are no specific sources mentioned for this particular claim; it is the author's own summation of the findings of a broad body of offensiveness research. Second, it appears to conflate homosexuality with perceived sexual inadequacy, without being explicitly critical of this comparison. Third, it presumes the targeted man is *not* homosexual; a man who is and is comfortable with his sexuality would not likely be offended by an accurate description of it. If the mere insinuation of homosexuality was the sole basis for the categorisation of 'gay' as taboo by both Jay (1992) and later Janschewitz (2008), then I did not consider it appropriate to use these studies as my basis for identifying LGBTQ+ slurs. I also noted that 'lesbo' – targeting women perceived to be homosexual, not men – did not appear anywhere in Jay (1992). I concluded that this word must have been included in Janschewitz's taboo stimuli at her own discretion, with the reason or source for this being unclear, which was not helpful for determining whether it should be included in my own stimuli.

These two studies highlight the main difficulties with reliably identifying LGBTQ+ slurs from existing research, but an even greater problem was that many previous taboo language processing experiments did not include examples of such language at all. Returning to some of the experimental studies I discussed in section 2.3.2. of my literature review (Chapter 2), neither Geer and Bellard (1996) nor Madan, Schafer, Chan, and Singhal (2017) included any LGBTQ+ related taboo words in their stimuli lists. Kensinger and Corkin (2003) did not provide a list of their taboo stimuli at all, nor a clear explanation of how these words were selected and categorised. Two studies from which Jay, Caldwell-Harris, and King (2008) sourced their stimuli (Belleza, Greenwald, & Banaji, 1986; and MacKay, Shafto, Taylor, Marian, Abrams, & Dyer, 2004) also did not include any LGBTQ+ related taboo words.

The studies I have discussed in this section are still extremely valuable to my research for what they reveal about taboo language and its processing; I simply consider that they do not help to identify taboo language relating to the LGBTQ+ community, as they have generally not considered it. Furthermore, where such words have been included, their selection and categorisation has typically been unclear and/or problematic. This motivated me to conduct my own study, in order to resolve both of these problems. However, I was able to use some findings to make some early predictions. In spite of the lack of focus given to them, the presence of some LGBTQ+ related words in Janschewitz (2008) and Jay (1992) nevertheless suggests these stand out as examples of taboo language. Therefore, I predicted that 'cocksucker', 'dyke', 'fag', 'homo' and 'queer' would be identified as LGBTQ+ slurs in my own survey. As I have discussed, I did not feel the same prediction could confidently be made for 'gay' and 'lesbo', but considered that these words might also be frequently identified in my survey responses.

3.3. Methodology

3.3.1. Using a Survey to Identify LGBTQ+ Slurs and Frequency Data

To identify LGBTQ+ slurs and their respective frequencies, I initially considered using the British National Corpus (BNC): a general, monolingual corpus of 100 million words taken from samples of

British English in the late 20th Century (Burnard, 2007). However, this approach presented several problems. First, it would require me to pre-identify LGBTQ+ slurs, so that I could then search for them in the corpus. Given my earlier criticism of including slurs solely at the discretion of the researcher, and my commitment to avoiding imposing my own assumptions in this research (see Chapter 1, section 1.4.), I considered this a methodological issue to be avoided.

Second, corpora like the BNC are designed with the purpose of being generally representative of a particular language variety (Leech, 1992; McEnery, Xiao, & Tono, 2006), and the BNC itself consists of significantly more written (90%) than spoken (10%) samples (Burnard, 2007). I believed such a corpus might underrepresent taboo language, for two reasons. There is evidence to suggest that taboo words typically account for less than 1% of the words in a variety of sampled utterances (Jay, 1980; Jay, 2019b; McEnery, 2006; Mehl & Pennebaker, 2003; Thelwall, 2008), so would likely account for a similarly small proportion of the words in a general language corpus. Furthermore, many of the attempts to quantify the rate with which speakers use taboo language have used spoken data (Jay, 2019b), while the composition of corpora like the BNC are far more focused on written texts.

These problems were compounded by two further issues: first, that the representativeness of corpora is recognised to be time-specific (Hunston, 2002) and the currently-available version of the BNC lacks 21st century data. Second, that I would only be interested in searching for taboo language targeting a select group of people, which imposes even further restriction on what would already be a very limited pool of corpus data. I therefore decided to use an open-response survey, both to identify LGBTQ+ slurs while avoiding personal bias, and to gather information about slur frequency by counting the number of times each is identified by separate respondents.

3.3.2. Sample

I recruited respondents through social media (Facebook and Twitter), particularly the social media pages of LGBTQ+ groups based in Nottingham, UK (with their permission). My sample was therefore a non-probability sample, using elements of purposive sampling (primarily targeting LGBTQ+ individuals) and snowball sampling (selecting key 'starting points' and encouraging subsequent distribution through respondents' own networks) to gather data. For a detailed discussion of these sampling methods, see Sarantakos (2005). In this regard, it offered some of the typical advantages of non-probability sampling: it enabled targeted recruitment, had flexible parameters, was low-cost, and was time-efficient (Sarantakos, 2005). Additionally, sampling via social media channels allowed me to efficiently distribute the online survey, with added convenience and privacy for respondents in that they did not need to contact me directly for access. However, I acknowledge that my approach was also subject to the typical limitations of non-probability sampling, in that it produced a smaller sample, the size of which was not statistically determined (Sarantakos, 2005).

One hundred and fifteen unique respondents completed the survey, and no responses were excluded. The only pre-requisites for participation were that respondents be over the age of 16 (to give their own informed consent to participate) and be native or near-native speakers of British English. The latter requirement kept language variety and native speaker status (see section 2.1.4. of Chapter 2) as consistent as possible, without excluding participants who do not consider British English their native language, but nevertheless have high levels of fluency and sociopragmatic knowledge (see section 2.1.4. of Chapter 2) of British English taboo expressions. I did not collect demographic information about respondents or restrict their participation accordingly, as the only aim of this study was to identify the most frequent LGBTQ+ slurs in general-population terms. This

included not requiring that respondents identify as LGBTQ+, although by targeting LGBTQ+ groups in my sampling I ensured that the research still reached members of this community, as I considered they might be familiar with a wider range of slurs.

3.3.3. Survey Design

The main body of the survey consisted of two open-response questions, the first compulsory and the second optional. The first question asked respondents to list as many slurs as they could think of which specifically target members of the LGBTQ+ community. The second question then gave respondents the space to provide further comments on any of the slurs they had listed, if they wished to.

3.3.4. Procedure

On the front page of the survey, I provided respondents with information about the overall research and the particular survey they were responding to. Reading this information was required to proceed, and after reading all respondents provided informed consent. They then provided their responses to one or both of the questions. Finally, they were thanked for their contribution and were able to save their responses and close the survey. The survey typically took 5-10 minutes to complete, depending on length of responses.

3.3.5. Data Analysis

Because the first question simply generated a list of LGBTQ+ slurs for each respondent, I manually collated these lists and counted the frequencies with which all slurs occurred. Given that I intended to use the slurs I identified in word recognition experiments (Chapters 5 to 7), and that syllable length has particularly been shown to affect word recognition (Chetail, 2014), I decided to treat any variations of a slur which differed in syllable length (e.g. 'fag' and 'faggot') as separate entries. I did not treat simple orthographic variations that did not affect syllable length (e.g. 'lezza' and 'lezzer') as separate entries.

For the optional second question, I decided to use a content analysis framework to analyse all additional comments given by respondents. I employed two types of content analysis: quantitative and qualitative. As a means of text analysis, quantitative content analysis "seeks to show patterns of regularities in content through repetition" (Ericson, Baranek, & Chan, 1991). In this regard, computer-based tools such as corpus linguistics software can be useful (Fielding & Lee, 1991; Hansen, 1995). Because of its focus on identifying concepts which occur and reoccur, quantitative content analysis is related to thematic analysis – a means of identifying and evidencing themes across a dataset which an analyst can then use to create a framework for comparison and discussion (Gomm, 2004). However, it is important that frequency of repetition is not taken as the only measure of significance; to do so would not only reduce the data to what the researcher considers quantifiable (May, 1997), but also risks overlooking individually significant words and phrases (Scott, 1990). There is therefore a need for qualitative content analysis too, an approach which allows the analyst to highlight particular aspects of the data as significant in regards to the specific (social and production) context(s) from which it has arisen (Ericson, Baranek, & Chan, 1991; May, 1997).

I began by using WordSmith Tools corpus software (Scott, 2012) to conduct a quantitative content analysis and identify frequency-based themes for my analysis. In accordance with the process set out by Gomm (2004), I produced a code for the themes that I identified, and annotated second question responses accordingly, enabling me to identify the presence of each theme across

responses. As noted by Hammersley and Atkinson (1995) and Dörnyei (2007), these 'log trails' constitute an important evidence base for qualitative analysis. Accordingly, I provide a coded and annotated version of my second question responses in Appendix 1. In order to identify any points of significance that might have been missed by the quantitative content analysis, I also conducted a qualitative content analysis via a close reading of all responses, in which I identified any statements which stood out as significant in their own right.

3.4. Results

3.4.1. Number of Unique LGBTQ+ Slurs Identified

Overall, the survey generated 164 unique slurs specifically targeting the LGBTQ+ community. Four of these were excluded due to not being English words, giving a final total of 160. I provide the full list of LGBTQ+ slurs and their respective frequencies in Appendix 2, but have identified the 20 most commonly identified LGBTQ+ slurs in Table 3.1, for illustration.

Table 3.1. – Most Frequently Identified LGBTQ+ Slurs

Slur	***N***
Dyke	86
Faggot	83
Tranny	78
Fag	74
Queer	74
Poof	69
Homo	38
Lesbo	38
Lezzer	37
Gay (Pejorative)	25
Bender	23
Gayboy	23
Batty-Boy	22
Poofter	20
Fairy	19
Shemale	19
He-She	18
Rug-Muncher	16
Fudge-Packer	15
Pansy	14

3.4.2. Semantic Fields of LGBTQ+ Slurs

Having identified a full list of LGBTQ+ slurs, I then observed that many belonged to at least one of three distinctive semantic fields. Pursuing this line of analysis provided valuable insights into the kinds of meaning that are indexed by slurs targeting individuals for their gender and/or sexual identity.

The most common semantic field consisted of slurs relating to sexual acts. I identified 46 of these slurs, meaning that this category represented over a quarter (28.75%) of total slurs identified. Table 3.2. lists all of these and the frequencies with which they were identified by respondents. I then identified a number of subcategories concerning the specific sexual acts referenced within this

semantic field. Table 3.3. lists these, alongside the number of unique LGBTQ+ slurs belonging to each subcategory, and the percentage of all references to sexual acts represented by each.

Table 3.2. – Slurs Relating to Sexual Acts

Word	***N***	**Word (cont.)**	***N***
Batty-Boy	22	Brown-Hatter	1
Rug-Muncher	16	Bum-Basher	1
Fudge-Packer	15	Bumder	1
Arse-Bandit	12	Butt-Boy	1
Shirt-Lifter	12	Butt-Muncher	1
Batty	8	Butt-Pirate	1
Bummer	8	Chutney-Ferret	1
Cock-Sucker	7	Cock-Jockey	1
Muff-Diver	7	Cock-Muncher	1
Bum-Boy	6	Cunt-Licker	1
Shit-Stabber	5	Cushion-Biter	1
Sodomite	5	Fanny-Fiddler	1
Bugger	4	Fish Fingers	1
Uphill-Gardener	4	Knobjockey	1
Batty-Man	2	Lettuce Licker	1
Butt-Fucker	2	Lickalotapuss	1
Muff-Muncher	2	Pillow-Princess	1
Pillow-Biter	2	Pussy-Muncher	1
Turd-Burglar	2	Rug-Diver	1
Anal-Architect	1	Rump-Roaster	1
Arse-Goblin	1	Self-Confessed Player of the Pink Oboe	1
Ass-Pirate	1	Sod	1
Backdoor-Warrior	1	Tunnel-Tester	1

Table 3.3 – Specific Sexual Acts Referenced within Sexual Acts Semantic Field

Sexual Act	***N***	**% of All Sexual Acts Referenced**
Anal Sex	32	69.57
Oral-Vaginal Sex	8	17.39
Fellatio	3	6.52
Digital-Vaginal Sex	2	4.35
Oral-Anal Sex	1	2.17

I then identified a second semantic field related to gender expression. This contained 31 slurs, representing just under a fifth (19.38%) of total slurs identified. Table 3.4. lists all of these and the frequencies with which they were identified by respondents. To explore what kinds of gender expression are particularly targeted, I then subcategorised these slurs, depending on whether they referenced feminine or masculine gender expressions. Table 3.5. provides these results, alongside the number of unique LGBTQ+ slurs belonging to each subcategory, and the percentage of all references to gender expression represented by each.

Table 3.4. – Slurs Relating to Gender Expression

Word	*N*	Word (cont.)	*N*
Fairy	19	Chichi	1
Pansy	14	Effeminate	1
Butch	8	Faggy	1
Bull-Dyke	6	Fey	1
Camp	5	Fruitcake	1
Queen	5	Fruity	1
Sissy	5	Ginger	1
Fruit	4	Jessie	1
Lemon	4	Limp-Wristed	1
Nancy	4	Mannish	1
Ponce	4	Mary	1
Nancy-Boy	3	Pussy	1
Diesel-Dyke	2	Queenie	1
Martha	2	Tomboy	1
Mincer	2	Trixie	1
Pussy-Boy	2		

Table 3.5. – Type of Gender Expression Referenced within Gender Expression Semantic Field

Type of Gender Expression	*N*	% of All References to Gender Expression
Femininity	26	83.87
Masculinity	5	16.13

Finally, I identified a semantic field of abnormality and deviance. This contained 17 slurs, representing over a tenth (10.63%) of total slurs identified. Table 3.6. lists all of these, alongside the frequencies with which they were identified by respondents. I then realised that these could be further subcategorised to indicate specific types of abnomality and deviance. Table 3.7. provides these subcategoriess, alongside the number of unique LGBTQ+ slurs belonging to each, and the percentage of all references to abnormality/deviance represented by each.

Table 3.6. – Slurs Relating to Abnormality/Deviance

Word	*N*
Queer	74
Bender	23
Bent	10
Pervert	6
Deviant	2
Invert	2
Perv	2
Unnatural	2
Confused	1
Freak	1
Ill	1
Mentally Ill	1
One of Them	1
Sick	1
Twist	1
Warped	1
Weirdo	1

Table 3.7. – Shared Semantic References Used to Indicate Abnormality/Deviance

Shared Semantic Reference	***N***	**% of All References to Abnormality/Deviance**
Strange/Unacceptable Behaviour	8	47.06
Malformation	5	29.41
Illness/Being Unwell	4	23.53

3.4.3. Second Question Responses

Of the 115 total respondents, 52 (45.22%) provided additional comments in the form of a response to the second, optional question. There are two main sets of results that I identified from this section of the survey. The first concerns the frequency with which specific LGBTQ+ slurs identified in first question responses also appeared in second question responses. 34 (65.38%) responses to the second question contained references to specific LGBTQ+ slurs, with 65 such references made overall. Table 3.8. lists these, by order of frequency.

Table 3.8. – References to Specific LGBTQ+ Slurs in Second Question Responses

Word	***N***
Queer	29
Gay	13
Dyke	10
Chichiman	2
Tranny	2
Arse-Bandit	1
Batty Bwoi	1
Batty Man	1
Bitch	1
Butch	1
Chichi/Cheechee	1
Faggot	1
Fairy	1
Genderqueer	1
Invert	1
It	1
Jessie	1
Lesbian	1
Pansy	1
Puffter	1
Queer Folk	1
Trans	1
Warped	1

The second set of results concerns my content analysis, as described in section 3.3.5. In Table 3.9., I list all of the items that I judged to be relevant for further analysis, taken from a corpus-generated word list of the top 50 most frequent words to appear in second question responses. To reduce data pollution, all non-lexical items found in the top 50 most frequently occurring words within the British National Corpus were excluded from this list.

Table 3.9. – Frequencies of Relevant Words in Top 50 Word List

Word	***N***
I	65
Queer	29
Used	28
Slurs	22
Slur	17
Gay	13
Reclaimed	13
Community	11
People	11
Use	11
Now	10
Dyke	10
Feel	6
Context	5
Heard	5
LGBTQ	5
Men	5

Comparing Tables 3.8. and 3.9., it is clear that the three LGBTQ+ slurs which appeared most frequently in second question responses ('queer', 'gay', 'dyke') were also some of the most frequent words to appear in these responses in general, emphasising their particular significance to respondents. I was also able to use the results in Table 3.7. to immediately identify references to words being 'reclaimed' (13 occurrences) as a distinctive theme for further analysis.

For many of the other words in the list, I needed to carry out further analysis to determine whether there were any themes underlying their inclusion. This was achieved through a collocation analysis, in which I used the same corpus software I used to produce my quantitative content analysis (WordSmith Tools, see section 3.3.5.) to determine pairs or groups of words that frequently co-occurred. For example, the most frequent word in the data (once stop list items were removed) was the personal pronoun 'I', occurring 65 times across responses. Because it was not possible to establish how the pronoun was being used by single-word frequency alone, searching for its collocations allowed exploration of its use in-context.

This collocation analysis revealed a number of additional themes for analysis. Three themes were drawn out by further exploring collocates of 'I'. 'I think' (10 occurrences) and 'I feel' (3 occurrences) were typically used to identify and distinguish personal feelings about a particular slur. 'I heard' (4 occurrences) and 'I use' (2 occurrences) were used to discuss individual experiences of slurs, from the perspective of either speaker or listener. 'I am' (5 occurrences) was typically used to indicate the relationship between a respondent's own identity and the slur(s) they were highlighting. This last theme was also evidenced by exploring collocates of 'community' (11 occurrences), all of which described particular groups of people in relation to their use and experience of slurs: 'LGBTQ(IA+) community' (5 occurrences); 'the community' (2 occurrences); 'African Caribbean community' (1 occurrence); 'lesbian community' (1 occurrence); 'queer community' (1 occurrence); 'working class community' (1 occurrence).

Another theme to emerge from collocation analysis was the idea of particular words being 'used as slurs' (4 occurrences) or 'used as a slur' (3 occurrences), suggesting that other uses of the same words are possible. I observed this idea in regard to other collocations of 'slur' and 'slurs': 'not

always slurs' occurred twice in responses. It was also noticeable in regards to the presence of the word 'context' (5 occurrences) in the frequency table, which itself collocated three times with 'depend' or 'depending' and twice with 'offensive', to discuss the importance of context in determining whether a word is judged as a slur (see section 2.1.3. and 2.2. of my literature review for discussion). Related to this was another theme revealed through collocation analysis: within sentences, the word 'now' (10 occurrences) was consistently used to discuss temporal (particularly generational) changes in the use of particular LGBTQ+ slurs (see section 2.1.3. and 2.2. of my literature review for discussion).

Based on these results, Table 3.10 lists the themes that I identified for my analysis of second question responses. As mentioned in section 3.3.5., I annotated my data to indicate wherever these themes occurred in responses (see Appendix 1). As Gomm (2004) recommends, I used these frequency-based themes as focus points with which to explore second question responses for important areas of discussion. Bearing in mind the criticisms of May (1997) and Scott (1990) discussed earlier, I also made sure to note any discussion that might have been missed by quantitative analysis which stood out as particularly significant in its own right. Such examples are integrated into the discussion that follows, as they become relevant.

Table 3.10. – Themes Identified through Content Analysis

Themes
Discussion of Specific Words/Slurs
References to Reclamation
Opinions/Experiences of LGBTQ+ Slurs
Relations between Identity and LGBTQ+ Slur Use
Multiple/Context-Dependent Uses of LGBTQ+ Slurs
Temporal/Generational Differences Regarding LGBTQ+ Slurs

3.5. Discussion

3.5.1. Insights from Identifying LGBTQ+ Slurs

The primary aim of this study was to identify LGBTQ+ slurs for use in my later studies. In this regard, the study was highly successful, identifying 160 unique British English LGBTQ+ slurs and providing an indication of word frequency for each. Consequently, I have ensured that stimuli selection for the studies described in Chapters 4 to 7 can be methodologically clear and data-driven. As discussed in section 3.2., this is something which distinguishes the identification of LGBTQ+ slurs in my research from that of previous taboo language processing experiments.

My identification of LGBTQ+ slurs yielded some very valuable findings, in terms of which slurs were identified, which were most frequent, and how my findings support or challenge the treatment of these words in previous taboo language research. I begin this discussion by focusing on the most frequently identified slurs, and on patterns across the whole dataset.

'Dyke' was the most frequently identified LGBTQ+ slur – included in just under 75% of all 115 responses – with 'fag', 'faggot', 'tranny' and 'queer' proving similarly familiar (see Table 3.1.). I note that while 'dyke' was clearly an extremely well-recognised LGBTQ+ slur, it may only have been the most frequently identified because of a lack of similarly common variations. Had I counted 'fag' and 'faggot' as the same slur (refer to section 3.3.5. for the reason I did not), they would have been the most frequently identified instead. This observation does not solely concern which slur was the most

frequent and recognisable; it was an early indication of the salience of slurs targeting LGBTQ+ men compared to LGBTQ+ women. This is an issue to which I return in section 3.5.2. of this discussion, and indeed throughout the rest of my research.

I make 3 additional observations about these findings. First and most straightforward, my results demonstrate that a wide variety of contemporary British English slurs exist which target members of the LGBTQ+ community. This derogatory and marginalising language is prolific, and is in active circulation. There are also many more LGBTQ+ slurs than existing literature would suggest; even sociolinguistic studies which explicitly focus on the use of LGBTQ+ slurs (e.g. Brontsema, 2004; Jones, 2012; Pascoe, 2005, 2007; Smith & Smith, 1998) have tended to consider only the slurs that my results suggest are the most recognisable examples, such as 'dyke', 'fag', and 'queer'.

Second, in comparison to the number of slurs that my survey identified, psycholinguistic studies of taboo language have tended to include only a small number of LGBTQ+ slurs as stimuli: only 7 of the 89 taboo words in Janschewitz's (2008) affective ratings database are LGBTQ+ slurs, for example. Considering that even this number is high compared to other experimental studies (see section 3.2. for additional examples), existing experimental studies of taboo language processing may not accurately or fully capture responses to LGBTQ+ slurs in particular. Again, I observed that many of the slurs which these studies have included were among the most frequently identified in my survey responses, although I acknowledge that this is not a disadvantage for psycholinguistic experiments, in which ensuring words are sufficiently recognisable is important.

Third, I observe that several of the most recognisable LGBTQ+ slurs overall are also those which are most frequently used to target particular LGBTQ+ subgroups: 'dyke' as the slur most commonly used to target women who are attracted to other women, 'fag' and 'faggot' as the slurs most commonly used to target for men who are attracted to other men, and 'tranny' as the slur most commonly used to target transgender people. This suggests that LGBTQ+ slurs are more often intended to target specific LGBTQ+ identities, rather than LGBTQ+ people as a whole group. However, the presence of 'queer' – which can be used to collectively reference all LGBTQ+ people, either positively or negatively (Livia & Hall, 1997) – among my most frequently identified LGBTQ+ slurs is a singular exception to this, which perhaps reflects its distinctive history and social salience (see section 2.2.3. for discussion).

My results also help to clarify some of the findings of previous studies. Referring back to Table 3.1., I found that 6 of the 7 LGBTQ+ related taboo words that appeared in Janschewitz (2008) were among the most frequently identified by my respondents: 'dyke', 'fag', 'queer', 'homo', 'lesbo', and 'gay'. This broadly supports Janschewitz's identification of these words as taboo, particularly 'gay' and 'lesbo', the inclusion of which was not explained beyond researcher discretion. The exception was 'cocksucker', included in Janschewitz's study due to Jay's (1992) claim that it is an especially offensive term to call a man, as it accuses him of homosexuality and – per his understanding – sexual inadequacy. I outlined all of the issues I take with this claim in section 3.2., but consider the fact that it was not one of the most frequently identified LGBTQ+ slurs in my study to be a final piece of evidence that the word is not as offensive as Jay claimed it to be. I acknowledge that part of the explanation for this finding could be that some respondents did not consider 'cocksucker' a term specifically targeting LGBTQ+ people; the act it describes is not specific to the LGBTQ+ community. However, as I go on to discuss in section 3.5.2., many other LGBTQ+ slurs reference fellatio, specifically those targeting non-heterosexual men.

While I consider the results of this study to broadly support the stimuli selection of previous studies, I include a caveat for the inclusion of 'gay' as a taboo word or LGBTQ+ slur. Whenever 'gay'

did appear in my first question responses, respondents consistently made it clear that they were referring only to pejorative uses, e.g. using "that is gay" to mean "that is bad".

Baker (2008, p. 129) describes this use of 'gay' as referring to "anything that is pathetic or useless", arguing that this extension in meaning is purposefully homophobic: "those who characterised gay people as pathetic probably extended the term to refer to *anything* that is pathetic" [italics in original]. This is in contrast to the use of 'gay' as a "historically typical sexual identity [label]" (Russell, Clarke, & Clary, 2009, p. 884), in this case for someone who experiences exclusively same-gender attraction. That my survey respondents needed to clarify which use they were referring to suggests both that this distinction is significant, and arguably that the pejorative use of the word is less common than its use as an identity label. For this reason, I argue that 'gay' can be considered a taboo word and/or slur, but only when being used pejoratively. I consider the pejorative use of 'gay' further in section 3.5.4.1., when analysing responses to my second survey question.

While I made a distinction between 'fag' and 'faggot' for entirely methodological reasons (see section 3.3.5.), it was interesting to observe that there was a difference between the two. As indicated in Table 3.1., 'fag' was less frequently identified than 'faggot', the former appearing in 74 (64.35%) responses while the latter appeared in 83 (72.17%). Indeed, 'faggot' was the second most common word to be identified as an LGBTQ+ slur overall, while 'fag' was the fourth. This prompted me to re-examine my data, to explore how often both slurs were actually listed as separate entries by individual respondents. I found that at least one of the two slurs was included in 98 (85.22%) of my responses. Within these, 58 (59.18%) identified both 'fag' and 'faggot', while 40 (40.82%) identified only one. Of the 58 which identified both, only 3 (5.17%) suggested any kind of interchangeability: "faggot (+ fag)", "Faggot/fag" and "Fag (Faggot)". Of the 40 which mentioned only one, 'fag' was identified less often, with 17 (42.50%) responses as opposed to 23 (57.50%) for 'faggot'. Collectively, these results show that it was more common to identify both 'fag' and 'faggot' as separate slurs than it was to identify only one of the two; that they were very rarely presented as interchangeable; and that when only one was identified, 'faggot' was more frequent than 'fag'. While both slurs are clearly well-known, I argue that they should be considered separately and not as interchangeable, that 'faggot' is more widely recognised as a slur in British English, and that it is perhaps the more offensive of the two forms.

In considering the apparent difference between 'fag' and 'faggot', I note that previous studies – both psycholinguistic and sociolinguistic – have not typically explored such a difference at all. For example, 'faggot' does not feature at all in the results of Jay (1992) or Janschewitz (2008), while 'fag' does. The difference between the two slurs is also not explained by examining more qualitative research into the use of 'fag' and/or 'faggot' (e.g. Smith & Smith, 1998; Rivers, 2001; Pascoe, 2005; Pascoe, 2007), because while such studies have tended to acknowledge the use of both words, they have also tended to treat them as interchangeable. Speculatively, a difference between the two in British English might be due to register; I suggest that 'fag' may be perceived as less aggressive due to a general perception that clipped forms are less formal and indicate less interpersonal distance than full forms of a word (Brooke, Wang, & Hirst, 2011). It is also possible that reclamation is also relevant here, which I discuss further in section 3.5.4.3.

3.5.2. Discussion of Semantic Fields of LGBTQ+ Slurs

As identified in section 3.4.2., some of the most valuable insights from this study came from identifying the semantic fields to which its LGBTQ+ slurs belonged. In this section of my discussion, I

consider the semantic fields of sexual acts, gender expression, and abnomality/deviance which emerged from my data.

3.5.2.1. Sexual Acts Semantic Field

Referring to Table 3.2., I observed some key morphological features of the slurs relating to sexual acts. Most noticeably, 37 (80.43%) of the slurs in this category were two-word compounds, such as 'arse-bandit' and 'knobjockey'. As with these two examples, 22 (59.46%) of these compounds were formed using at least one taboo word, including some instances of both words being taboo, e.g. 'butt-fucker'. This is perhaps to be expected; as noted in section 2.1.1. of my literature review, references to sex and body parts are extremely typical of taboo language and are often included within list-based definitions of it (e.g. Jay, 1992, 2000; Janschewitz, 2008; Jay et al. 2008; Madan et al., 2017). Nevertheless, some compound slurs did not include any taboo words (e.g. 'pillow-biter'), suggesting that these are not a required component. This – combined with the fact that a majority of compound phrases did include at least one taboo word – suggests that the inclusion of already taboo language in these compound phrases is deliberate, perhaps to portray particular sex acts and those who engage in them as obscene. This is supported by the further observation that 22 (59.46%) compound slurs included the derivational morpheme <-er>, used to derive nouns from verbs to describe the performer of an action. In such instances, the LGBTQ+ target is explicitly characterised in terms of the sex acts they are perceived to engage in, which are considered perverse from a heteronormative perspective, and reference to which often incorporating already-taboo language.

As shown in Table 3.3., anal sex was by far the most commonly referenced sexual act within the sexual acts semantic field, accounting for just under 70% of all such references. Slurs relating to anal sex could also be subcategorised: 16 referenced the giving partner (50.00%), 5 (15.63%) referenced the receiving partner and 11 (34.38%) were ambiguous or could be applied to both. I considered this a surprising result; given these slurs primarily target men who have sex with men, I initially expected that more slurs would target the receiving partner, as it is not the sexual role men are heteronormatively expected to satisfy. The literature supports this: in describing some of the metaphors used to discuss the taboo of sex, Crespo-Fernández (2019) observes that several of these (e.g. horse-riding, weapon-wielding) position the giving partner as dominant over the receiving partner, who is characterised as inferior or lacking in agency. Moreover, he notes that the giving partner is traditionally and most often perceived to be a man. By contrast, a receiving male partner is seen to possess "a penetrated masculinity" (Pascoe, 2005, p. 329), reflecting a belief that "to be penetrated is to abdicate power" (Bersani, 1987, p. 212). That my data contained more LGBTQ+ slurs referencing anal sex which targeted the giving partner went against this expectation. Perhaps, when the sex is between two men, what would otherwise be acceptable (even encouraged) displays of male dominance become displays of deviance; Baker (2005) notes the common characterisation of non-heterosexual men as predatory in negative discourses of homosexuality. If so, there may also be a perception that the giving partner is the more active participant in the deviant act and therefore more 'responsible' for it, although such a view is problematic for precluding the agency and consent of the receiving partner. Alternatively, the targeting of the giving partner may also relate to the historical, homophobic perception of non-heterosexual men as predatory; an image more congruent with someone in the 'active' sexual role. I argue such perceptions would explain why I found that slurs relating to anal sex targeted the giving partner more frequently than the receiving one.

Returning to the tendency to form slurs as two-word compounds, 26 (81.25%) of the slurs referencing anal sex were formed in this way. These were distinctive from compound slurs referencing other sex acts, in their much higher tendency to invoke metaphor. Two recurrent target/source combinations were 'PENETRATOR as CRIMINAL' (e.g. 'arse-bandit', 'butt-pirate', 'turd-burglar') and 'PENIS as WEAPON' (e.g. 'backdoor warrior', 'bum-basher', 'shit-stabber'). Characterising people who have anal sex as criminals has a historical precedent, especially considering that the decriminalisation of sex between men in the United Kingdom was as recent as 1967. Therefore, the breaking of this taboo has carried extremely serious social and legal consequences within living memory, so it is unsurprising to still see the prominence of the criminality metaphor. Describing the penis as a weapon is also quite a common metaphor; Cameron (1992) observed the use of this imagery by heterosexual men to name the penis, commenting on the patriarchal conflation of sex, masculinity and warfare. Cameron also noted some non-derogatory, even positive uses of this metaphor (e.g. 'love pistol', 'passion rifle', 'purple-helmeted love warrior'), but I argue that it is significant that these uses were always in reference to penis-in-vagina sex, and still conflated sex, war, and masculinity as Cameron criticised. Both kinds of imagery lend further explanation to why the giving partner in anal sex might be more prominently targeted by LGBTQ+ slurs: metaphorically, they are the criminal and/or the weapon-wielder.

Oral-vaginal sex was the second most common sexual act referenced within the sexual acts semantic field, though it accounted for a smaller proportion of slurs in this category than references to anal sex. Alongside digital-vaginal sex, I consider LGBTQ+ slurs relating to oral-vaginal sex to predominantly target women who have sex with women. I therefore argue that the relative cultural erasure and invisibility of women who have sex with women (see Chirrey, 2007; Turner, 2014) may have contributed to the reduced number of references to these sex acts in my data. As one example of this erasure, while sex between men has a long history of illegality in the UK, sex between women has had the opposite problem in being given no status within the law at all; a statutory age of consent in the UK for sex between women was only introduced as recently as 2000. Though the law does not dictate what is and is not taboo, taboos are closely related to what is made illegal or otherwise prohibited (Allan, 2019; Hutton, 2019); the fact that sex between women has barely been recognised as sex in law until very recently suggests a denial or ignorance of its practice in wider society. Another possibility is that within the judgement of a hetero-patriarchal society, the degree to which certain acts are tabooed is linked to the degree to which they satisfy the male gaze. Gill (2008, p. 50) notes a contemporary proliferation of representations of lesbianism in advertising – "…sexualised display[s] designed primarily for heterosexual men, frequently drawing on well-established codes from pornography". It is possible that sex acts between women are less frequently made the subject of taboo language because they possess appeal to a socially privileged heterosexual male population, whereas those performed between men do not.

It was an unexpected finding that more references to fellatio did not appear, especially given Jay's (1992) specific highlighting of 'cocksucker' as a taboo term. Again, I consider this evidence that Jay's claim that 'cocksucker' is especially offensive is unreliable. Besides 'cocksucker', there were only two other LGBTQ+ slurs referencing fellatio: the similar 'cock-muncher' and the highly metaphorical 'self-confessed player of the pink oboe'. As with the references to anal sex, I consider these slurs to particularly target men who have sex with men; though sexual acts are always somewhat tabooed regardless of who is performing them, such acts (in this case, fellatio) are more permissible when performed between a woman and a man, due to heteronormativity. The inclusion of the phrase 'self-confessed' in the last example is particularly telling; Baker (2005) noted a link

between homosexuality and perceptions of shamelessness in the negative discourses he analysed – including use of the term 'self-confessed' – with the notion of shameless homosexuality necessarily implying that sexual activity between men is something to be ashamed of. Arguably, the idea of being self-confessed also implies a rejection of this shame, an attitude underpinning many reclamation efforts (see section 2.2.2. of this thesis). However, this is not to argue 'player of the pink oboe' itself is beginning to undergo a reclamation; if it were, its frequency of identification would have been much higher than a single response.

3.5.2.2. Gender Expression Semantic Field

As I illustrated in Tables 3.4. and 3.5., 26 (83.87%) of the slurs relating to gender expression indexed femininity, while only 5 (16.13%) indexed masculinity. This is a stark contrast, for which I argue there is not a sole explanation. However, all of the explanations that I provide relate to the punishment of femininity in those perceived to be men, and of masculinity in those perceived to be women. Again, one possibility is the erasure and relative invisibility of women who do not conform to heteronormative expectations (Chirrey, 2007; Turner, 2014), as I discussed earlier in relation to slurs referencing sex acts between women. It is therefore possible that slurs targeting masculinity in those perceived to be women were less frequent, because such individuals are themselves less frequently acknowledged by society. Another possibility concerns the concept of 'hegemonic masculinity': "[an] ideology based on a hierarchy of dominant alignment roles, especially men over women, but also men over other men" (Kiesling, 1998, p. 71). Importantly, hegemonic masculinity is the product of a system which "maintains, legitimises and naturalises the interests of powerful men while subordinating the interests of others, notably the interests of women and gay men" (Coates, 2011, p. 264). While hegemonic masculinity neither recognises nor approves of masculinity in those perceived to be women, it arguably punishes femininity in those perceived to be men more severely, as this behaviour threatens the dominance of masculinity (in particular masculine men) over femininity. Feminine men are therefore punished for failing to uphold this dominance, which may subject them to more prolific linguistic derogation. Additionally, under hegemonic masculinity, any display of femininity is seen as a 'marked' behaviour, while displays of masculine behaviour – even by people who are not men – are more likely to be regarded as 'unmarked' or neutral.

Referring to Table 3.4. in particular, I observed a subcategory of slurs indexing femininity: women's names (e.g. 'jessie', 'nancy', 'nancy-boy', 'martha', 'mary', 'trixie'). It is clear how invoking women's names can be used to negatively ascribe femininity to a man. However, the practice has its own history in the culture of gay and bisexual men that is worth comparing. Alongside pronoun-switching (e.g. using 'she'/'her'/'hers' instead of 'he'/'him'/'his'), gay men referring to other gay men using women's names has long been recognised as a historical feature of gay men's speech, for both secretive (preventing identification of a conversational subject) and social (establishing camaraderie through comedic and theatrical language) purposes (Hayes, 1976; White, 1980; Baker, 2008). The giving of a 'camp name' (typically a woman's name) was also an important feature of Polari, an underground language used primarily by British, working-class LGBTQ+ people, which peaked in use during the 1950s and 1960s prior to the decriminalisation of homosexuality in the UK in 1967 (Baker, 2008; Baker, 2019). An individual was 'christened' with their camp name rather than choosing it themselves, being applied to an individual by other people, although the giving of a camp name was used to signal group acceptance and inclusion rather than rejection (Baker, 2019). Though the practice is still observed today, its use as a means of anonymisation and identity protection has largely dissipated, due to the emergence of the Gay Liberation movement, the subsequent shift in

social attitudes to LGBTQ+ people, and the eventual recognition of the practice by mainstream society (Raban, 1973; Baker, 2008). It is difficult to establish whether the decision by gay and bisexual men to use women's names to identify each other was a direct response to women's names being used to derogate feminine men; existing literature does not make such a connection, although Baker (2019) suggests that the practice of giving a camp name still had an underlying sarcastic or mocking quality, which may indicate that it was more a reflection of hegemonic masculinity than it was a subversion. Nevertheless, I argue this history is relevant to the prevalence of women's names as LGBTQ+ slurs, and that it is still possible that the use of women's names among gay and bisexual men represented a reclamation project of sorts.

3.5.2.3. Abnormality/Deviance Semantic Field

I now consider the results presented in Tables 3.6. and 3.7., concerning slurs which reference a perceived abnormality or deviance. The most common shared reference was to someone exhibiting strange or socially unacceptable behaviour (e.g. 'freak', 'pervert', 'queer', 'weirdo'). To understand the prevalence of this, I refer to Becker's (1963) theoretical understanding of *deviance*. Becker (1963, p. 9) argued that "…deviance is not a quality of the act the person commits but rather a consequence of the application by others of rules and sanctions". In other words, no act is inherently deviant, but is marked as deviant in contrast to a socially-constructed norm, or set of norms. A means by which this deviance is often marked is *labelling*, which is itself an exercise of power; Becker (1963, p. 3) states that a focus for researchers should be on "who applies the label of deviant to whom", and what the consequences of this labelling might be for the person (or people) so labelled. Regarding this latter question, Lemert (1967), notes that this labelling itself leads to further stigmatisation and punishment, through the demarcation of individuals as 'other'. I argue that this understanding of deviance underpins all of the LGBTQ+ slurs I list in Table 3.6., but especially those referencing someone whose behaviour is seen to be unacceptable. This also makes sense with respect to the primary purpose of slurs to derogate those with less social status (see section 2.2. of this thesis).

There were two other subcategories of slurs referencing abnormality/deviance. Some slurs (e.g. 'bender', 'bent', 'twist' and 'warped') suggest malformation, juxtaposing the LGBTQ+ target against the idea of 'straightness', in reference to both physical form and to heterosexuality. Indeed, the use of 'straight' to mean 'heterosexual' itself perpetuates this metaphor of correct and incorrect physical form. A second and slightly less common subcategory indexed medical dysfunction (e.g. 'confused', 'ill', 'mentally ill', and 'sick'). This reflects a history (especially in the UK) of the medicalisation of sex, and of non-heterosexual sex in particular (Plummer, 1995). Lastly, I argue that these slurs characterise their target as lacking agency; the metaphor of "DIFFERENCE as SICKNESS" suggests an individual whose difference is caused by some external, corrupting force.

3.5.3. Further Observations on First Question Responses

In addition to the findings I discussed in sections 3.5.1. and 3.5.2., there were some individual findings from the slur-identification stage of the survey which stood out as significant. I observed 20 slurs (12.50%) specifically targeting members of the transgender community. Morphologically, several transphobic slurs were constructed as two-word compounds juxtaposing differently-gendered words, such as 'he-she' or 'ladyboy'. This related to a broader tendency for transphobic slurs to index confusion or deception with regard to gender, such as 'gender-bender' and 'trap'.

While there was only one instance of a slur which I felt explicitly targeted bisexual individuals ('trysexual'), a number of terms raised in responses drew on stereotypically biphobic tropes, such as 'cheater', 'confused' and 'greedy' (see Turner, 2008; 2014 for discussion).

Finally, I also observed a tendency among respondents to identify certain linguistic practices – rather than specific words or phrases – as causing otherwise acceptable words to function as LGBTQ+ slurs. For example, 3 respondents suggested that noun uses of adjectival identity labels (e.g. 'a transgender', 'the gays') should be regarded as slurs, or at least as derogatory. There is some precedent for this in existing research: it has been argued that noun uses position gender and/or sexual identities as both monolithic and character-defining, whereas adjectival uses recognise them as one aspect of an individual among many (Wierzbicka, 1986; Zwicky, 1997). Similarly, one respondent used their response to highlight that deliberately using the incorrect pronouns to describe a transgender person (e.g. referring to a transgender woman using masculine pronouns) could be considered as derogatory as using a slur word or phrase.

3.5.4. Discussion of Second Question Responses

I now turn my discussion to responses to the optional second question of my survey, which invited respondents to comment further on the LGBTQ+ slurs which they identified. While 'dyke' was the word most frequently identified as a slur overall (see Table 3.1.) and was discussed further in just under a fifth of all responses to the second question (see Table 3.8.), it was not the word that was most elaborated on by respondents. Instead, 'queer' (joint fourth in terms of frequency, see Table 3.1.) was most frequently discussed in responses to the second question, appearing in just under half of all second question responses (see Table 3.8.). Additionally, some words that were frequently identified as slurs across responses were very rarely discussed in responses to the second question, if indeed they appeared at all: 'fag', 'faggot' and 'tranny' being the clearest examples of this. These results are further explained in the discussion which follows, but it was clear from the start of this analysis that respondents were not providing further commentary for only the most frequently-identified LGBTQ+ slurs.

3.5.4.1. When Is a Slur Not a Slur?

A subtle but common finding across responses was the suggestion that the status of a word as a slur is not constant. As noted throughout section 2.1. of my literature review, this is consistent with the current belief among taboo language scholars that no word is taboo, insulting or inappropriate to all people in all situations (e.g. Allan, 2019; Burridge & Benczes, 2019; Culpeper, 2019; Jay, 2019b). In my second question responses, this was often signalled through the use of adverbs, particularly of 'sometimes' and '(not) always':

"Butch…is not always a slur"
"Some…are therefore not always slurs"
"…it is sometimes still used as a slur.."
"…words that have been and sometimes still are used as slurs"
"These are not always slurs…"

This observation was similarly signalled through the use of modal verbs, particularly 'can' and 'could':

"'Correct' terms e.g. 'gay' aren't slurs in themselves but can be used as such…"
"…it can still be seen as a slur…"
"[queer]…could still be a slur depending on context"

Two factors particularly affected judgements about whether a word was being used as a slur or not. The first, though not discussed very frequently, was word class: as also indicated by some first question responses (see section 3.5.3.), two second question responses suggested that some words (e.g. 'gay' or 'queer') were unacceptable when used as nouns (e.g. "a gay", "the queers") as opposed to adjectives. The second – and much more frequently commented upon – was what respondents tended to refer to as 'context'. One particularly striking comment from a respondent was that "context, performance and delivery are all-important in coding [words] as insults". These comments are consistent with the current thinking in taboo language research that I outline in section 2.1. of my literature review (Chapter 2).

A key line of analysis, then, concerned what respondents meant by 'context' in these instances. Four respondents raised the idea of geographical differences:

"I'm Scottish and I think Jessie may be a Scottish term"
"it will depend on…location"
"…my experience varies in different parts of the country…"
"Chichi/cheechee/chichiman I heard from members of London's African Caribbean community…"

I consider this discussion to reflect the importance of community membership to tabooness judgements, as I discussed in section 2.1.2. of my literature review. Notably, the final example also identifies intersections of gender, sexual identity and race as relevant to tabooness judgements. This particular comment therefore raises two further considerations for my research. First, despite restricting my research sample to native speakers of British English in order to control variation, experiences of LGBTQ+ slurs can still differ regionally among British English speakers. Second, I believe this example stresses the need to approach taboo language research with an appreciation of the intersections that can exist between multiple aspects of a person's identity and their lived experiences. As noted by Besnier and Philips (2014, p. 133), 'individual' parameters of difference such as gender and sexuality are "variously inform[ed], reinforce[d], or def[ied]" by multiple other parameters of difference, including but not limited to factors like age, race and class. I argue this also justifies my theoretical and methodological decision to avoid overly-essentialising identity in my research, as I discuss in Chapter 1 of the thesis.

Four respondents also made specific mention of 'intent' as a factor, which I argue supports my discussion of the same in sections 2.1.3 and 2.2.3. of Chapter 2:

"it will depend on…their intention"
"whether or not the words are slurs often depends on who uses them, when and why"
"…I see them as slurs (when intended as such)"
"…I would not consider them slurs unless they were clearly intended for that purpose".

As signals of intent, aspects of prosody and use within discourse seem to be recognised as important too:

"…performance and delivery are all-important in coding them as insults"
"For the ones in some situations, it's often the tone used behind them"
"Of course, some of these words depend on context and tone of voice"
"I also feel like a slur is implied through conversations..."

Despite the consideration respondents gave to speaker intent, there was a broader line of discussion in second question responses which was often felt to supercede it in the case of LGBTQ+ slurs: the relationship between a speaker's identity and the perceived acceptability of their slur use.

3.5.4.2. The Positioning of Identity with Regard to 'Acceptable' Uses of Slurs

When discussing the importance of intent behind slur use, one respondent mentioned slurs they identified being used 'ironically' by people outside of the LGBTQ+ community – in their words, by "straight middle-class men". They suggested – with apparent disapproval – that such individuals would deny that such uses qualify as slurs, but that to the respondent, they are considered "distinctly othering". This response reflects a particular tension concerning community membership that was identified in a number of comments from respondents. I counted 8 instances where respondents had made explicit connections between the identity of a speaker and the perceived acceptability of them using particular words, with one such example stating that "…if the terms are used by someone outside the LGBTQ+ community, they are more likely to be interpreted as a slur". I suggest that the use of "more likely" in this comment is important for suggesting that while clearly important, speaker identity is not wholly or consistently a determiner of slur status, and that the factors mentioned earlier in this discussion are also relevant. However, I also make the observation that no second question responses suggested that speaker identity *did not* matter.

This comment provides a clear link to social identity theory, ingroups and outgroups, and the intergroup status hierarchy (Tajfel & Turner, 1986; Giles & Giles, 2013), as discussed in section 2.1.2. of my literature review. Using this theory as a lens, such comments directly relate the perception of language as derogatory and oppressive to the privileging of certain social groups over others. I therefore argue that when responses suggested LGBTQ+ ingroup membership was necessary for acceptable use of LGBTQ+ slurs, that also signalled a perception that LGBTQ+ slurs are used to subordinate the LGBTQ+ ingroup to a cisgender heterosexual outgroup. Second, I argue such comments therefore frame the use and reclamation of LGBTQ+ slurs as an act of resistance. Finally, I argue that such comments suggest it is not considered possible for perceived outgroup members – occupying a more privileged position within the intergroup status hierarchy – to participate in this resistance. As such, their slur use is only likely to be perceived as an oppressive exercise of hegemonic power.

This overall argument is strengthened by the fact that some survey responses suggested this difference in intergroup status was not only recognised by LGBTQ+ ingroup members, but by cisgender heterosexual outgroup members as well. Furthermore, respondents also indicated that some cisgender heterosexuals explicitly avoid using LGBTQ+ slurs for this precise reason. Two respondents, for example, made specific mention of their heterosexual identity, with one such respondent stating that they "would use none of [the] terms" that they had identified because of it.

3.5.4.3. Reclaiming Language

To return to one of the findings illustrated in Table 3.10., it was significant that 'queer' was the word that most frequently occurred in responses to the second survey question. I identified one possible reason for this by exploring second question responses for clear references to linguistic reclamation. I counted 23 references to reclamation in responses to the second question, 16 of which (69.57%) used the terms 'reclaim(-ed/-ing)' or 'reclamation', and 7 of which (30.43%) described the process in other words, e.g. "used by the LGBTQ+ community to self-identify" or "use it as an identity rather than a slur". This is a valuable finding in its own right, as it demonstrates that there is some general awareness of linguistic reclamation as a phenomenon, and that it seems common to use the term 'reclamation' to describe it.

Of those 23 references, 16 (69.57%) were found to discuss reclamation with specific regard to the word 'queer'. This number included the majority (but not all) of the 16 respondents who had explicitly referred to 'reclamation', suggesting a specific social awareness of reclamation in relation to the word 'queer'. While other slurs were also raised as examples of reclamation (e.g. 'dyke', 'gay', 'faggot' and 'tranny'), I focus the rest of this discussion on 'queer', for the fact that it was centred in the majority of responses concerning reclamation, and for its particular social and theoretical importance (see section 2.2.4. of my literature review for discussion).

Importantly, while respondents frequently highlighted 'queer' as a key example of linguistic reclamation within the LGBTQ+ community, they often qualified its reclamation to some degree: the word was "becoming a reclaimed term", "pretty reclaimed", "somewhat reclaimed", subject to "partial reclaiming" etc., rather than simply "reclaimed". Similarly, I counted 11 instances where 'queer' was discussed in relation to broader personal opinion/experience rather than to reclamation specifically. Many of these discussions emphasised strictly personal judgements of whether the word should be considered offensive or a slur, through use of words and expressions like "personally…", "my understanding…", "I see…", "I think…", "I don't feel…".

Though clearly not reclaimed to the extent that it is no longer considered offensive at all within the LGBTQ+ community, responses identified a wide variety of reclaimed uses for 'queer'. A number of comments spoke in general terms about its use as an identity marker, with 2 examples noting that the word was being used "…by the LGBTQ+ community to self-identify" and as a "legitimate [word] for people and behaviours and orientations in the LGBTQ community".

Some comments from respondents were more specific as to the identities that 'queer' can describe. One response gave two such examples: "queer folk", reflecting the use as an "umbrella term" for gender and sexual minorities as described by Livia and Hall (1997); and within the term "genderqueer", referencing what Baker (2008) identified as flexibility, fluctuation and/or resistance to strict categorisation with respect to gender. Another response indicated that the word can serve both of these purposes at the same time: "[I] personally reclaim queer as an umbrella term as my sexuality is fluid and [I] don't feel like I can isolate it to one specific term".

Despite having a variety of reclaimed uses, I observed that 'queer' was no less the subject of discussions about legitimate and illegitimate use. One respondent remarked that they knew 'queer' "…has been used by cishets [cisgender heterosexuals] to be derogatory", while another expressed a clear judgement that the reclamation of 'queer' (and also of 'dyke') "by their communities…doesn't mean that it's ok for people outside the community to use it". Such comments strongly recall the ingroup/outgroup dichotomy discussed previously. The use of the blend word 'cishets' in one response is particularly interesting in this regard; it exhibits the same concerns with noun phrase

identity labels discussed earlier in sections 3.5.3. and 3.5.4.1., but applied to a socially dominant rather than marginalised group. Importantly though, responses also suggested that LGBTQ+ ingroup membership alone does not entirely alleviate the tensions surrounding 'queer'. One LGBTQ+ respondent said that they "vehemently disagree" with the reclamation of 'queer', describing the word as "still very much a slur" to people who, like them, are older and from a working-class background. This tension is acknowledged in the literature; Zwicky (1997, p. 23) identifies a "generational clash…with many older speakers finding it irredeemable and many younger speakers preferring it to *gay*" [italics in original]. This reflects two points of discussion from my review of the literature (Chapter 2): that perceptions of a word's tabooness can be subject to sociolinguistic changes over time (sections 2.1.3.) and that individuals can simultaneously belong to multiple ingroups which might jointly affect their tabooness judgements (section 2.1.2. and 2.1.3.). Referring again to the importance of intersectional perspectives that I discussed in section 3.5.4.2., it should therefore be acknowledged that interactions between the multiple ingroups to which an individual belongs can lead to very different experiences of and attitudes towards LGBTQ+ slurs, including both their derogatory and reclaimed uses. This is an important insight for understanding individual responses to LGBTQ+ slurs and their reclamation, but as I go on to demonstrate in Chapter 4, it should not be seen to detract from the existence of broader patterns in attitudes towards the same.

3.6. Conclusion

The study I describe in this chapter was a necessary and valuable first step in my research, which yielded a number of useful insights in its own right. The first of these was simply discovering which words can be identified as LGBTQ+ slurs. I have explained that within taboo language processing research especially, there has been a failure to properly include and consider this language. When LGBTQ+ slurs have been included, they have typically only appeared as a small subset of a much larger list of stimuli, have rarely been identified as slurs, and their inclusion has rarely been sufficiently explained or justified. This is a problem which I have successfully avoided in my own research: from 115 survey responses, I identified 160 unique LGBTQ+ slurs. Even though I will not use all of these in my experiments to follow, I have evidenced that all would be considered LGBTQ+ slurs, and can now make evidence-based judgements about which to include.. Additionally, the results of my survey have demonstrated that the number of LGBTQ+ slurs in contemporary British English is much greater and diverse than the literature might suggest.

My analysis of all of the LGBTQ+ slurs identified by respondents was valuable for two main reasons. I was able to identify the most recognisable slurs and some associated issues with these, in particular: that 'dyke' was the most frequently identified LGBTQ+ slur, that both 'fag' and 'faggot' were also highly frequent but appeared to differ from each other, and that 'gay' can be considered a taboo word or slur but only when being used pejoratively. My analysis of first question responses also illuminated the specific social issues and taboos that LGBTQ+ slurs tend to draw upon. I found that semantic fields of sexual acts, unconventional gender expression and ideas of abnormality/deviance were particularly relevant to LGBTQ+ slurs.

Finally, I was able to use additional comments from survey respondents to provide a more in-depth understanding of some of the LGBTQ+ slurs I am investigating. Respondents appeared to support the current view of taboo language scholars that perception of a word as taboo (and by extension as a slur) is not universal or constant. I argue in particular that speaker identity, listener identity, speaker intent and the relationships of these factors to social power hierarchies are critical to understanding whether a particular use of a word is perceived to be a slur or not. I stress that

certainly at the level of the individual, this is not purely a question of whether or not one identifies as LGBTQ+; other factors like class, age and race are also relevant, particularly with regard to the ways these identities interact and intersect with gender and sexuality. Importantly, responses suggested that these factors are just as key to understanding attitudes to linguistic reclamation as they are to understanding perceptions of offensiveness and tabooness. Responses also particularly highlighted the word 'queer' as a case of reclamation, with a number of comments indicating both a rich variety of reclaimed uses, yet also a clear underlying tension surrounding this phenomenon.

In the next chapter, I describe how I used the findings of this study as the basis for another survey, which gathered ratings for emotional and non-emotional properties and measures of reclamation behaviour for LGBTQ+ slurs. This survey, in turn, was for the purposes of the lab-based taboo language processing experiments which I describe in Chapters 5 to 7 of my thesis.

Chapter 4 – Word Norms and Measures of Linguistic Reclamation for LGBTQ+ Slurs

4.1. Introduction

This chapter focuses on the second stage of my research, in which I used a survey to produce a set of normed ratings for LGBTQ+ slurs from British English. Respondents to the survey rated slurs for emotional and non-emotional word properties relevant to taboo language processing, and for measures of reclamation behaviour. The resulting database is novel, both for its specific focus on slurs and for its inclusion of quantitative data on slur reclamation. I was able to use this data to reveal a consistent pattern of relationships between particular word properties and reclamation behaviours, as well as noteworthy differences in ratings depending on respondent age, gender, and sexual identity. These findings are valuable in their own right, but also provide me with measures of these word properties and reclamation behaviours, which I will use to analyse their impact on the cognitive processing of LGBTQ+ slurs in the experimental studies described in Chapters 5 – 7.

4.2. Research Background

Databases of word norms – averaged, self-report ratings of linguistic stimuli for particular properties – are highly valuable to experimental studies of language; they provide standardised measurements for use as independent variables, enable comparisons between different properties of words and allow for greater replicability of experimental research within and across research labs (Bradley & Lang 1999).

In the case of databases of affective word norms – those concerning emotional measures like valence and arousal – Braun, Goudbeek, & Krahmer (2020, p. 1) note a potential limitation: they can assume individual words have an inherent affective meaning, despite findings from corpus and cognitive linguistics suggesting that words can acquire positive or negative connotations from frequent collocation with other words, which can be retained out of context (Louw, 1993; Stubbs, 1995). Evidence from their own study showed that valence ratings differed when words were rated with and without contextualisation in a sentence, and that when presented with context, valence ratings of individual words shifted considerably toward the valence rating of the sentence containing them. Braun et al. therefore advise that affective ratings taken for individual words not be regarded as essential truths, but rather as an "attempt at determining an "*a priori* affect" (2020, p. 2), which captures generalised patterns of valence and other aspects of emotionality across a language and across contexts. They argue that it is in this regard that databases of affective ratings have been proven useful and advantageous in experimental research in particular, as Bradley & Lang (1999) observed. They also recommend creating databases of affective word norms tailored to the specific needs of the research questions, rather than assuming generalised databases of affective ratings will apply equally to all studies.

With this in mind, I observed in the previous chapter (section 3.2.) that much of the existing norming data for taboo words is insufficient for experiments seeking to focus specifically on LGBTQ+ slurs. Similarly, I noted that LGBTQ+ slurs are frequently absent from the stimuli lists of existing taboo language experiments (e.g. Belleza, Greenwald, & Banaji, 1986; Geer & Bellard, 1996; MacKay, Shafto, Taylor, Marian, Abrams, & Dyer, 2004; Madan, Schafer, Chan, & Singhal, 2017). Combined with the fact that no existing norming data includes measures of linguistic reclamation for taboo words, these factors motivated me to produce a novel database of normed word property and reclamation behaviour ratings for LGBTQ+ slurs, which could inform the experimental stages of my research.

Though not able to provide all of the data required for my own studies, existing norming data for taboo words were extremely useful in identifying the most important word properties for which to gather ratings. One of the most comprehensive of these was produced by Janschewitz (2008), whose study produced normed ratings for taboo and non-taboo words for the following properties: word familiarity, frequency of personal use, valence, arousal, offensiveness, tabooness and imageability. The relevance of each of these word properties to performance in taboo language experiments has since been proven out in the literature. As discussed in section 2.3.2., Madan et al. (2017) found that a combination of 'non-emotional' properties (e.g. familiarity, personal use) and tabooness best explained taboo language performance in lexical decision tasks, while 'emotional' properties like arousal, valence, offensiveness and tabooness best explained performance in the free recall tasks. Their study also included normed Age of Acquisition ratings (henceforth AoA; the age at which a word is first learned) as one of the non-emotional properties relevant to lexical decision performance, based on research conducted after the production of Janschewitz's norming study (Brysbaert & Cortese, 2011; Juhasz, Yap, Dicke, Taylor, & Gullick, 2011).

Therefore, this chapter describes a study I conducted in order to produce my own database of word norms for LGBTQ+ slurs, which gathered averaged ratings for their non-emotional (familiarity, personal use, AoA) and emotional (arousal, valence, offensiveness, tabooness) properties, as well as measures of their reclamation. The next section outlines the sample, design, procedure and analytical approach of this study.

4.3. Methodology

4.3.1. Sample

The survey had 155 respondents. Of these, 127 respondents were recruited through social media, particularly through the social media pages of LGBTQ+ groups based in the North-West and East Midlands of England. The remaining 28 were a combination of students at the University of Nottingham – recruited either as a means of gaining module credit or via posters advertising the study – and members of the public recruited via posters placed in local community centres, with permission. To aid recruitment, respondents were offered a chance to enter a prize draw for a small online shopping voucher upon completion of the survey, excepting the students participating for module credit instead. No contact information was retained once the winner had been selected.

All respondents were required to be over the age of 16 and a native or near-native speaker of British English. I collected additional respondent demographic data in order to make group comparisons based on age group, gender, cisgender status (i.e. whether their gender matches the one they were assigned at birth) and sexual identity (see Table 4.1.). For the gender and sexual identity questions, respondents were able to give free and open responses, from which I created categories for analysis.

Descriptive analysis indicated that a majority of respondents were aged between 16 and 29, a slight majority identified as women, a large majority of respondents were cisgender and a slight majority of respondents were not heterosexual.

4.3.2. Survey Design

Despite the LGBTQ+ slur identification survey generating 160 unique slurs, a majority of these (51.85%) were only identified by single respondents. As a result, I decided to only include those which had been identified by a minimum of 5 respondents in the preliminary survey; this threshold allowed me to eliminate the most low-frequency slurs, while still leaving enough to use in my lab experiments (Chapters 5 – 7). As I discuss later in this section, respondents to the norming survey

Table 4.1. – Overall Respondent Demographic Data

Age Group	*N*	%	Gender	*N*	%	Cisgender?	*N*	%	Sexual Identity	*N*	%
16-19	36	23.23	Woman	82	52.90	Yes	140	90.32	Straight/Heterosexual	69	44.52
20-29	60	38.71	Man	66	42.58	No	15	9.68	Gay/Lesbian/Homosexual	43	27.74
30-39	17	10.97	Non-Binary	4	2.58	Total	155	100.00	Bisexual/Pansexual/Polysexual	31	20.00
40-49	15	9.68	Agender	1	.65				Queer	5	3.23
50-59	17	10.97	Unclear Response	2	1.29				Asexual	4	2.58
60-69	9	5.81	Total	155	100.00				Unclear Response	3	1.94
70+	1	.65							Total	155	100.00
Total	155	100.00									

were asked not to provide ratings for slurs they did not know; none had an excess of non-responses, demonstrating the threshold was not too low. Accordingly, 41 slurs were included in the norming survey (see Table 4.2).

Table 4.2. – LGBTQ+ Slurs Meeting Frequency Requirement

Slur	*N*	Slur (cont.)	*N*
Dyke	86	Shirt-Lifter	12
Faggot	83	Bent	10
Tranny	78	Batty	8
Fag	74	Bummer	8
Queer	74	Butch	8
Poof	69	Gender-Bender	8
Homo	38	Ladyboy	8
Lesbo	38	Trap	8
Lezzer	37	Cock-Sucker	7
Gay (Pejorative)	25	Gaylord	7
Bender	23	Muff-Diver	7
Gayboy	23	Bull-Dyke	6
Batty-Boy	22	Bum-Boy	6
Poofter	20	Pervert	6
Fairy	19	Camp	5
Shemale	19	Hermaphrodite	5
He-She	18	Queen	5
Rug-Muncher	16	Shit-Stabber	5
Fudge-Packer	15	Sissy	5
Pansy	14	Sodomite	5
Arse-Bandit	12		

The first set of survey questions asked respondents to rate LGBTQ+ slurs for 7 word properties identified by both Janschewitz (2008) and Madan et al. (2017): familiarity, personal use, arousal, valence, offensiveness, tabooness and AoA. Although included as a word property in both Janschewitz (2008) and Madan et al. (2017), I decided not to ask for imageability ratings. Because my study was focusing explicitly on highly offensive expressions, I anticipated that respondents may be unwilling to provide an answer if asked how easily they could picture a target of such language. Considering some of my participants could themselves be targets of such language, I also considered this to be a potentially unethical question to ask due, to the degree of discomfort it could cause.

I decided to ask respondents for negative valence ratings specifically, for four key reasons. Firstly, slurs being highly derogatory words by definition (Chen, 1998; Croom, 2011) means their

valence ratings are very likely to skew negative, and if participants felt a word was not at all negative, they could provide a rating to indicate this. Secondly, more precise ratings could be given if all points on the scale related to just negative valence, rather than to both positive and negative valence. Thirdly, this decision kept the formatting of the rating scales more consistent across the survey. Finally, the types of words to which I intended to compare slurs in my lab experiments (Chapters 5 – 7) would all have normed ratings indicating negative valence, so this decision ensured valence measurements would be more consistent across all experimental stimuli.

Regarding the wording of questions, I deliberately avoided describing any of the words as slurs, as I felt doing so may prime participants to rate words as more arousing, negative, offensive, and taboo. I took care to describe word properties in a way that would be accessible, without losing focus on what they represented. For example, the arousal question asked respondents how shocking they considered a word to be. I acknowledge that this is not a perfect translation of what arousal – as an unconscious emotional reaction – represents, but considered it the best word choice for the purpose of making a conscious, self-reported judgement of the same. I was careful to clearly distinguish properties that might be easily confused, such as offensiveness and tabooness. Here, respondents were simply asked how offensive they personally considered a word to be, and how frequently they would expect a word to be met with disapproval, respectively.

The second set of survey questions asked respondents to rate the same slurs for measures of reclamation behaviour. Because no existing taboo language studies have included these, I needed to create my own. In approaching this, I was strongly guided by the literature on linguistic reclamation. Particularly, I was mindful of Brontsema's (2004) observations that within communities targeted by a particular slur, reclamation of that slur may be practised by some members and not by others, but most members will have an awareness of the process taking place. I was also informed by an understanding that some slurs can be reclaimed in reference to both oneself and others, as in the case of the word 'queer' (Baker, 2008; Livia & Hall, 1997; Stephens, 2011). As such, I decided to gather ratings for four different contexts of reclaimed use: positive uses of slurs by individuals in reference to themselves (self-self reclamation), by individuals in reference to others (self-other reclamation), by other individuals in reference to themselves (other-self reclamation) and by other individuals in reference to others (other-other reclamation). To help ensure that respondents were aware of the different contexts of use being specified for the reclamation questions, bold text was used to highlight the relevant speaker and referent, as well as to make clear that respondents were only being asked about positive uses of the words. I argue that approaching the measurement of reclamation in this way also enabled me to ask questions about whether certain contexts of reclaimed use are more commonly experienced than others, and therefore may be better predictors of performance in my experiments.

Some word property questions required specific adjustments to be made and communicated to respondents. In general, respondents were asked to leave a word unrated if they had never heard of it before, so that they would not give misrepresentative data. However, this was not the case for the familiarity question in particular, as instances of total unfamiliarity were relevant to the word property. For just the self-self reclamation question, respondents were provided with an additional response option to declare that they did not consider that a word applied to them at all. I reasoned that if respondents were reporting never using a slur to describe themselves positively – as the question asked – because it simply could not refer to them, this was different to never reclaiming a slur because a respondent chose never to do so. I also wanted to enable respondents to indicate whether they felt a word applied to them or not, rather than making assumptions based on their demographic information (see Chapter 1, section 1.4.).

Initially, all ratings were to be given on a 1-9 Likert scale, in keeping with the length of scale used by Janschewitz (2008). However, survey piloting suggested that respondents found these scales

too long. Accordingly, I reduced all scales to a 1-5 Likert scale instead. Piloting also revealed that – due to the number of slurs in the study and the number of word property ratings needed for each – when the survey included the full list of 41 slurs, it was too time-consuming and discouraged completion. To resolve this, I split the survey into two versions, such that one version was randomly assigned 20 of the slurs and the other contained the remaining 21. These versions were then alternated between respondents, with Version 1 receiving 78 responses and Version 2 receiving 77. As illustrated in Tables 4.3. and 4.4., the distributions of respondents according to all four demographic categories were similar in both versions of the survey, both when compared to each other and to the overall demographic spread presented in Table 4.1. As such, I judged that the responses to both versions could still be combined for analysis. For illustration, I provide a full example version of the survey in Appendix 3.

Table 4.3. – Version 1 Respondent Demographic Data

Age Group	***N***	**%**	**Gender**	***N***	**%**	**Cisgender?**	***N***	**%**	**Sexual Identity**	***N***	**%**
16-19	16	20.51	Woman	42	53.85	Yes	69	88.46	Straight/Heterosexual	32	41.03
20-29	34	43.59	Man	33	42.31	No	9	11.54	Gay/Lesbian/Homosexual	21	26.92
30-39	7	8.97	Non-Binary	1	1.28	Total	78	100.00	Bisexual/Pansexual/Polysexual	20	25.64
40-49	8	10.26	Agender	1	1.28				Queer	1	1.28
50-59	8	10.26	Unclear Response	1	1.28				Asexual	2	2.56
60-69	4	5.13	Total	78	100.00				Unclear Response	2	2.56
70+	1	1.28							Total	78	100.00
Total	78	100.00									

Table 4.4. – Version 2 Respondent Demographic Data

Age Group	***N***	**%**	**Gender**	***N***	**%**	**Cisgender?**	***N***	**%**	**Sexual Identity**	***N***	**%**
16-19	20	25.97	Woman	40	51.95	Yes	71	92.21	Straight/Heterosexual	37	48.05
20-29	26	33.77	Man	33	42.86	No	6	7.79	Gay/Lesbian/Homosexual	22	28.57
30-39	10	12.99	Non-Binary	3	3.90	Total	77	100.00	Bisexual/Pansexual/Polysexual	11	14.29
40-49	7	9.09	Agender	0	0.00				Queer	4	5.19
50-59	9	11.69	Unclear Response	1	1.30				Asexual	2	2.60
60-69	5	6.49	Total	77	100.00				Unclear Response	1	1.30
70+	0	0.00							Total	77	100.00
Total	77	100.00									

I encountered one unexpected issue with splitting the survey: after completion, some respondents expressed concern that they had not seen some slurs which they strongly expected to have seen, or had seen more slurs pertaining to certain members of the LGBTQ+ community than to others. This was due to slurs being randomly assigned to each version of the survey and to respondents not being aware that there was more than one version. Although I was able to explain to respondents that they had only seen half of the slurs used in the full study, if future studies required a survey to be similarly split into multiple versions, I would adjust the distribution of slurs following random assignment to ensure more equal representation.

4.3.3. Procedure

The survey was conducted online. The front page provided respondents with an information sheet, detailing the aims of the research and outlining the participation process. Respondents had to tick a box to indicate they had read this information before they could proceed. On the next page, they had to read a statement of informed consent and tick a box to indicate their agreement with it.

Respondents were then asked to provide demographic information, before responding to each slur rating question. Instruction pages with example illustrations were provided where appropriate to assist with completion. Following completion, eligible respondents were given the option to provide a contact email address to enter the shopping voucher prize draw, after which they could save their results and close the survey.

4.3.4. Data Analysis

My data analysis had 3 stages. First, producing a database of normed word property ratings and reclamation behaviours for every LGBTQ+ slur in the study. Second, exploring relationships between word properties and reclamation behaviours across the normed database. Third, identifying any demographic group differences in word property and reclamation behaviour ratings. All data analysis was conducted using IBM SPSS Statistics v.26. In this section, I outline the analytical procedures used for each of these stages. All participants were found to have met the inclusion criteria and completed the survey correctly, so at no stage of my analysis was any participant's data excluded.

4.3.4.1. Producing the Database

The procedure for creating the database was straightforward. Each of the 41 LGBTQ+ slurs in the study received a group of ratings for each of the word properties and reclamation behaviours. Depending on whether a slur was featured in Version 1 or Version 2 of the experiment, it received a maximum of 78 or 77 ratings for each property/behaviour. The total number of ratings that each slur received varied slightly, as respondents did not provide ratings for any slurs they were completely unfamiliar with, but any missing ratings were excluded from analysis. Overall, both versions of the survey generated a combined total of 34,949 ratings.

Using this data, I calculated the mean rating for every property and behaviour, for every slur. I then computed three additional word properties from the mean reclamation behaviour data: *self-reclamation* (mean self-self reclamation + mean self-other reclamation / 2), *other-reclamation* (mean other-self reclamation + mean other-other reclamation / 2) and *overall reclamation* (mean self-self reclamation + mean self-other reclamation + mean other-self reclamation + mean other-other reclamation / 4). I produced these combined ratings so that I could make broader comparisons between self-referential and other-referential reclamation behaviours, as well as to an overall measure of reclamation behaviours.

In total, 574 mean ratings were generated (41 LGBTQ+ slurs x 14 word properties). These were then compiled into a complete database (Appendix 4).

4.3.4.2. Correlation Analyses

Because all word property and reclamation behaviour ratings were ordinal, I chose Spearman's rank-order correlation analysis (Field, 2013; Spearman, 1904) as the best statistical procedure for identifying correlations between them. Each property/behaviour had a potential correlation with 13 others, so I ran 182 Spearman's correlation analyses in total. Half of these were not unique (e.g. familiarity x personal use was the same as personal use x familiarity), meaning I tested 91 unique correlations in total.

For all significant correlations, I carried out two further validity checks. First, I produced a scatterplot of the relevant data in SPSS, in order to confirm by visual inspection that the relationship was monotonic (i.e. if there was a positive or inverse correlation). Second, I calculated upper and lower confidence intervals for the value of the r_s coefficient, to check that the range between the two did not pass through 0, thereby ensuring there was no uncertainty as to the value of the coefficient being positive or negative. In all cases, these procedures showed that all the relationships

were monotonic, and that none of the ranges between upper and lower confidence intervals passed through 0.

Though not an indication of causality, I also determined how close the relationship between two particular properties/behaviours was, by calculating as a percentage how much of the variance was directly shared between those which correlated significantly. This is achieved by calculating the r_s^2 value – the square of the value of the r_s coefficient (Field, 2013).

4.3.4.3. Demographic Group Analyses

I selected the Mann-Whitney U test – a nonparametric equivalent to the Analysis of Variance (ANOVA) test – to identify whether there were group differences between word property and reclamation behaviour ratings according to respondent demographic data. As with the correlation analyses, I chose this test because all of the ratings were measured on an ordinal scale (Lehmann, 2006; Mann & Whitney, 1947).

Each property/behaviour rating for each LGBTQ+ slur was the dependent variable of its own Mann-Whitney U test. Three sets of these were conducted, in which the independent variable was either respondent age group, gender or sexual identity. Although I collected demographic data on whether or not respondents were cisgender, I could not make a statistically valid comparison between cisgender and non-cisgender respondents (see Table 4.1.). Overall, I conducted 1,722 Mann-Whitney U tests (41 slurs x 14 property/behaviour ratings per slur x 3 independent variables).

Within each independent variable, respondents were further divided into one of two groups. Groups were created from the demographic data presented in Table 4.1., adjusted so that there would be enough respondents within each to make valid comparisons. I only compared women and men for the gender comparison, as there were not enough non-binary respondents in the sample to constitute their own group (see Table 4.1.). Table 4.5. lists the groups created, and how many respondents belonged to each.

Table 4.5. – Adjusted Demographic Groups and Distributions

Age Group	***N***	**%**	**Gender**	***N***	**%**	**Sexual Identity**	***N***	**%**
16-29	96	61.94	Woman	82	55.41	Non-Heterosexual	79	53.38
30+	59	38.06	Man	66	44.59	Heterosexual	69	46.62
Total	155	100.00	Total	148	100.00	Total	148	100.00

An important preliminary analysis when running a Mann-Whitney U test is to check whether distributions of the two groups of the independent variable have a similar shape. Similarly-shaped distributions enable precise comparisons of the median values in both groups, while dissimilarly-shaped distributions only allow for comparison of mean ranks, which indicate whether values were generally higher or lower in one group compared to the other (Hart, 2001). I visually inspected distributions for all of the Mann-Whitney U tests I conducted, and found that none were similarly shaped. This was unsurprising, because the range of ratings that could be given for each word property was small. As such, all significant findings were based on comparisons of mean ranks between groups, rather than medians.

To clarify what mean rank values represent, a higher value indicates that ordinal ratings in that group were generally higher than in the other. For example, in Tables 4.8. – 4.11, I make some age group comparisons for particular LGBTQ+ slurs. The values in the mean rank columns do not denote a specific age nor rating, but represent whether ratings skewed comparatively higher or lower in that age group. For ordinal data such as mine, this information can be more useful than the

descriptive medians, especially for shorter rank scales; the medians for both groups may be identical if participants only have five options to choose from, but both groups may still have significantly differed in providing generally higher or lower ratings, which only mean rank values are able to indicate.

Regarding determining statistical significance, another important decision when conducting a Mann-Whitney U test is whether to use the exact or asymptotic *p*-value. The latter was more appropriate, for the main reason that the calculation for the exact *p*-value in SPSS does not correct for ties in the data, where two or more respondents provide identical values for the dependent variable. As all of my dependent variables were ordinal, I anticipated that such ties would be very common. I did not consider using the asymptotic *p*-value as opposed to the exact one a problem, as it is generally considered that when both groups of the independent variable have more than 20 cases (as mine consistently did), the asymptotic *p*-value is a good approximation to the real *p*-value (Dineen & Blakesley, 1973).

4.4. Results

4.4.1. Correlations between Word Properties and Reclamation Behaviours

The Spearman's correlation analyses revealed that the ratings for every word property and reclamation behaviour had a significant correlation with those of every other property/behaviour. Furthermore, two distinct groups emerged, in which properties/behaviours in the same group shared positive correlations with each other, but inverse correlations with those in the other group. Table 4.6. illustrates these findings. Because of the number of correlations observed, here I have only represented whether the correlation was positive or negative and the degree of statistical significance.

Table 4.6. – Correlations between Word Properties and Reclamation Behaviours

	Familiarity	Personal Use	Arousal	Negativity	Offensiveness	Tabooness	Age of Acquisition	Self-Self Reclamation	Self-Other Reclamation	Other-Self Reclamation	Other-Other Reclamation	Self-Reclamation	Other-Reclamation	Overall Reclamation
Familiarity		+ +	-	-	-	- -	- -	+ +	+ +	+ +	+ +	+ +	+ +	+ +
Personal Use	+ +		- -	- -	- -	- -	- -	+ +	+ +	+ +	+ +	+ +	+ +	+ +
Arousal	-	- -		+ +	+ +	+ +	+	- -	- -	- -	- -	- -	- -	- -
Negativity	-	- -	+ +		+ +	+ +	+	- -	- -	- -	- -	- -	- -	- -
Offensiveness	-	- -	+ +	+ +		+ +	+	- -	- -	- -	- -	- -	- -	- -
Tabooness	- -	- -	+ +	+ +	+ +		+	- -	- -	- -	- -	- -	- -	- -
Age of Acquisition	- -	- -	+	+	+	+		- -	- -	- -	- -	- -	- -	- -
Self-Self Reclamation	+ +	+ +	- -	- -	- -	- -	- -		+ +	+ +	+ +	+ +	+ +	+ +
Self-Other Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +		+ +	+ +	+ +	+ +	+ +
Other-Self Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +	+ +		+ +	+ +	+ +	+ +
Other-Other Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +	+ +	+ +		+ +	+ +	+ +
Self-Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +	+ +	+ +	+ +		+ +	+ +
Other-Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +	+ +	+ +	+ +	+ +		+ +
Overall Reclamation	+ +	+ +	- -	- -	- -	- -	- -	+ +	+ +	+ +	+ +	+ +	+ +	

+ = positive correlation, $p < .05$; + + = positive correlation, $p < .01$; - = negative correlation, $p < .05$; - - = negative correlation, $p < .01$.

Eighty-four (92.31%) of the 91 unique correlations found were significant at the $p < .01$ level. Seven (7.69%) reached statistical significance at the $p < .05$ level: the 3 inverse correlations of familiarity with arousal, negativity and offensiveness, and the 4 positive correlations of AoA with arousal, negativity, offensiveness and tabooness.

Regarding the groups indicated by these results: familiarity, personal use and all measures of reclamation behaviour correlated positively with each other. Similarly, arousal, negativity, offensiveness, tabooness and AoA all correlated positively with each other. Where a property/behaviour from one group correlated with a property/behaviour from another (e.g.

familiarity x tabooness), this correlation was always negative, typically reaching a high level of significance as well.

The results of the r_s^2 analysis, showing how much variance was shared between properties/behaviours which correlated significantly, are presented in Table 4.7.

Table 4.7. – r_s^2 Values for Significant Correlations

	Familiarity	Personal Use	Arousal	Negativity	Offensiveness	Tabooness	Age of Acquisition	Self-Self Reclamation	Self-Other Reclamation	Other-Self Reclamation	Other-Other Reclamation	Self-Reclamation	Other-Reclamation	Overall Reclamation
Familiarity		59%	14%	14%	14%	17%	58%	49%	67%	64%	66%	69%	66%	67%
Personal Use	59%		31%	36%	34%	30%	28%	58%	83%	72%	76%	81%	74%	77%
Arousal	14%	31%		85%	88%	83%	12%	34%	41%	42%	44%	38%	41%	42%
Negativity	14%	36%	85%		92%	85%	12%	28%	44%	45%	50%	41%	48%	46%
Offensiveness	14%	34%	88%	92%		83%	15%	31%	42%	40%	44%	41%	46%	42%
Tabooness	17%	30%	83%	85%	83%		12%	20%	34%	37%	44%	31%	38%	38%
Age of Acquisition	58%	28%	12%	12%	15%	12%		40%	42%	42%	41%	46%	44%	44%
Self-Self Reclamation	49%	58%	34%	28%	31%	20%	40%		71%	62%	58%	74%	69%	66%
Self-Other Reclamation	67%	83%	41%	44%	42%	34%	42%	71%		88%	86%	98%	92%	90%
Other-Self Reclamation	64%	72%	42%	45%	40%	37%	42%	62%	88%		94%	88%	90%	96%
Other-Other Reclamation	66%	76%	44%	50%	44%	44%	41%	58%	86%	94%		86%	88%	98%
Self-Reclamation	69%	81%	38%	41%	41%	31%	46%	74%	98%	88%	86%		94%	92%
Other-Reclamation	66%	74%	41%	48%	46%	38%	44%	69%	92%	90%	88%	94%		94%
Overall Reclamation	67%	77%	42%	46%	42%	38%	44%	66%	90%	96%	98%	92%	94%	

Considering these results alongside those presented in Table 4.6., I concluded that word properties tended to share more variance when they also shared a positive correlation. The only exception to this seems to have been AoA, but this may be explained by the fact that correlations between AoA and other properties were generally weaker in terms of their significancece (see Table 4.6.). The relationships between familiarity and arousal, negativity, offensiveness and tabooness were also quite weak, all sharing <20% of their variance with each other. Although it is unsurprising that self-reclamation, other-reclamation and overall reclamation shared >90% of their variance with the word properties they were each calculated from, it is worth noting that all reclamation behaviours shared >80% of their variance with each other, regardless of who was using the slur and in reference to whom. Overall, these results emphasise the strength of the relationships found between familiarity, personal use and reclamation behaviours, and between arousal, negativity, offensiveness and tabooness. Interestingly, the r_s^2 values in the latter group were particularly high.

4.4.2. Results of Demographic Group Comparisons

In this section of the results, I detail the findings of the Mann-Whitney U tests that compared every word property rating for each LGBTQ+ slur by age group, gender and sexual identity. In total, 378 of the 1,722 (21.95%) demographic comparisons conducted were significant. Because of the volume of significant findings, I have selected the most notable for consideration.

4.4.2.1. Age Group (16-29 vs. 30+) Comparisons

In total, I found 73 significant differences in word property and reclamation behaviour ratings as a function of age group. 24 (32.88%) of these were significant differences in AoA ratings. In all but one of these cases, the mean ranks were lower in the 16-29 group than in the 30+ group, indicating that AoA ratings were higher in the 30+ group for these slurs. As a reminder of the discussion in section 4.3.4.3., the values in the mean rank columns do not indicate the average age at which participants in that group reported acquiring a slur, but a higher value in one group indicates that group generally reported later acquisition of a slur than the other did.

Some of the LGBTQ+ slurs with the starkest age-related AoA differences fitting the pattern described above were also those with some of the highest familiarity ratings, as indicated in Table 4.8. These results suggest that slurs which are more widely recognised are also being acquired earlier by younger speakers.

Table 4.8. – Age of Acquisition Differences by Age Group for High-Familiarity LGBTQ+ Slurs

Slur	**Normed Familiarity Rating**	**Mean Rank (16-29)**	***N***	**Mean Rank (30+)**	***N***	***U***	***z***	***p***
Dyke	3.4	27.69	40	46.73	31	952.50	3.33	< .001
Fag	3.81	31.37	45	48.85	31	1018.50	3.73	< .001
Faggot	3.66	32.10	45	47.79	31	985.50	3.33	= .001
Tranny	3.63	31.40	49	52.30	28	1058.50	4.11	< .001

The presence of 'tranny' in this set of results is noteworthy, because none of the other properties/behaviours which differed by age group for this slur correlated with the AoA difference in the manner expected from Table 4.6., although they correlated with each other in accordance with this trend in every other respect. Table 4.9. outlines these results.

Table 4.9. – Other Age Group Differences for 'Tranny'

Word Property	**Mean Rank (16-29)**	***N***	**Mean Rank (30+)**	***N***	***U***	***z***	***p***
Arousal	44.55	49	29.29	28	414.00	-2.97	< .01
Negativity	44.21	49	28.13	27	381.50	-3.23	= .001
Offensiveness	44.53	48	28.16	28	382.50	-3.37	= .001
Personal Use	34.28	45	41.38	28	752.50	1.97	= .05
Self-Other Reclamation	36.84	49	42.79	28	792.00	2.01	.04

While Table 4.6. indicates that arousal, negativity and offensiveness typically correlated positively with AoA, so would therefore be expected to have higher ratings in the group that reported a later AoA for 'tranny' (30+), these results show that ratings for these properties were actually higher in the group that reported an earlier AoA (16-29). Similarly, Table 4.6. shows that personal use and self-other reclamation typically correlated negatively with AoA, so would therefore be expected to have lower ratings in the group that reported a later AoA for 'tranny', but the results in Table 4.9. show that ratings for these properties were actually lower in the group that reported an earlier AoA.

AoA was the only property of 'queer' that differed significantly by age group, but not in the same way as all of the other AoA age differences. 'Queer' was the only slur where AoA ratings were significantly lower in the 30+ group (N = 28, mean rank = 32.27) than in the 16-29 group (N = 50, mean rank = 43.55), U = 497.50, z = -2.30, p = .02.

Two related slurs – 'batty' and 'batty-boy' – also differed significantly for AoA by age group at the $p \leq .001$ level. AoA ratings for 'batty' were significantly lower in the 16-29 group (N = 36, mean rank = 20.43) than in the 30+ group (N = 24, mean rank = 45.60), U = 794.50, z = 5.66, p = < .001. Similarly, AoA ratings for 'batty-boy' were significantly lower in the 16-29 group (N = 34, mean rank = 19.66) than in the 30+ group (N = 24, mean rank = 43.44), U = 742.50, z = 5.47, p = < .001. Though they reached a particularly high level of statistical significance, I do not include this finding in my broader discussion section, as I argue there is a straightforward explanation for it: both terms come

from the Multicultural London English (MLE) *multiethnolect,* which itself developed from Jamaican Creole (Cheshire, Kerswill, Fox, & Torgersen, 2011; Clyne, 2000; Kerswill, 2014). The former has largely been observed in academic research in just the last decade (Cheshire et al. 2011; Kerswill & Sebba, 2011; Kerswill, 2014), while the latter was only observed in London in the mid-1980s (Hewitt, 1986; Kerswill, 2014; Sebba, 1993). Because both varieties of British English are relatively new, it is not surprising that older speakers first encountered them later in their lives than younger speakers did. This result could have been explored further had I also collected demographic information on respondents' ethnicity, as this may also have affected familiarity ratings with slurs from a multiethnolect like MLE.

Another property that particularly tended to show significant age group differences was familiarity, accounting for 16 (21.92%) of all age group differences in word property ratings. Twelve (75.00%) of these indicated less familiarity in the 16-29 group, while 4 (25.00%) indicated more familiarity in the 16-29 group. Tables 4.10. and 4.11. indicate which slurs were less and more familiar in the 16-29 group than in the 30+ group. Where certain LGBTQ+ slurs had significant age group differences for multiple word property ratings, their relationships to each other tended to be consistent with the general word property rating correlations in the data. However, there were some cases in which this was not true.

There were a few cases where slurs that were reported as being acquired at an earlier age by 16-29 year olds were also reported as being less familiar to the same group, which is inconsistent with the inverse correlation between these two properties across the entire dataset. This was true for 'arse-bandit', 'bull-dyke', 'fudge-packer' and 'rug-muncher'. I provide an explanation for these exceptions at the end of section 4.5.2.

Table 4.10. – LGBTQ+ Slurs with Lower Familiarity Ratings in 16-29 Group

Slur	**Mean Rank (16-29)**	***N***	**Mean Rank (30+)**	***N***	***U***	***z***	***p***
Arse-Bandit	30.18	46	52.08	31	1118.50	4.44	< .001
Bull-Dyke	32.10	46	49.24	31	1030.50	3.73	< .001
Bum-Boy	33.38	48	47.29	28	918.00	2.72	< .01
Cock-Sucker	35.86	50	46.00	28	882.00	1.96	= .05
Fairy	33.04	46	46.87	30	941.00	2.79	= .01
Fudge-Packer	32.69	50	56.11	28	1040.50	3.77	< .001
Lezzer	35.56	50	46.54	28	897.00	2.10	.04
Muff-Diver	33.14	50	50.86	28	1018.00	3.54	< .001
Pansy	34.58	46	45.16	31	904.00	2.10	.04
Poof	33.87	46	46.61	31	949.00	2.53	.04
Rug-Muncher	35.33	50	46.95	28	908.50	2.27	= .01
Shirt-Lifter	32.16	49	50.96	28	1021.00	3.95	< .001

Table 4.11. – LGBTQ+ Slurs with Higher Familiarity Ratings in 16-29 Group

Slur	**Mean Rank (16-29)**	***N***	**Mean Rank (30+)**	***N***	***U***	***z***	***p***
Fag	45.74	46	29.00	31	403.00	-3.37	= .001
Faggot	44.90	46	30.24	31	441.50	-2.93	< .01
Gay (pejorative)	43.24	46	32.71	31	518.00	-2.21	.03
He-She	43.32	50	32.68	28	509.00	-2.08	.04

4.4.2.2. Gender (Women vs. Men) Comparisons

In total, I found 118 significant differences in ratings as a function of gender. In 117 (99.15%) of these results, women provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings. In Tables 4.12. and 4.13., I provide the results for 'bent' and 'bum-boy', which illustrate this difference.

Table 4.12. – Gender Differences for 'Bent'

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***z***	***P***
Familiarity	32.43	42	45.09	33	927.00	2.57	= .01
Personal Use	29.74	35	38.66	32	709.00	2.28	.02
Negativity	36.88	32	28.12	32	372.00	-1.96	= .05
Offensiveness	37.73	33	28.12	32	372.00	-2.10	.04
Tabooness	37.65	33	28.20	32	374.50	-2.06	.04
Self-Reclamation	28.74	34	38.56	32	706.00	2.60	= .01
Overall Reclamation	28.82	34	38.47	32	703.00	2.27	.02

Table 4.13. – Gender Differences for 'Bum-Boy'

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***z***	***p***
Familiarity	31.75	40	43.36	33	870.00	2.39	.02
Personal Use	27.00	33	34.78	27	561.00	3.08	< .01
Offensiveness	35.56	31	23.84	28	261.50	-2.74	= .01
Tabooness	35.34	31	24.09	28	268.50	-2.62	= .01
Self-Other Reclamation	29.50	33	33.78	29	544.50	2.19	.03
Other-Self Reclamation	29.50	32	33.63	30	544.00	2.12	.03
Self-Reclamation	28.50	33	34.91	29	577.50	2.72	= .01
Overall Reclamation	28.79	33	35.53	30	601.00	2.29	.02

The singular exception to this finding (representing 0.85% of gender differences) was a lower arousal rating among women (N = 27, mean rank = 23.50) than among men (N = 28, mean rank = 32.34) for 'batty', U = 499.50, z = 2.11, p = .04. However, familiarity ratings for 'batty' were noticeably lower among women (N = 40, mean rank = 30.95) than among men (N = 31, mean 0rank = 42.52), U = 822.00, z = 2.43, p = .02. Although in all other cases a lower familiarity rating tended to present alongside a higher arousal rating, given that this was the only exception, it is possible that fewer women respondents found the word highly shocking because fewer women in the sample knew the word at all. As explained in section 4.3.2., all respondents had to provide familiarity ratings, but were asked not to provide further word property ratings for words which they simply didn't know. This possibility is further supported by the N difference in women's familiarity and arousal ratings: the difference in N count between familiarity and arousal suggests that 13 (32.5%) of the 40 women who were asked to rate 'batty' didn't provide an arousal rating for it, suggesting that they did not know the word. Another possibility is that there may have been some interference of the non-taboo sense of 'batty', meaning eccentric or strange.

With the exception of 'batty', I note that this pattern of gender differences is completely consistent with the correlations illustrated in Table 4.6. Indeed, in almost all cases where individual slurs exhibited significant gender differences in multiple property/behaviour ratings, the relationship between these differences matched the overall correlation pattern (see Table 4.14. for example). For

some slurs explicitly targeting LGBTQ+ women (e.g. *bull-dyke, lesbo, lezzer, rug-muncher*), there were no gender differences at all, which I discuss further in section 4.5.4.

Table 4.14. – Gender Differences for 'Muff-Diver'

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***z***	***p***
Personal Use	24.07	29	31.48	25	462.00	2.66	= .01
Offensiveness	27.36	22	18.83	23	157.00	-2.26	.02
Tabooness	28.41	22	20.12	25	178.00	-2.19	.03

A final gender comparison was the finding that, of the 10 results (8.47%) that reached statistical significance at the $p \leq .001$ level, 7 (70%) concerned one of the reclamation behaviours. All differences were for three of the LGBTQ+ slurs in the dataset: 'fag', 'faggot' and 'gayboy'. In all cases, the reclamation behaviours were consistently lower among women than among men. This is perhaps unsurprising, given that these slurs are each gendered so as to target men, but the degree of statistical significance is nevertheless striking. Tables 4.15. – 4.17. provide these results.

Table 4.15. – Gender Differences for 'Fag', p ≤ .001.

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***z***	***p***
Self-Other Reclamation	31.51	40	42.73	32	839.50	3.19	= .001
Self-Reclamation	31.36	40	42.92	32	845.50	3.28	= .001

Table 4.16. – Gender Differences for 'Faggot', p ≤ .001.

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***Z***	***P***
Other-Self Reclamation	30.96	41	43.82	31	862.50	3.34	= .001
Self-Reclamation	30.86	39	42.27	32	824.50	3.34	= .001
Other-Reclamation	31.06	41	43.69	31	858.50	3.21	= .001
Overall Reclamation	31.13	41	44.52	32	896.50	3.25	= .001

Table 4.17. – Gender Differences for 'Gayboy', p ≤ .001.

Word Property	**Mean Rank (Women)**	***N***	**Mean Rank (Men)**	***N***	***U***	***z***	***p***
Self-Reclamation	28.62	37	42.38	32	828.00	3.49	< .001

4.4.2.3. Sexual Identity (Non-Heterosexual vs. Heterosexual) Comparisons

In total, I found 187 significant differences in word property and reclamation behaviour ratings as a function of sexual identity. This meant that the sexual identity group comparisons had the highest number of significant differences in property/behaviour ratings. This was perhaps to be expected, given that all the words being rated were LGBTQ+ slurs.

One hundred and thirty-two (70.59%) of these were differences in reclamation behaviour ratings. In every case, the reclamation ratings were higher in the non-heterosexual group, regardless of the type of reclamation behaviour. Of these, 52 of these (39.39%) were significant at the $p \leq .05$ level, 34 (25.76%) were significant at the $p \leq .01$ level and 46 were (34.85%) significant at the $p \leq .001$ level. These results reflect existing thought on linguistic reclamation, in that they suggest the

non-heterosexual group is able to engage in the reclamation process to an extent that the heterosexual group is not. This might also explain why the majority of these differences reached high levels of statistical significance, with 80 (61.07%) reaching either $p \leq .01$ or $p \leq .001$ level.

Regarding the specific distributions of results, Table 4.18. shows how many sexual identity group differences were found for each type of reclamation behaviour.

Table 4.18. – Numbers of Significant Sexual Identity Group Differences for Reclamation Behaviours

Reclamation Behaviour	**Number of Significant Sexual Identity Differences**
Self-Self Reclamation	2 (1.51%)
Self-Other Reclamation	21 (15.91%)
Other-Self Reclamation	27 (20.45%)
Other-Other Reclamation	14 (10.61%)
Self-Reclamation	23 (17.42%)
Other-Reclamation	22 (16.67%)
Overall Reclamation	23 (17.42%)

The two sexual identity differences in self-self reclamation behaviour were for pejorative 'gay' and 'homo'. However, I suggest that these results should be discounted from analysis. While the data indicates a small number of people outside of the LGBTQ+ community reported using these words positively in self-reference, by my definition (see section 2.2.2.) I do not consider this an instance of reclamation. Additionally, in neither case were the number of responses for both sexual identity groups statistically comparable (34 non-heterosexual, 8 heterosexual for pejorative gay; 34 non-heterosexual, 5 heterosexual for homo). All other reclamation behaviours were relatively similar in this regard.

More generally, sexual identity differences for property/behaviour ratings behaved very similarly to the gender differences, in terms of how they reflected the general correlation pattern in the dataset. With only one exception, where such differences occurred, heterosexual respondents provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings. In this regard, sexual identity differences were almost completely consistent with the correlation pattern illustrated in Table 4.6. However, I have provided the results for 'gayboy' and 'homo' in Tables 4.19. and 4.20. to illustrate this.

I argue this represents a very clear ingroup/outgroup difference, one which I note was not present among the gender differences, which indicated no pattern of results that depended on whether the slur specifically targeted men or women.

The sole exception to these findings was 'tranny'. The sexual identity group differences for this slur are presented in Table 4.21. and indicate that the heterosexual group reported lower familiarity ratings, but also lower arousal, negativity and offensiveness ratings compared to the non-heterosexual group.

Table 4.19. – Sexual Identity Differences for 'Gayboy'

Word Property	Mean Rank (Non-Heterosexual)	*N*	Mean Rank (Heterosexual)	*N*	*U*	*z*	*p*
Familiarity	44.87	42	27.83	32	981.50	3.51	< .001
Personal Use	38.36	40	30.36	29	714.50	2.08	.04
Arousal	28.95	37	39.31	29	368.00	-2.29	.02
Tabooness	29.90	41	40.46	26	365.00	-2.26	.02
Age of Acquisition	30.49	41	39.54	26	389.00	-2.02	.04
Self-Other Reclamation	37.50	42	29.65	26	672.00	2.05	.04
Self-Reclamation	39.20	42	28.46	27	743.50	2.63	= .01

*Table 4.20. – Sexual Identity Differences for 'Homo'**

Word Property	Mean Rank (Non-Heterosexual)	*N*	Mean Rank (Heterosexual)	*N*	*U*	*z*	*p*
Familiarity	42.58	37	32.42	37	872.50	2.13	.03
Personal Use	44.83	36	26.91	35	948.00	3.86	< .001
Offensiveness	29.58	36	44.22	37	399.00	-3.02	< .01
Age of Acquisition	29.76	36	44.04	37	405.50	-3.14	< .01
Self-Other Reclamation	48.21	35	25.42	37	1057.50	5.33	< .001
Other-Self Reclamation	48.33	35	25.31	37	1061.50	4.85	< .001
Other-Other Reclamation	47.57	34	25.36	37	1022.50	4.86	< .001
Self-Reclamation	50.39	36	23.97	37	1148.00	5.92	< .001
Other-Reclamation	49.92	36	24.43	37	1131.00	5.28	< .001
Overall Reclamation	51.49	37	23.51	37	1202.00	5.72	< .001

* As per the discussion above, the Self-Self Reclamation difference for 'homo' has not been included.

Table 4.21. – Sexual Identity Differences for 'Tranny'

Word Property	Mean Rank (Non-Heterosexual)	*N*	Mean Rank (Heterosexual)	*N*	*U*	*z*	*p*
Familiarity	44.15	42	28.77	32	951.50	3.17	< .01
Arousal	41.76	42	31.91	32	851.00	2.01	.04
Negativity	42.02	41	30.56	32	862.00	2.43	.02
Offensiveness	41.07	42	31.48	31	822.00	2.07	.04

4.5. Discussion

The main objective of this study was to produce a normed database of word property ratings for LGBTQ+ slurs, which included measures of reclamation behaviours for each, for use in the experimental studies to follow. I achieved this aim, producing the database provided in Appendix 4. However, this data itself revealed a number of important findings, which both develop knowledge of the slurs identified in Chapter 3, and inform the designs and analysis of the experimental studies described in Chapters 5 – 7.

4.5.1. Patterns of Correlation between Word Properties and Reclamation Behaviours

As discussed in section 4.2., Madan et al. (2017) found that certain groupings of non-emotional (e.g. familiarity and personal use) and emotional (e.g. arousal, offensiveness, tabooness) word properties were best at explaining performance in taboo language processing tasks, as long as tabooness was

included in both models. Referring to Table 4.6., My results replicated these findings: familiarity and personal use correlated positively, while arousal, negativity, offensiveness and tabooness all correlated positively with each other. If two properties/behaviours did not share a positive correlation, then they always shared a negative one, suggesting two clearly delineated groups.

My correlation patterns developed existing findings by identifying where both AoA and measures of reclamation behaviour fit into these groups. First, I identified that AoA correlated positively with arousal, negativity, offensiveness and tabooness, but negatively with familiarity, personal use and all reclamation behaviours. Second, I identified that all reclamation behaviours correlated positively with familiarity and personal use, but negatively with arousal, negativity, offensiveness, tabooness and AoA. Overall, this suggested two key groups of properties/behaviours, sharing positive correlations within the group but inverse correlations between groups. The first group consisted of familiarity, personal use and all reclamation behaviours. The second group consisted of arousal, negativity, offensiveness, tabooness and AoA.

The results of my r_s^2 analysis (Table 4.7.) also developed existing findings by identifying which properties/behaviours shared the most variance in their ratings, and therefore may be more closely related. Generally, they suggested that word properties with positive correlations tended to share the most variance, but particularly that arousal, negativity, offensiveness and tabooness all shared >80% of their variance with each other. The same was true of the variance shared between all of the reclamation behaviours. Of all the properties in the data, AoA seems to have shared the least amount of variance with the others, regardless of the type of correlation. However, familiarity also shared similarly little variance with arousal, negativity, offensiveness and tabooness.

Some of these relationships seem intuitive. For example, it seems natural that greater personal use of an LGBTQ+ slur would be associated with greater familiarity. Similarly, the relationship of reclamation behaviour to familiarity and personal use makes sense; a reclaimed use of a slur is, after all, an instance of personal use. However, I note that it was not only the reclamation behaviours reflecting respondents' own uses (self-self reclamation, self-other reclamation, self-reclamation) that demonstrated this association; all reclamation behaviours did. For these specific correlations, it did not matter whether respondents were reclaiming the slur themselves or were experiencing others doing the same.

Similarly expected was that slurs rated more offensive and more taboo also tended to be reported more arousing and more negative. I consider this to replicate findings from existing literature; as discussed in sections 2.3.1. and 2.3.2., arousal (e.g. Anderson & Phelps, 2002; Bayer, Sommer, & Schacht, 2012; Brown & Kulik, 1977; Kensinger & Corkin, 2003; Madan et al. 2017) and valence (Bayer et al. 2012; Kensinger & Corkin, 2003) are closely linked to the processing of emotional language, with highly taboo language tending to demonstrate the most exaggerated versions of these effects (Jay, Caldwell-Harris, & King, 2008; Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Madan et al. 2017).

An unanticipated finding was that those who reported first encountering a slur later in their lives generally also reported it being more arousing, negative, offensive and taboo. Considering this alongside the finding that later AoA ratings tended to be given alongside lower familiarity and personal use ratings, it is possible that these results suggest a desensitisation effect of acquiring a word earlier in life.

Because of the demonstrated importance of arousal, offensiveness and tabooness to taboo language processing, the most important finding concerning the impact of reclamation was that all of these properties correlated negatively with every measure of reclamation behaviour in the data. Therefore, respondents' ratings for arousal, negativity, offensiveness and tabooness tended to be lower when their ratings for the various reclamation behaviours were higher. I consider this evidence for the possibility that, by its nature, linguistic reclamation might lessen perceptions of

arousal, negativity and tabooness for slurs, and may affect the way they are processed compared to other kinds of taboo language. This is a possibility which I hypothesised in sections 2.2.3. and 2.3.3. However, I also acknowledge that the reverse could be true; an increase in reclamation might be a consequence of (rather than a cause for) a slur starting to be viewed as less arousing, negative, offensive, or taboo in general.

4.5.2. Age Group Differences: Age of Acquisition and Familiarity Ratings

The comparison between respondents aged between 16-29 and 30+ revealed that almost a third of age group differences were differences in AoA, and just over a fifth were differences in Familiarity. In the case of the AoA differences, the results indicated that in all but one case of age difference, there were earlier ages of acquisition reported in the 16-29 group compared to the 30+ group.

Of all my findings in this study, these are the ones to treat with most caution. Arguably, AoA might be expected when age itself is the variable being compared. Similarly, AoA ratings might be expected to be lower and less variable in the younger group; those in the 16-29 group were not old enough to be able to select outside of the first three possible answers to the AoA question.

However, I argue that differences in age do not presuppose differences in AoA; if the use of a slur had remained constant intergenerationally, the age at which the slur is first encountered should be the same, regardless of the age of the respondent. This was not the case for many LGBTQ+ slurs, suggesting that the use of these words is not constant, and that generally speaking these words are being acquired at an earlier age by younger speakers. I argue that this is because LGBTQ+ identities have become much more visible in recent decades, and suggest two possible effects of this on age of acquisition. First, if LGBTQ+ identities are made more generally visible – particularly if this is facilitated by more frequent representations in the media – it is likely that awareness of LGBTQ+ people becomes less restricted to adults, and younger people have more opportunities to develop this awareness. Consequently, they may learn vocabulary associated with LGBTQ+ people at a younger age, including that which is derogatory. I also suggest that increased visibility does not automatically equate to increased social acceptance, but does mean that LGBTQ+ people are more socially recognisable. When it is easier to identify LGBTQ+ people in general, it follows that there are more opportunities for slurs to be used to target them, potentially also causing them to be acquired earlier.

As indicated in Tables 4.6. and 4.7., AoA had one of its strongest correlations with familiarity, which was a negative one. Table 4.8. also indicates that the 4 LGBTQ+ slurs with the most significant age group differences for AoA also had some of the highest familiarity ratings in the data. Together, these results suggest that slurs which exhibited AoA differences might also have increased in use over time.

The sole exception to the pattern of AoA differences was ‘queer’, for which AoA was the only property that differed significantly by age. In this case, the 30+ group instead reported an earlier age of acquisition than the 16-29 group. This was an unexpected finding, but I suggest a possible explanation comes from its particular history of reclamation, as I discussed in section 2.2.4. of my literature review. I consider that the various reclaimed senses of ‘queer’ may have led to it becoming more specialised to LGBTQ+ ingroup use over time, and by extension become less widespread in the general population. This includes the use of the word in formal or academic contexts – e.g. ‘queer theory’ – which is certainly unlikely to be encountered early in life, although is so specialised that it is unlikely to be the first reclaimed use of ‘queer’ a person ever encounters. If age group is the only demographic variable being considered, as was the case in this comparison, such changes in use might generally cause younger speakers to acquire it later. Another possibility is that ‘queer’ used to have a number of meanings largely unrelated to its use as an LGBTQ+ slur, such as ‘strange’ or ‘unusual’, which have since declined (Brontsema, 2004; Gaucher, Hunt, & Sinclair,

2015). If the word used to have entirely non-taboo meanings – which are less restricted in use than taboo ones and therefore likelier to be acquired sooner – that have become obsolete, this might also explain why the word has started to be encountered for the first time at an older age.

Regarding familiarity, the word property with the second highest number of age differences, these results were useful for determining which slurs might be increasing or declining in use over time. In sections 3.4.3. and 3.5.2. of the previous chapter, I identified and discussed LGBTQ+ slurs belonging to semantic fields of sexual acts and non-normative gender expressions. The results in Table 4.10. suggest that many of the slurs that were more familiar to older respondents and less familiar to younger ones belonged to one of these groups, e.g. 'fudge-packer', 'rug-muncher', 'bull-dyke', 'pansy'. As shown in Table 4.11., the slurs higher in familiarity for younger respondents are more concerned with possessing an LGBTQ+ identity than they are with displaying behaviours associated with being LGBTQ+. Of these examples, it is particularly notable that the more recent use of 'gay' as a pejorative was reflected in the age-related familiarity comparisons; as discussed in section 3.5.1. of the previous chapter, the word has started to be used to refer to "anything that is pathetic or useless" by direct association with homophobic attitudes to gay people (Baker, 2008, p. 129).

At the end of section 4.4.2.1., I noted four slurs which my results suggested were being acquired at an earlier age by 16-29 year olds, but were also less familiar to the same group: 'arse-bandit', 'bull-dyke', 'fudge-packer' and 'rug-muncher'. This runs against the overall correlation pattern, in which less familiar slurs are acquired at a later age. One possible explanation for might be that these particular slurs have declined in general use, but have been retained as 'playground insults', explaining a generally younger AoA. However, a better explanation is provided by the results discussed in this section. My results suggest that all LGBTQ+ slurs are generally being acquired at an earlier age by younger people. These four slurs belong to semantic fields of sexual acts and non-normative gender expressions, which my results indicate are less familiar to younger people. Taking these results together, it seems entirely possible that some slurs might be less familiar to younger people but still acquired earlier by them, even if this is not the general trend.

Collectively, I consider all of these results to affirm the emphasis placed on "specifiable times" in Allan and Burridge's (2006, p. 11) definition of taboo language, as discussed in sections 2.1.1. and 2.1.3. of my literature review (Chapter 2). I do not suggest that social and linguistic changes over time cannot manifest cross-generationally, but do argue that the presence of key age differences in my data indicates that usage and perception of LGBTQ+ slurs has not remained static over time, even just in recent decades.

4.5.3. Differences in Ratings by Gender and Sexual Identity

As discussed in section 4.4.2.2., where differences occurred between women and men, almost without exception women provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings. Similarly, where differences occurred between non-heterosexual and heterosexual respondents, heterosexual respondents almost always provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings (see section 4.4.2.3.).

Both of these patterns in gender and sexual identity differences mirror the overall correlations found in the data, as set out in Table 4.6. That is to say, where such demographic differences existed, if one demographic group had higher ratings in one group of properties/behaviours (e.g. familiarity, personal use, reclamation behaviours), the same demographic group would also tend to have lower ratings in the other (e.g. arousal, negativity, offensiveness, tabooness, AoA). I had anticipated that the broader correlation pattern might change

when looking within specific gender and sexual identity groups, as LGBTQ+ slurs are strongly related to both of these factors. I argue that the fact that it did not change speaks to the strength and significance of the overall pattern.

Although the gender and sexual identity differences both followed this pattern, the manner in which each did so raised distinct issues, to which I now turn my discussion.

4.5.4. Gender Differences and the Gendering of Slurs

Women consistently reported lower familiarity, personal use and reclamation behaviour ratings. Indeed, the majority of the most statistically significant gender differences were differences in reclamation behaviour, although all of the slurs with gendered differences in reclamation specifically targeted LGBTQ+ men (see Tables 4.15. – 4.17.). Women also tended to rated LGBTQ+ slurs as more arousing, more negative, more offensive and more taboo (see Tables 4.12. and 4.13.). Combined, these results reflect the broader relationships between ratings in Table 4.6. This was the case for most slurs which explicitly target LGBTQ+ women, but for some, there were no differences between women and men whatsoever. Given the broader pattern, this suggests two possibilities. Either gender simply did not affect responses to these particular slurs, or the cause of the general difference in women's responses may not be as present when women are the target of the slur.

As to what this cause may be, it is possible that these results speak to a broader social expectation regarding women and taboo language. A number of well-known studies in language and gender research have identified a particular burden on women to use forms of language which are felt to be more polite and indirect (e.g. Brown, 1980; Coates, 1988; Fishman, 1980; Holmes, 1995). There has been much debate over the course of recent decades as to why this is the case, but more contemporary arguments stress that where gendered speech norms do develop, they are often a result of pervasive, patriarchal and heteronormative socialisation from an early age (Cameron, 2006). This body of work offers an explanation as to why women may have generally reacted less favourably to LGBTQ+ slurs. However, it does not explain why women no longer exhibited this more negative response for some slurs which explicitly target them. Here, I note that more of the slurs in Table 4.2. typically target LGBTQ+ men rather than LGBTQ+ women. I also note that my sexual identity comparisons revealed those not targeted by LGBTQ+ slurs provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings. This was the same pattern that gender differences followed, suggesting that differences in women's ratings may also have been reflecting their judgements of slurs which do not typically target them. This would help explain why, for some of the slurs which did explicitly target women, these gender differences disappeared.

4.5.5. Sexual Identity and Reclamation Behaviour Ratings

As raised in sections 4.5.3. and 4.5.4., my sexual identity comparisons revealed that compared to non-heterosexual participants, heterosexual participants provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and AoA ratings. This pattern again mirrors the overall correlation pattern in Table 4.6., and suggests that those not targeted by LGBTQ+ slurs reported less experience with and greater objection to them. This is an extremely interesting finding, as it suggests that the group which reports the least experience of and strongest objections to slurs is the group which is not targeted by them, raising a question about who most reinforces the taboos surrounding marginalised social groups and the language which targets them.

I also found that 132 of the 187 (70.56%) of the sexual identity differences were differences in reclamation behaviours. It is perhaps unsurprising that it was always the non-heterosexual group

reporting higher reclamation behaviour for LGBTQ+ slurs (see Tables 4.19 and 4.20 for specific examples), given that the reclamation of slurs is a process which operates specifically within the targeted group (Bianchi, 2014; Herbert, 2015). This might also explain why the majority of differences between the two groups reached high levels of statistical significance, with 80 (61.07%) reaching either $p \leq .01$ or $p \leq .001$ level.

Unlike the gender differences, where there was no effect of the specific gendering of a slur, sexual identity did appear to have affected responses to slurs targeting non-heterosexual individuals. I suggest this finding indicates a much clearer issue of ingroup/outgroup membership and hegemonic power, which justifies the emphasis on these concepts in existing literature regarding their relationship to slurs and reclamation (see sections 2.1.2. and 2.2.).

It is interesting that other-self reclamation was the most common sexual identity difference in reclamation behaviour, rather than any of the reclamation behaviours focused on individuals using the slurs themselves (see Table 4.18.). With the relevance of ingroup/outgroup membership in mind, I suggest as a possible explanation that ingroup access may itself be required in order to experience other LGBTQ+ people using an LGBTQ+ slur to describe themselves positively.

4.5.6. Differences in Ratings for the Transphobic Slur 'Tranny'

A final set of findings concerned the transphobic slur 'tranny', for which ratings often misaligned with the correlation and demographic patterns described thus far. As illustrated in Table 4.8., 'tranny' had one of the highest average familiarity ratings in the data, and was typically reported as being acquired earlier by respondents in the 16-29 age group, which was consistent with other age group differences. However, as shown in Table 4.9., none of the other age differences for 'tranny' reflected the general correlation pattern in terms of how they related to AoA differences. The pattern in Table 4.6. suggests that an earlier AoA should have meant that ratings for personal use and self-reclamation would be higher in the 16-29 group, while ratings for arousal, negativity and offensiveness would be lower in the same. However, Table 4.9. shows the opposite was true. Despite first encountering the word at an earlier age, younger respondents appeared to object more to its use – the only slur for which this was the case.

A similar breaking of pattern was found in the sexual identity comparisons for 'tranny'. Referring to Table 4.21., while the non-heterosexual group generally reported higher levels of familiarity with the slur, they also gave it higher ratings for arousal, negativity and offensiveness. Again, this is not consistent with the results described in Table 4.6. and section 4.5.3. These results suggest that despite the non-heterosexual group being more familiar with the word, they too objected to it more strongly.

With reference to the demographics of my respondents, it is worth noting at this point that just over 90% of the respondents in this study were cisgender, as shown in Table 4.1. These results therefore cannot speak to the perceptions of 'tranny' among those who the slur targets most; a similar study to this with a statistically valid sample of both cisgender and transgender respondents could and should explore whether membership of the trans community affects these findings.

However, I argue that these results suggest two perceptions of transphobic slurs among cisgender people. First, that there may be a greater awareness of and subsequent caution toward explicit transphobia among younger cisgender people. This possibility is supported by recent literature suggesting more unified recognition of and opposition to transphobic discrimination among younger people (e.g. Crissman, Czuhajewski, Moniz, Plegue, & Chang, 2019) compared to older people (Callahan & Zukowski, 2019; Elischberger, Glazier, Hill, & Verduzco-Baker, 2016; Parent & Silva, 2018). Second, these results suggest that among cisgender people, membership of the LGBTQ+ community seems to simultaneously increase awareness of and opposition to transphobic slurs; of my largely cisgender respondents, those who were also heterosexual and therefore not

members of the LGBTQ+ community were less familiar with 'tranny' but also made less strongly negative judgements of it.

4.6. Conclusion

The principal goal of this study was to obtain a set of normed ratings for LGBTQ+ slurs, both for a number of word properties relevant to the processing of taboo language, and for various measures of linguistic reclamation. In this, the study succeeded: it generated a combined total of 34,949 individual ratings, which were collectively used to produce normed ratings for all 14 properties/behaviours, for all 41 LGBTQ+ slurs in the study. The resulting database is a novel contribution to the field in two regards: it is the only such database of word norms specifically for LGBTQ+ slurs, and it is the only one to include slurs which also includes normed measures of slur reclamation. This information is essential for measuring the effects of these word properties and of linguistic reclamation on the processing of LGBTQ+ slurs in experimental research, including the studies I describe in Chapters 5-7 of this thesis.

However, this was not the only successful outcome of the study: further analysis of the data produced by the study yielded important findings of its own. I consider the most significant of these to concern the relationships found between word properties and reclamation behaviours. Every property/behaviour in the study was found to have a significant correlation with every other property/behaviour, and there was a clear pattern to these correlations. Familiarity, personal use and all types of reclamation behaviour all correlated positively with each other. Additionally, arousal, negativity, offensiveness, tabooness and age of acquisition all correlated positively with each other. Importantly, properties/behaviours in one group were always found to have correlated negatively with any property/behaviour in the other group. Although these findings cannot be used to argue causation, I argue they provide early support for my hypothesis that increased reclamation of slurs might be associated with reduced arousal, negativity and tabooness judgements, and may therefore have an effect on performance in taboo language processing experiments.

My study also identified a number of important demographic differences in responses. Results generally indicated that the age of acquisition of LGBTQ+ slurs was generally earlier in younger (16-29) respondents, which I argue may be attributable to increasing visibility of LGBTQ+ people, and by extension the words used to describe them. Several age group differences were also found for Familiarity ratings, providing insight on which LGBTQ+ slurs may be increasing or decreasing in use cross-generationally.

Differences in word property ratings between women and men, as well as between non-heterosexuals and heterosexuals, were found to have aligned with the general word property correlation pattern. Typically, both women and heterosexual respondents provided lower familiarity, personal use and reclamation behaviour ratings, but higher arousal, negativity, offensiveness, tabooness and age of acquisition ratings. Significantly, these gender differences were not found to have been affected by the gendering of LGBTQ+ slurs, and the most pronounced differences in terms of sexual identity were for reclamation behaviours. Although there are several possible explanations for these findings, I particularly consider that these results reflect ingroup/outgroup differences, as well as different gendered expectations regarding use of taboo language more broadly.

Finally, my results suggested that patterns of demographic differences in responses to transphobic slurs may not be the same as those to slurs targeting other groups within the LGBTQ+ community, at least for cisgender individuals. I suggest that for cisgender LGBTQ+ people, both age group and membership of the LGBTQ+ community affects recognition of and opposition to transphobic language, but note that the database is limited in its ability to draw such conclusions. This issue could be explored more thoroughly by a study which explicitly compares transphobic slurs

to homophobic ones, and which has a better balance of cisgender to transgender and non-binary respondents.

Chapter 5 – Comparing LGBTQ+ Slurs to High-Arousal and Low-Arousal Negative Words in a Lexical Decision and Free Recall Task

5.1. Introduction

This chapter focuses on the third study of my research: a lab-based experiment to compare the processing of LGBTQ+ slurs as a type of taboo language to high- and low-arousal non-taboo words. Given the emphasis of my research on the reclamation of LGBTQ+ slurs, I was particularly interested in participants' responses to these words and the degree to which reclamation affected them.
I begin by revisiting some of the findings from the literature that are relevant to this experiment. I then discuss in detail my participant sample, experiment design, experiment procedure and methods of data analysis, before outlining my results. I then consider some of the most significant findings to emerge, with particular regard to the differences found between types of word and demographic groups, as well as the variables – including measures of reclamation behaviour – that were best at predicting performance in the experiment tasks.

5.2. Research Background

Given the focus of this experiment, the most relevant findings from existing literature are those concerning the impacts of both arousal and tabooness on language processing. I discuss these in detail throughout section 2.3. of my literature review, but revisit some of the most important existing findings in this section. I also outline again the potential impact of linguistic reclamation – which I define as non-derogatory uses by members of targeted groups of the slurs which members of dominant groups use to target them – as discussed in sections 2.2.3. and 2.3.3.

Generally, two key processing effects have been observed for taboo language. The first is increased recallability, even relative to non-taboo, high-arousal stimuli (Buchanan, Etzel, Adolphs, & Tranel, 2006; Jay, Caldwell-Harris, & King, 2008; Kensinger & Corkin, 2003; Madan, Schafer, Chan, & Singhal, 2017). These findings are related in part to those which suggest all kinds of emotionally arousing stimuli – both verbal and non-verbal – are associated with increased neurological (Anderson & Phelps, 2002; Bzdok, Laird, Zilles, Fox, & Eickhoff, 2013; LaBar & Phelps, 1998; Packard & Cahill, 2001) and physiological (Cahill, Prins, Weber, & McGaugh, 1994; Sharot & Phelps, 2004) responses which contribute to memory activation and retention, such as activation of the amygdaloid complex and adrenergic system.

However, this does not account for the significant increase in memorability for taboo words specifically. One explanation for this is that taboo language is simply the most arousing type of emotionally arousing language, and therefore demonstrates the most exaggerated effect on word memory (Janschewitz, 2008; Jay et al., 2008). However, more recent studies have suggested that differences in the intensity of arousal responses cannot by themselves explain enhanced memorability for taboo words, and have instead suggested that tabooness itself might enhance word memorability (Madan, Caplan, Lau, & Fujiwara, 2012; Madan et al., 2017).

The second taboo language processing effect is impaired lexical accessibility relative to non-taboo words, whether the non-taboo words are emotionally arousing or not (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005; Williams & Evans, 1980). These findings contrast those of earlier lexical accessibility research, which – although not focusing on taboo language – tended to find that among non-taboo language, emotional arousal increases lexical accessibility (Graves, Landis, & Goodglass, 1981; Strauss, 1983; Williamson, Harpur, & Hare, 1991). One suggested explanation for these results has been the idea of a 'perceptual defence' mechanism – an urge to

avoid the processing of troubling stimuli – causing processing delays for taboo language not seen in non-taboo language (Williams & Evans, 1980; Madan et al., 2017). This is in turn related to the concept of *approach-avoidance motivation* in behavioural psychology (for review see Higgins, 1997; Elliot, 2006; Elliot, Gable, & Mapes, 2006; van Berkum, 2019), in which *approach motivation* describes "the energization or direction by, or the direction of behaviour toward, positive stimuli" and *avoidance motivation* describes "the energization or direction by, or the direction of behaviour away from, negative stimuli" (Elliot, 2006, p. 112). Taboo language may represent a negative stimulus to be avoided, so research has suggested that tabooness might prompt an avoidance of processing, subsequently impairing lexical access (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005).

A key experimental study I have cited often in both my literature review and already in this section is Madan et al. (2017), whose experiment designs, procedures and analyses strongly informed how I conducted the experiments reported in both this chapter and Chapter 6. Importantly, their study looked at both lexical access and free recall, making it a particularly important study to summarise in this section. They conducted a lexical decision task, followed by a free recall task, followed by an affective ratings task, in order to test the lexical accessibility and recall likelihood of the following types of word: highly arousing taboo words, moderately arousing positive-valence words, moderately arousing negative-valence words, and emotionally neutral words. All of these words were selected from Janschewitz's (2008) normed database, which I discussed critically and in detail in sections 3.2 & 4.2., as a key contribution to the literature. Using Analyses of Variance (ANOVAs), they found that lexical decision response times were slower to taboo words than to almost all other word categories (excluding the negative valence words) in the lexical decision task, and recalled more than any other word category in the free recall task. Using linear regression models in combination with the affective ratings they had gathered, they found that a combination of 'non-emotional' properties (imageability, familiarity, personal use, letters, syllables, word frequency, contextual diversity, and age of acquisition) and tabooness best explained performance in the lexical decision task. Conversely, a combination of 'emotional' properties (specifically tabooness, arousal, and offensiveness) best explained free recall probability, with tabooness making the strongest contribution to this model. These findings were then retested using a significantly larger dataset, and were successfully replicated.

However, as I noted in section 4.2. of Chapter 4, Madan et al.'s study included no LGBTQ+ slurs as part of its taboo stimuli, nor did it consider any potential impact of linguistic reclamation on the slurs that were present in the experiment. It is not at all unusual in these regards; few psycholinguistic studies of taboo language include many examples of LGBTQ+ slurs, and there appears to be a total absence of literature which explicitly considers whether linguistic reclamation may affect responses to arousing and/or taboo language. However, the results of my norming study in Chapter 4 suggest that all kinds of reclamation behaviour had a significant inverse correlation with arousal and tabooness judgements for LGBTQ+ slurs: when any reclamation behaviour rating increased, arousal ratings decreased (see Table 4.6.).

Relating to this, there is some theoretical basis – though not extensive – for believing that reclamation might affect tabooness judgements as well. Research into both taboo language judgements (e.g. Allan, 2019; Allan & Burridge, 2006; Culpeper, 2019; Jay, 2019b) and linguistic reclamation (e.g. Bianchi, 2014; Herbert, 2015; Sinfield, 1994) has positioned ingroup/outgroup membership and intergroup status (Giles & Giles, 2013; Tajfel & Turner, 1986) as an issue central to both. It is possible that the ability to reclaim slurs within particular social groups might also affect judgements of the tabooness of those same slurs in those same groups. Additionally, Culpeper (2011; 2019) suggests that an individual's awareness of tabooness 'metadiscourse' - broader

knowledge of and discussion about the associations between taboo expressions and the communicative contexts in which they occur – is key to their judgements of that language as being taboo. It is possible that an individual's metadiscursive knowledge of a slur might be shifted by awareness and experience of contexts in which that slur is both positive and permitted, i.e. when it is being reclaimed. Again, the findings of my norming study in Chapter 4 support this possibility: all kinds of reclamation behaviour had a significant inverse correlation with tabooness judgements for LGBTQ+ slurs.

Based on the extant literature and my own findings, I made the following three predictions. First, that the LGBTQ+ slurs in my study would generally be associated with reduced lexical accessibility but improved recallability, relative to both types of non-taboo words. Second, that any such effects for the LGBTQ+ slurs might be reduced by greater participation in and awareness of the linguistic reclamation of particular slurs, given that reclamation may lessen arousal responses and tabooness judgements for those slurs. Finally, that any effect of reclamation on the processing of the LGBTQ+ slurs should be strongest among the non-heterosexual participants, as the likeliest group to be engaging in and aware of their reclamation.

I tested my predictions using a lab experiment, in which lexical accessibility was measured via a lexical decision task, and recallability via a free recall task. I used all of the normed word property and reclamation behaviour ratings collected and outlined in Chapter 4 to explore how these affected performance in these tasks, but also used a new survey to collect reclamation ratings from the lab participants themselves. This survey also collected demographic information about my participants, to determine if and how these broader factors might have affected performance in the tasks. I now outline the methodological considerations for each stage of designing and running the experiment.

5.3. Methodology

5.3.1. Sample

Fifty-nine participants took part in the experiment. Data from 5 participants were excluded: 4 for not having met the inclusion criteria (outlined below) and 1 for providing an excess of incorrect responses to the lexical decision task (< 80% response accuracy). As such, a final total of 54 participants were entered into the analyses.

Of these, 29 were recruited through a combination of posts on social media (Facebook and Twitter) and research posters placed around the University of Nottingham campus and Nottingham-based community centres, with permission. Participants recruited via these methods were able to take part in a raffle for a £10 shopping voucher as a participation incentive, though this was optional. The remaining 25 participants were undergraduate students at the University of Nottingham, who were recruited in exchange for course credit. These participants were not eligible to take part in the voucher raffle.

All participants were aged between 16 and 29 years old and were all native or near-native speakers of British English. This age range was selected to control for effects of participant age, while also acknowledging the differences in LGBTQ+ slur ratings between 16-29 year olds and 30+ year olds that were identified from my norming survey in Chapter 4. I collected additional respondent demographic data in order to enable demographic comparisons (see section 5.3.2.3. for details). Demographic data is presented in Table 5.1. There was a mean participant age of 20.15 (SD: 2.45), the majority of participants identified as women, almost all participants were cisgender, and a slight

majority of participants were heterosexual rather than non-heterosexual. These categories were later refined for the purposes of statistical analysis, as discussed in section 5.3.4. and illustrated in Table 5.3.

Table 5.1. – Overall Participant Demographic Data

Age	***N***	**%**	**Gender**	***N***	**%**	**Cisgender?**	***N***	**%**	**Sexual Identity**	***N***	**%**
16	1	1.85	Woman	38	70.37	Yes	51	94.44	Straight/Heterosexual	29	53.70
18	10	18.52	Man	13	24.07	No	3	5.56	Gay/Lesbian/Homosexual	10	18.52
19	20	37.04	Non-Binary	2	3.70	Total	54	100.00	Bisexual/Pansexual/Polysexual	11	20.37
20	8	14.81	Agender	1	1.85				Queer	2	3.70
21	2	3.70	Total	54	100.00				Asexual	2	3.70
22	6	11.11							Total	54	100.00
23	1	1.85									
24	1	1.85									
25	2	3.70									
26	1	1.85									
27	1	1.85									
28	1	1.85									
Total	54	100.00									

5.3.2. Materials

The experiment involved three main stages: a lexical decision task, a free recall task, and a survey task to collect participant demographic information and reclamation behaviour ratings. In this section, I outline the materials involved in each stage of the experiment.

5.3.2.1. Lexical Decision Task

For the lexical decision task, stimuli fell into three experimental conditions, based on word type: 25 LGBTQ+ slurs, 12 negative high-arousal words and 12 negative low-arousal words. Where I later refer to experimental 'conditions' as an independent variable, I am referring to these 3 word type categories. All stimuli had negative valence, to ensure that arousal was the main manipulation in the experiment. For the same reason, stimuli were balanced across conditions in terms of character length and number of syllables. All stimuli ranged from 3 to 7 characters in length, and contained a maximum of 2 syllables. The task also included 49 non-words as filler items.

LGBTQ+ slurs were selected from those identified by my LGBTQ+ slur identification survey (see Chapter 3), which had met the frequency threshold for continued use in my research (see Table 4.2. for complete list). Both the negative high-arousal and negative low-arousal words were selected from the normed database produced by Janschewitz (2008), as these had already been normed for both arousal and valence. Non-words were generated using Wuggy Pseudoword Generator, version 0.1.9. (Keuleers & Brysbaert, 2010). Using this software ensured that all non-words were pronounceable and both phonotactically and morphologically plausible for English, so that it would not be too easy for participants to identify them as non-words in the experiment. The software was also able to use the real words in the experiment in order to generate the non-words, which helped ensure the non-words were not too obviously distinct from the real words, and that non-words proportionally matched real words for character and syllable length. Table 5.2. provides examples of the stimuli used in the experiment, divided by condition. The full list of experimental stimuli can be found in Appendix 5.

Table 5.2. – Examples of Lexical Decision Task Stimuli

LGBTQ+ Slurs	Negative High Arousal Words	Negative Low Arousal Words	Non-Words
Bender	Assault	Alone	Berfer
Dyke	Chaos	Carcass	Calsy
Faggot	Guilty	Crime	Deke
Lesbo	Gun	Illness	Filmot
Queer	Torture	Lonely	Gabtord
Tranny	Toxic	Sick	Perdort

5.3.2.2. Free Recall Task

As the goal of the free recall task was to recall as many real words from the lexical decision task as possible within a set time frame (see section 5.3.3. for detailed discussion of the procedure), the materials for participants to recall were identical to those used in the lexical decision task (see section 5.3.2.1.), excluding the non-words.

5.3.2.3. Survey Task

The survey task was split into two parts: a set of four demographic questions, and a set of four reclamation behaviour rating tasks. Although my norming survey had generated normed reclamation behaviour ratings for all LGBTQ+ slurs used in this experiment (see sections 4.3., 4.4. and Appendix 4), I was still keen to gather each individual participants' own reclamation ratings for two reasons. The first was to include individual experience of reclamation behaviours in my analysis, as opposed to solely a broader, general-population measure of reclamation behaviours. The second was to avoid making the assumption that each participant's experience of reclamation would necessarily be comparable to others', and instead acknowledge participants' individual linguistic behaviours and practices (Mendoza-Denton, 2004; Potter, 2000; Schegloff, 1997).

The demographic questions asked participants to state their age, their gender, whether they were cisgender (their gender matched the one they were assigned at birth), and their sexual identity. With the exception of the third question – which was either a yes or no response – all other questions were open-response. Although some categorisation of participants by identity would be necessary to enable demographic group comparisons, here I decided to follow the recommendation of other scholars doing quantitative identity research that the analyst "look to participants' own explicitly signalled orientations and formulations" (Mendoza-Denton, 2004, p. 478) in order to determine what identities are actually present in my research. Though it resulted in more work to create a balanced sample for my analyses, this decision had the benefit of avoiding the imposition of my own identity judgments and categories onto my participants, risking both bias and exclusion (Mendoza-Denton, 2004; Potter, 2000; Schegloff, 1997).

The reclamation behaviour ratings questions were the same as those used in the norming survey, as were the instructions to participants for providing their responses (see section 4.3.2. for details of both). The sole difference was that participants were only asked to provide their own reclamation ratings for the 25 LGBTQ+ slurs used in the experiment. This is the reason that the survey task was the last stage of the experiment; I did not want this list of LGBTQ+ slurs to prime participants' responses to either the lexical decision or free recall tasks.

5.3.3. Procedure

Upon arrival, participants gave informed consent before beginning the study. They then sat in front of the computer on which the lexical decision task was running, and were given time to adjust the desk, chair and keyboard positions for comfort.

The lexical decision task was built and run using OpenSesame, version 3.3.3. (Mathôt, Schreij, & Theeuwes, 2012). The experiment consisted of a short practice block containing unrelated words, in order to teach participants how to complete the task correctly, followed by the main experiment block. The practice block contained 6 real words not used in the main experiment and 6 non-words. The experiment block contained all 98 experimental stimuli. Practice and main blocks were each preceded by an instruction screen reminding participants how the experiment worked. Additional information relevant to the completion of the subsequent parts of the experiment was included on the screen following completion of the experimental block.

All text was presented in black Monospace font, size 32 px, on a light grey background. The order in which stimuli were presented was randomised for each participant, to control for any effect of presentation order. Presentation order was still recorded for each participant, in case response time or recall likelihood was different for stimuli presented earlier or later in the lexical decision task, but no such effects were ultimately found in any of my analyses. Before a stimulus was presented, participants were shown a fixation point (a black circle) in the exact centre of the screen for 500ms, to align their gaze to where the stimulus would then be presented. The stimulus remained on-screen until participants gave a keyboard response to indicate whether they thought it was a real word or not. I decided not to cap the time in which a participant could give a keyboard response; given that pronounced delays to lexical decision are characteristic of taboo language (Geer & Bellard, 1996; Bertels, Kolinsky & Morais, 2009; Madan et al., 2017), I did not want to risk limiting my ability to measure the full extent of these delays.

Responses were given by pressing either the ‘Z’ key to indicate a real word, or the ‘M’ key to indicate a non-word. These keys are at opposite ends of the bottom row of a standard QWERTY keyboard, so were chosen to ensure participants’ hands would not be positioned in such a way that might risk them pressing the wrong key. In the experiment block, the software recorded both the speed with which the key press was given (to use as a measure of processing speed), and whether the response was correct or not (to allow me to filter out incorrect responses).

When participants entered the lab and began the lexical decision task, they were first presented with an instruction screen detailing how the task worked. This informed them that during the task, they would see a string of letters on-screen, and told them which buttons on their keyboard they would have to press to indicate whether they thought the string was a real word or not. The instruction screen advised participants to get their hands into position before the experiment began, so that response times to the first word in the block would not be delayed. The instruction screen also told participants how stimuli would be presented to them: the fixation point in the centre of the screen that they were required to focus on, then the letter string, then the fixation point again once they had given their keyboard response. Participants then completed the practice block, followed by the experimental block after being informed the practice block was complete.

Following the procedure used by Madan et al. (2017), the free recall task was presented to participants immediately following completion of the lexical decision task, and participants were given 5 minutes upon starting the task to correctly recall as many real words as they could from the

lexical decision task. If participants stopped being able to recall any more words before the full 5 minutes had passed, they could ask to end the task.

Responses were provided using an online survey presented on-screen, in which participants listed all words they could remember by typing them into a free-response text box. Participants were asked to separate each word with a line break, for ease of analysis. Before beginning, participants were given brief instructions on how to provide their responses, including being shown an example response using words unrelated to the task. The 5 minute period in which participants could provide responses only began once they had left the instruction screen and begun the free recall task. Once this period was over – or sooner if participants had stopped being able to recall more words – participants were told to stop the task and click on a 'finish' button to submit their responses.

Finally, participants were given the survey task, and were encouraged to read through the instructions and respond to the questions at their own pace. If they had any questions while completing any part of this task, they were able to pause and ask. Following completion, participants were thanked for their involvement in the research.

At this stage, those who were not participating for course credit were given the option to provide an email address to enter the raffle to win a shopping voucher worth £10. All participants were then given a final opportunity to ask questions about the research. Overall, the procedure took between 15-25 minutes.

5.3.4. Data Analysis

Identifying the results of the lexical decision and free recall tasks involved two main stages of data analysis. The first of these was identifying whether there were significant differences between the independent variables: experimental condition (LGBTQ+ slurs vs. high-arousal negative words vs. low-arousal negative words), gender (women vs. men), and sexual identity (non-heterosexual vs. heterosexual), as well as any interaction effects between these. The second stage aimed to identify which combinations of factors (gender, sexual identity, stimulus character length, stimulus syllable number, normed word and reclamation ratings, participant reclamation ratings) best predicted lexical decision response times and likelihood of recall. All data analysis was conducted using IBM SPSS Statistics v.26, except for the analyses described in section 5.3.4.1., which were conducted using JASP v.0.16.2.

Prior to analysis, I visually inspected the data for outliers, such as abnormally fast or slow response times for single participants, or high amounts of incorrect responses. Two participants consistently provided response times > 2000ms, which were considerably slower than all other participants, so their responses were removed from all analyses. For the free recall results, I also ran a validity check to ensure that participants were recalling words that had actually been included in the lexical decision task. This was especially important given that the LGBTQ+ slurs belonged to a common semantic field; I needed to be confident that participants were not simply reporting words which they thought could have been included in the task. I found that 91.01% of all words recalled were those which had been included in the lexical decision task, so was satisfied that participants were accurately and appropriately completing the recall task.

Before conducting any analyses involving participant demographic data (see Table 5.1.), I first needed to recategorize participants to ensure that groups were more balanced. Participants were all of a similar age for control purposes, and I could not make balanced comparisons based on whether participants were cisgender, because over 90% of participants were. Therefore, I only made

comparisons based on gender and sexual identity. For gender, I decided only to compare cisgender women and men, and for sexual identity I decided to make comparisons between heterosexual and non-heterosexual participants. I also decided not to include asexual participants within the non-heterosexual category, in order to avoid conflating asexuals with non-heterosexual allosexuals (those who experience sexual attraction, see Antonsen, Zdaniuk, Yule, & Brotto, 2020). Instead, I excluded data from asexual participants from just the sexual identity comparisons. Table 5.3. shows the new distributions of participants in the new categories.

Table 5.3. – Revised Gender and Sexual Identity Group Distributions

Gender	***N***	**%**	**Sexual Identity**	***N***	**%**
Woman	38	74.51	Non-Heterosexual	23	44.23
Man	13	25.49	Heterosexual	29	55.77
Total	51	100.00	Total	52	100.00

5.3.4.1. By-Subjects and By-Items Analyses

I began to identify significant differences between groups of the independent variables by conducting two separate repeated-measures Analyses of Variance (ANOVAs): one where response time and recall likelihood data was organised by-subjects (averaged across each participant), and the other by-items (averaged for each target word in the experiment). A Greenhouse-Geisser sphericity correction was applied to both repeated-measures ANOVAs. This procedure also allowed me to look for interactions between the independent variables in the 3 x 2 x 2 structure (experimental condition x gender x sexual identity). Where any interactions were significant, I conducted post-hoc tests to identify which specific groups significantly differed in terms of the interaction. Beginning by carrying out both by-subjects and by-items analyses allowed me to explore whether any significant differences and/or interaction effects were more likely arising from the participants themselves, or from the specific items used in the experiment. If the same patterns were found in both analyses, this would suggest a more generalisable effect. The results of these analyses are reported in section 5.4.1.

Past this stage, my analyses focused on exploring differences between participants, rather than differences between specific items. This is because both my hypotheses for this experiment (see sections 5.1. and 5.2.) and my overall research questions (see section 1.2.) particularly sought to investigate the potential impact of individuals' own identity and their own engagement with reclamation behaviours on their responses. Beyond exploration of broader, categorical differences (e.g. LGBTQ+ slurs versus non-taboo, high-arousal words), the present study sought less to explore differences in responses to specific items (e.g. between specific LGBTQ+ slurs). However, the patterns I identified in the by-items analysis informed my analysis and discussion of all other results, and indications of differences between specific LGBTQ+ slurs are identified and discussed in both Chapters 3 and 4 of this thesis.

5.3.4.2. Detailed Analyses of Participant Responses

After conducting by-subjects/by-items analysis of differences between the independent variables and their interaction effects, I conducted more detailed analyses of the differences between participants, using all of their responses rather than their averages. For this, I used the most

appropriate statistical measures for separate analysis of each independent variable, rather than a repeat-measures ANOVA which analysed all of them collectively.

For the lexical decision results, participant response times represented a continuous dependent variable. Because there were three experimental conditions, I explored response time differences between each using a one-way ANOVA procedure (Dörnyei, 2007; Field, 2013). Gender and sexual identity were both divided into two groups, so for these I conducted an independent-samples t-test instead (Dörnyei, 2007; Field, 2013).

Once I had analysed group differences for the full dataset, I then decided to explore whether there were differences in response times based on gender and sexual identity for each individual condition. This involved isolating the data pertaining to one condition, then running independent-samples t-tests for both gender and sexual identity comparisons on that data. Of particular interest was whether there were gender or sexual identity differences in response times to just the LGBTQ+ slurs, but I ran this procedure for all three conditions.

Because I treated free recall data as a dichotomous dependent variable (in which I recorded a word as having either been recalled or not), I needed to use different analytical procedures to look for differences based on condition, gender, and sexual identity. Because stimuli were again divided into three conditions, I selected the Chi-Square Test of Homogeneity (Agresti, 2007; Fleiss, Levin, & Paik, 2003) for that analysis. As gender and sexual identity were each split into two groups, I analysed both of these using a Test of Two Proportions (Agresti, 2007; Fleiss et al., 2003). Additionally, I ran four further Chi-Square Tests of Homogeneity looking at differences between conditions in just women's responses, just men's responses, just non-heterosexual participants' responses and just heterosexual participants' responses.

5.3.4.3. Producing Models for Lexical Decision Data

I ran a series of hierarchical multiple regressions to explore which independent variables best predicted response times. This procedure was appropriate for generating a model for a continuous dependent variable such as response time, while also allowing me to test whether the model could be improved with the addition of further independent variables (Gelman & Hill, 2008).

To make sure any model that I generated met the assumptions required for this analysis, I checked for and removed any outliers, as indicated by data points where the values of the standardised residuals and/or studentized deleted residuals were greater than ±3 standard deviations (Berry, 1993; Cook & Weisberg, 1982; Fox, 1991). I then checked the leverage values for all data points, and consistently found that all were safely below a value of .20 (Huber, 1981). Lastly, I checked to make sure that no data points were overly influential, as would be indicated by a Cook's Distance value above 1.00 (Cook & Weisberg, 1982), which none were found to be.

Having checked for the above, I then used partial regression plots and a plot of studentized residuals against predicted values to check that there was linearity in the data, which there consistently was (Berry, 1993; Fox, 1991).

There was some evidence of multicollinearity between some of my independent variables, as evidenced by correlation coefficients greater than .70. Besides a predictably positive correlation between number of characters and number of syllables, all the other independent variables which correlated above this threshold were word properties belonging to the same group of word properties identified in the correlation analysis described in Chapter 4, e.g. familiarity and personal use. To address this, I made sure that when an independent variable was entered into the model, no

other independent variables with which it strongly correlated were also entered into the model. To be sure that I was not being preferential about which independent variables I selected, I inspected all possible variations of the model in order to establish which combination actually produced the best fit.

5.3.4.4. Producing Models for Free Recall Data

I ran a series of binomial logistic regressions to explore which independent variables best predicted likelihood of word recall. This procedure was the most appropriate for generating models for a dichotomous dependent variable such as recall likelihood, for which participants could only have recalled or not recalled a word (Collett, 2003; Hilbe, 2016).

For all models that I generated using this procedure, I first checked the linearity of any continuous independent variables in the model – with respect to the logit of the dichotomous dependent variable – using the Box-Tidwell (1962) procedure. For the two recall likelihood models I describe in section 5.4.3, I applied Bonferroni correction using all five terms in both models, resulting in statistical significance being accepted when $p < .01$ (Tabachnick & Fidell, 2014). I found that in both models, all continuous independent variables in the final model were linearly related to the logit of the dependent variable, $p > .01$. Casewise diagnostics for both models revealed that there were no cases with standardised residuals greater than ±3, so no cases needed to be considered outliers and removed (Berry, 1993; Cook & Weisberg, 1982; Fox, 1991).

As all independent variables were the same as those used in the lexical decision model, there was again evidence of multicollinearity between some of these. I took the same approach in addressing this issue as I described in section 5.3.4.3.

5.4. Results

5.4.1. Results of By-Subjects and By-Items Analyses

For the by-subjects analysis, Table 5.4. provides the results for both response time and recall likelihood in the 3 x 2 x 2 (condition x gender x sexual identity) structure. Table 5.5. provides the same for the by-items analysis. In both tables, mean recall likelihood values represent a percentage, where a value of 1.00 would indicate a 100% likelihood of recall within a particular group of an independent variable. Significant differences are indicated in both tables.

Table 5.4. – By-Subjects Analysis of Response Time and Recall Accuracy

	Group	Mean RT	SD	Mean Recall Likelihood	SD
Condition	LGBTQ+ Slurs	976.96	287.74	0.35[c]	0.48
	Negative High-Arousal	969.49	262.86	0.16[c]	0.37
	Negative Low-Arousal	973.01	262.78	0.11[c]	0.32
Gender	Women	993.02[a]	259.50	0.24	0.42
	Men	927.09[a]	286.23	0.25	0.44
Sexual Identity	Non-Heterosexual	1011.52[b]	338.24	0.26	0.44
	Heterosexual	941.87[b]	205.38	0.24	0.43

[a] Difference was significant, $p < .001$.

[b] Difference was significant, $p < .001$.

[c] Post-hoc tests revealed differences between all categories were significant, all $p < .001$.

Table 5.5. – By-Items Analysis of Response Time and Recall Accuracy

	Group	Mean RT	SD	Mean Recall Likelihood	SD
Condition	LGBTQ+ Slurs	971.07	29.77	0.39[c]	0.16
	Negative High-Arousal	955.88	10.42	0.15[c]	0.12
	Negative Low-Arousal	967.92	36.17	0.12[c]	0.08
Gender	Women	999.42[a]	24.68	0.27	0.20
	Men	933.73[a]	55.83	0.26	0.19
Sexual Identity	Non-Heterosexual	998.63[b]	43.06	0.24	0.19
	Heterosexual	934.53[b]	28.16	0.28	0.22

[a] Difference was significant, $p < .001$.

[b] Difference was significant, $p < .001$.

[c] Post-hoc tests revealed significant differences between LGBTQ+ slurs and Negative High-Arousal words ($p < .001$), and LGBTQ+ slurs and Negative Low-Arousal words ($p < .001$).

For the by-subjects analysis, interactions analyses revealed no significant 3-way nor 2-way interactions between the independent variables, for neither response time nor recall likelihood. For the by-items analysis, interactions analyses revealed a significant 2-way interaction of Gender * Sexual Identity on lexical decision response time ($F(1, 48) = 16.37$, $p < .001$, partial $\eta^2 = .26$) and on recall likelihood ($F(1, 48) = 13.06$, $p < .001$, partial $\eta^2 = .22$). Post-hoc tests for the interaction effect on response times revealed non-heterosexual women had significantly slower response times to all words than non-heterosexual men (mean difference 103.65ms, SD: 13.50ms, $p < .001$), heterosexual women (95.97ms, SD:11.23ms, $p < .001$), and heterosexual men (126.22ms, SD: 11.95ms, $p < .001$). Post-hoc tests for the interaction effect on recall likelihood revealed non-heterosexual women were likelier to recall all words than non-heterosexual men (mean difference 9.50%, SD: 3.00%, $p < .05$); non-heterosexual men were less likely to recall all words than heterosexual men (mean difference -12.30%, SD: 3.40%, $p < .01$); and heterosexual women were less likely to recall all words than heterosexual men (mean difference -7.70%, SD: 3.00%, $p = .05$).

5.4.2. Results of Detailed Analyses of Participant Responses

Here, I report the results of the detailed analyses of participant responses described in section 5.3.4.2., separated by analyses of response time in the lexical decision task, and of likelihood of recall in the free recall task.

5.4.2.1. Lexical Decision Task Responses

Response accuracy to all target words across all in the lexical decision task was high, with an overall response accuracy of 92.11%. Accordingly, the results reported in this section are not considered to have been affected by the general accuracy of participants' lexical decisions.

I conducted a one-way ANOVA to test whether there were any significant differences in response time for each condition. The analysis found no significant differences in response times between different conditions, $F(2, 2389) = .15$, $p = .86$.

I then conducted two independent-samples t-tests to check for response time differences based on participant gender and sexual identity.

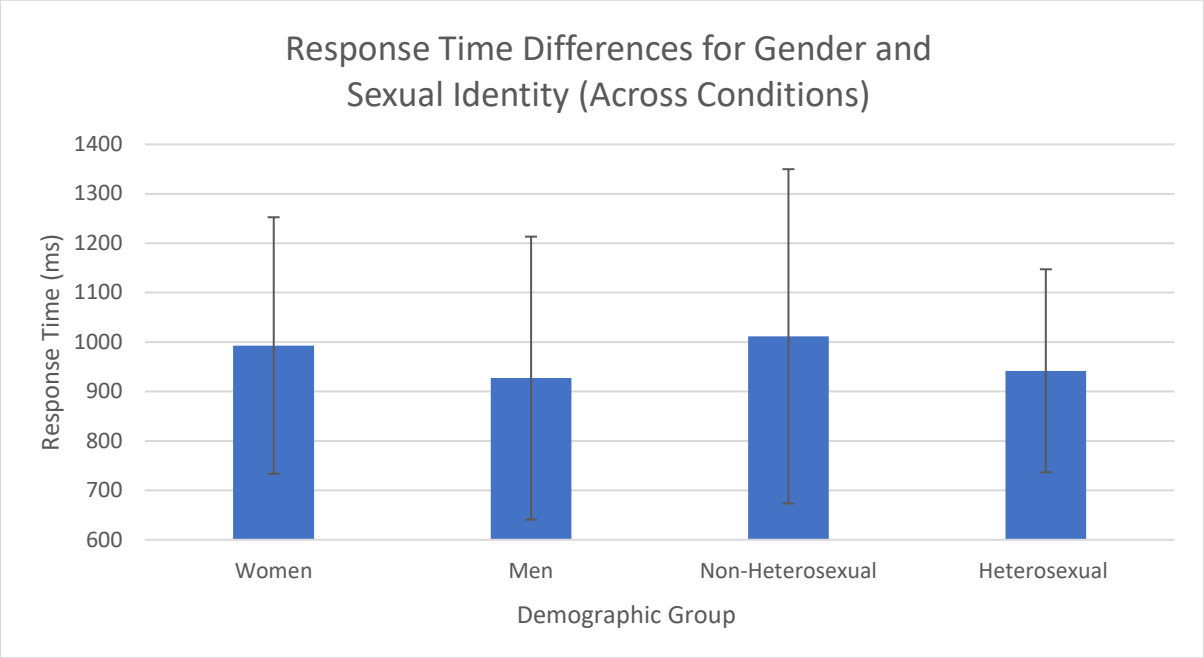
For gender, a Welch t-test was run due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances ($p < .001$). There were 38 women and 13 men. Results indicated that response times were slower among women (993.02ms, SD:

259.50ms) than among men (927.09ms, SD: 286.23ms), a significant difference of 65.93ms (95% CI, 38.97ms to 92.88ms) $t(856.10) = 4.80$, $p < .001$.

For sexual identity, a Welch t-test was run due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances ($p < .001$). There were 23 non-heterosexual participants and 29 heterosexual participants. Results indicated that response times were slower among non-heterosexual participants (1011.52ms, SD: 338.24ms) than among heterosexual participants (941.87ms, SD: 205.38ms), a significant difference of 69.65ms (95% CI, 46.17ms to 93.11ms) $t(1654.97) = 5.82$, $p < .001$. The results of both t-tests are visualised in Figure 5.1.

I then examined response times for each condition, and investigated whether there were significant response time differences based on both gender and sexual identity within each. This involved running six independent-samples t-tests in total: a gender and sexual identity comparison for each of the three conditions. All were run as Welch's t-tests, due to the assumptions of homogeneity of variances being violated, as assessed by Levene's tests for equality of variances (all $p < .05$). There were 38 women and 13 men in all gender comparisons, and 23 non-heterosexual participants and 29 heterosexual participants in all sexual identity comparisons.

Figure 5.1. – Response Time Differences for Gender and Sexual Identity (Across Conditions)



For the LGBTQ+ slurs, results indicated that response times were slower among women (998.08ms, SD: 271.34ms) than among men (931.94ms, SD: 300.76ms), a significant difference of 66.14ms (95% CI, 25.59ms to 106.69ms) $t(419.52) = 3.21$, $p = .001$. Results also indicated that response times were slower among non-heterosexuals (1018.01ms, SD: 359.27ms) than among heterosexuals (940.57ms, SD: 207.03ms), a significant difference of 77.44ms (95% CI, 41.96ms to 112.92ms) $t(773.25) = 4.29$, $p < .001$.

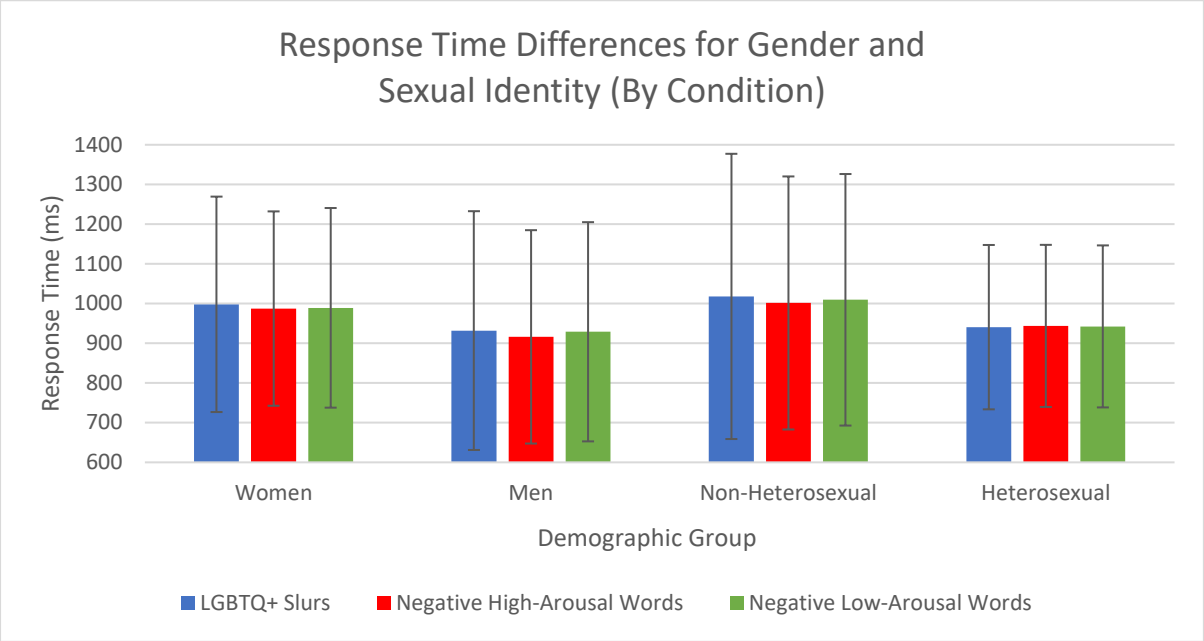
For the negative high-arousal words, results indicated that response times were slower among women (987.38ms, SD: 244.83ms) than among men (916.12ms, SD: 268.75ms), a significant difference of 71.26ms (95% CI, 21.24ms to 121.26ms) $t(221.01) = 2.81$, $p = .01$. Results also indicated that response times were slower among non-heterosexuals (1001.50ms, SD: 318.74ms) than among

heterosexuals (943.72ms, SD: 204.19ms), a significant difference of 57.78ms (95% CI, 13.85ms to 101.74ms) $t(449.94) = 2.59$, $p = .01$.

For the negative low-arousal words, results indicated that response times were slower among women (989.23ms, SD: 251.47ms) than among men (928.94ms, SD: 276.11ms), a significant difference of 60.29ms (95% CI, 8.31ms to 112.26ms) $t(212.55) = 2.29$, $p = .02$. Results also indicated that response times were slower among non-heterosexuals (1009.50ms, SD: 316.83ms) than among heterosexuals (942.45ms, SD: 204.07ms), a significant difference of 67.05ms (95% CI, 22.86ms to 111.24ms) $t(434.55) = 2.98$, $p < .01$. The results of all three sets of t-tests are visualised in Figure 5.2.

Collectively, these results suggest that women consistently provided slower response times in all conditions when compared to men, and the greatest of these differences was in response times for negative high-arousal words. Non-heterosexual participants also consistently provided slower response times in all conditions when compared to heterosexual participants, and the greatest of these differences was in response times for LGBTQ+ slurs.

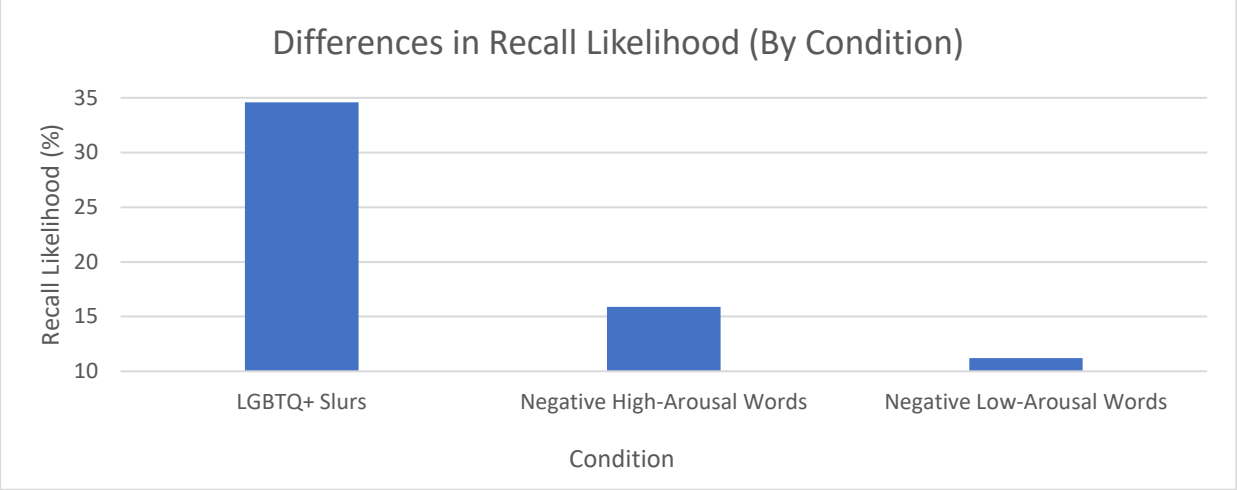
Figure 5.2. – Response Time Differences for Gender and Sexual Identity (By Condition)



5.4.2.2. Free Recall Responses

I conducted a chi-square test of homogeneity to look for differences in recall likelihood based on condition. Results indicated that 34.60% of LGBTQ+ slurs were recalled, compared to 15.90% of negative high-arousal words, and 11.20% of negative low-arousal words, a significant difference in proportions, $p < .001$. Post hoc analysis involved pairwise comparisons using the z-test of two proportions with a Bonferroni correction. All pairwise comparisons between conditions were found to be significant. These results are visualised in Figure 5.3.

Figure 5.3. – Differences in Recall Likelihood (By Condition)



I then conducted two chi-square tests of homogeneity to check if there were any differences in recall likelihood based on participant gender and sexual identity groups.

Gender did not have a significant effect: women recalled 23.50% of words, while men recalled 24.30% of words ($p = .67$). Sexual identity also did not have a significant effect: non-heterosexual participants recalled 25.80% of words, while heterosexual participants recalled 23.50% of words ($p = .17$).

Finally, I conducted four further chi-square tests of homogeneity to test if there were any recall likelihood differences between conditions, within each individual participant demographic group: women, men, non-heterosexual and heterosexual.

For women, results indicated that 33.30% of LGBTQ+ slurs were recalled, compared to 16.40% of negative high-arousal words and 10.10% of negative low-arousal words, a significant difference in proportions, $p < .001$. Post hoc analysis involved pairwise comparisons using the z-test of two proportions with a Bonferroni correction. All pairwise comparisons between conditions were found to have been significant, $p < .05$.

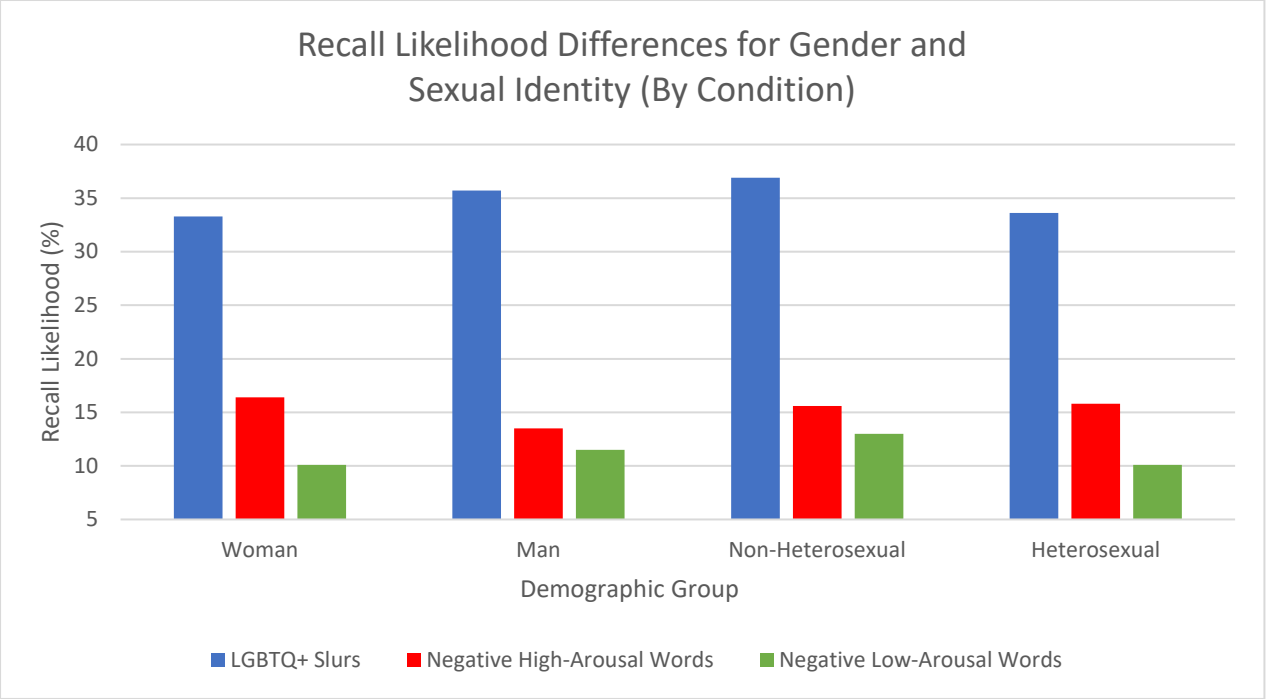
For men, results indicated the same pattern as for women: 35.70% of LGBTQ+ slurs were recalled, compared to 13.50% of negative high-arousal words and 11.50% of negative low-arousal words, a significant difference in proportions, $p < .001$. Post hoc analysis involved pairwise comparisons using the z-test of two proportions with a Bonferroni correction. Among men, the proportion of words recalled was found to be significantly higher for LGBTQ+ slurs than for negative high-arousal or negative low-arousal words, $p < .05$. However, there was no significant difference in the proportion of words recalled between negative high-arousal and negative low-arousal words, $p > .05$.

For non-heterosexual participants, results indicated that 36.90% of LGBTQ+ slurs were recalled, compared to 15.60% of negative high-arousal words and 13.00% of negative low-arousal words, a significant difference in proportions, $p < .001$. Post hoc analysis involved pairwise comparisons using the z-test of two proportions with a Bonferroni correction. Among non-heterosexual participants, the proportion of words recalled was found to be significantly higher for LGBTQ+ slurs than for negative high-arousal or negative low-arousal words, $p < .05$. However, there was no significant difference in the proportion of words recalled between negative high-arousal and negative low-arousal words, $p > .05$.

For heterosexual participants, results indicated the same pattern as for non-heterosexual participants: 33.60% of LGBTQ+ slurs were recalled, compared to 15.80% of negative high-arousal words and 10.10% of negative low-arousal words, a significant difference in proportions, $p < .001$. Post hoc analysis involved pairwise comparisons using the z-test of two proportions with a Bonferroni correction. Among heterosexual participants, the proportion of words recalled was found to be significantly higher for LGBTQ+ slurs than for negative high-arousal or negative low-arousal words, $p < .05$. However, there was no significant difference in the proportion of words recalled between negative high-arousal and negative low-arousal words, $p > .05$. The results of all four chi-square tests of homogeneity are visualised in Figure 5.4.

Collectively, these results suggest that LGBTQ+ slurs were over twice as likely to be recalled as either negative high-arousal or negative low-arousal words, and this was true regardless of gender and sexual identity, for which there were no significant differences across conditions. Generally, there were no significant differences in recall likelihood between negative high-arousal and negative low-arousal words, except among women, who were significantly likelier to recall negative high-arousal words compared to negative low-arousal words.

Figure 5.4. – Recall Likelihood Differences for Gender and Sexual Identity (By Condition)



5.4.3. Results of Modelling Lexical Decision Responses

I found that the best model of response times across all words in the lexical decision task included just participant gender and sexual identity. The model significantly predicted response times and explained 1.0% of the variance in response times, $F(2,2082) = 9.11$, $p < .001$, adj. $R^2 = .01$. Both variables added significantly to the prediction, $p < .01$. The model predicted that response times were faster when participants identified as a man compared to as a woman, and when participants identified as heterosexual compared to non-heterosexual. I provide the regression coefficients and standard errors in Table 5.6.

Table 5.6. – Model of Response Times to All Words in Lexical Decision Task

Response Times (All Words)	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	*UL*				
Model						.01	.01***
Constant	1031.21***	990.53	1071.89	20.78			
Gender	-34.77***	-55.56	-13.99	10.61	-.07		
Sexual Identity	-25.96**	-44.57	-7.35	9.51	-.06		

Note. Model = "Enter" method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardised coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2. ** $p < .01$, *** $p \leq .001$.

The best model of response times to the LGBTQ+ slurs specifically in the lexical decision task was found to include participant gender, sexual identity and participants' own overall reclamation behaviour ratings. The model significantly predicted response times and explained 2.0% of the variance in response times, $F(3,994) = 7.25$, $p < .001$, adj. $R^2 = .02$. All three variables added significantly to the prediction, $p < .05$. The model predicted that response times to LGBTQ+ slurs were faster when participants identified as a man compared to as a woman, when participants identified as heterosexual compared to non-heterosexual, and when participants gave higher personal overall reclamation behaviour ratings for a slur. I provide the regression coefficients and standard errors in Table 5.7.

Table 5.7. – Model of Response Times to LGBTQ+ Slurs in Lexical Decision Task

Response Times (LGBTQ+ slurs)	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	*UL*				
Model						.02	.02***
Constant	1097.38***	1030.53	1164.22	34.07			
Gender	-35.27*	-65.64	-4.91	15.47	-.07		
Sexual Identity	-42.73**	-70.41	-15.04	14.11	-.10		
Participant Overall Recl	-25.19**	-40.82	-9.56	7.97	-.10		

Note. Model = "Enter" method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardised coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2; Participant Overall Recl = mean of participant's own reclamation behaviour ratings for an LGBTQ+ slur. * $p < .05$, ** $p < .01$, *** $p < .001$.

5.4.4. Results of Modelling Free Recall Responses

The logistic regression model which best explained overall recall likelihood included normed Familiarity and Negativity word ratings. The interactions described by the model were significant, $\chi^2(2) = 67.90$, $p < .001$. A Hosmer and Lemeshow goodness of fit test indicated that the model was not a poor fit, $p > .05$. The model explained 4.0% (Nagelkerke R^2) of the variance in recall likelihood and correctly classified 75.70% of recall results. Sensitivity was .00%, specificity was 100.00%, positive predictive value was 100.00% and negative predictive value was 75.74%. Increased familiarity and decreased negativity ratings for a word were found to be jointly associated with an increased likelihood of recall, as indicated in Table 5.8.

Table 5.8. – Model of Recall Likelihood for All Words in Free Recall Task

Recall likelihood (All Words)	*B*	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							LL	*UL*
Constant	-1.78	.47	14.44	1	< .001	.17		
Familiarity	.51	.09	31.65	1	< .001	1.67	1.40	2.00
Negativity	-.28	.08	14.04	1	< .001	.75	.65	.87

The logistic regression model which best explained recall likelihood for just the LGBTQ+ slurs included normed Tabooness and Self-Reclamation ratings. The interactions described by the model were significant, $\chi^2(2) = 55.61$, $p < .001$. A Hosmer and Lemeshow goodness of fit test indicated that the model was not a poor fit, $p > .05$. The model explained 6.0% (Nagelkerke R^2) of the variance in recall likelihood and correctly classified 67.70% of recall results. Sensitivity was 14.80%, specificity was 95.60%, positive predictive value was 64.15% and negative predictive value was 68.01%. Increased self-reclamation and tabooness ratings for a slur were found to be jointly associated with an increased likelihood of recall, as indicated in Table 5.9.

Table 5.9. – Model of Recall Likelihood for LGBTQ+ Slurs in Free Recall Task

Recall likelihood (LGBTQ+ Slurs)	*B*	SE	Wald	*df*	*P*	Odds Ratio	95% CI for Odds Ratio	
							LL	*UL*
Constant	-5.57	.80	48.49	1	< .001	.004		
Self-Reclamation	1.34	.19	51.84	1	< .001	3.83	2.66	5.51
Tabooness	.91	.17	27.25	1	< .001	2.48	1.76	3.49

5.5. Discussion

5.5.1. Lexical Decision Results

In this section of my discussion, I have grouped the key findings from the lexical decision task. I begin by discussing the results of the by-subjects and by-items analyses, then consider the results of the more focused comparisons I made by experimental condition (LGBTQ+ slurs, negative high-arousal words and negative low-arousal words), gender, and sexual identity. I then discuss what is indicated by the models which predicted response time variance, first to all words in the experiment, and then specifically to the LGBTQ+ slurs.

5.5.1.1. By-Subjects and By-Items Analyses

For the lexical decision responses, the results of the by-subjects and by-items analyses largely showed the same patterns: regardless of whether response times were averaged for individual participants or individual target words, there were no significant differences between the three experimental conditions, but the response times of women and non-heterosexual participants to all target words were slower than those of men and heterosexual participants. I discuss each of these findings in more detail in sections 5.5.1.3. – 5.5.1.5., in relation to the more focused analyses of these variables that I conducted. However, it is important at this stage to note that these differences were significant in both types of analysis. This suggests that these differences represent more generalisable patterns, and cannot be solely attributed to differences between the individual participants in my sample, nor between the specific items that I selected to include in my

experiment. Accordingly, I argue that my subsequent discussions of these results in relation to broader patterns of social behaviour cannot be dismissed as products of the experiment design.

There was one notable difference between the by-subjects and by-items analyses of lexical decision responses: the by-items analysis revealed a two-way interaction effect of gender and sexual identity, which reached trend but not statistical significance in the by-subjects analysis. In line with the broader differences between groups of gender and sexual identity, the interaction effect suggested that non-heterosexual women gave significantly slower response times than all other groups, with the biggest difference being between non-heterosexual women and heterosexual men. Because the interaction effect mirrored these broader patterns, it is possible that it arises from a combination of reasons for the gender and sexual identity differences in my data, which I discuss in sections 5.5.1.4 and 5.5.1.5. However, I note that this interaction effect was only significant when results were averaged for individual items in the experiment, suggesting it is more closely attributable to differences between the specific items that were included. I argue that the explanation for this is likely quite straightforward: half of the items in the experiment were sexuality-based slurs, the majority of which are also specifically gendered (particularly toward men, as I discuss further in section 5.5.1.4). It may therefore be expected that a combined effect of gender and sexuality may have been present when examining average responses to individual items, that was not observed for average differences in response times for participants themselves. Accordingly, while I do consider this interaction effect in relation to the other results observed in my study, I do not consider it an explanation for the more generalisable, separate differences observed for gender and sexual identity in my results.

5.5.1.2. Differences between Conditions

As indicated by my by-subjects and by-items analyses, and clarified by the results outlined in section 5.4.2.1., there were no significant differences in response times between the three experimental conditions across all participant responses. This finding is inconsistent with those of previous studies. As I discussed in section 5.1., the literature indicates that taboo words like the LGBTQ+ slurs should be associated with significantly slower response times than non-taboo words (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005; Williams & Evans, 1980). It was also unexpected that there were no differences found between the two non-taboo conditions that differed by arousal; research suggests that among non-taboo language, emotional arousal should be associated with faster response times in lexical decision tasks (Graves, Landis, & Goodglass, 1981; Strauss, 1983; Williamson, Harpur, & Hare, 1991). On this basis, my lexical decision results did not support the first of my hypotheses set out in section 5.2.: that there would be a significant difference in response times between the LGBTQ+ slurs and the two non-taboo conditions in the experiment.

As I note in section 2.3.1. of my literature review, the body of published work on taboo language and lexical decision is notably smaller than that for either word recallability or attention while reading. Nevertheless, it is unexpected that my experiment did not replicate either of the key findings I describe in the previous paragraph. I consider there to be two main potential explanations for this.

The first is the number of variables that can affect response times in lexical decision tasks. For example, Madan et al. (2017) demonstrated that lexical decision times in their study were best predicted by a model which included word familiarity, personal use of a word, word frequency, contextual diversity, number of letters, number of syllables and tabooness of a word. I detail all of

the ways in which I attempted to control for these variables in section 5.3.2.1, but it is possible that I was not able to do so to an extent which enabled a valid comparison by condition alone. It is worth noting that of all the factors which were included in Madan et al.'s model, only familiarity and tabooness appeared in any of my own, although for free recall rather than lexical decision. Possibly, these factors in particular should be more strictly controlled in future lexical decision experiments, so as to more accurately determine if and how they influence results. However, they are both quite difficult to control for beyond establishing general population norms, as both frequency of exposure to a word and personal judgments of its social acceptability can vary greatly from person to person.

The second is that the only taboo words in this experiment were LGBTQ+ slurs. I acknowledge that this cannot account for the lack of difference between the two non-taboo conditions in the experiment, for which a difference in processing time was also expected. However, it may be able to account for a lack of difference between taboo and non-taboo words across all responses, considering first that slurs may present a unique case compared to other kinds of taboo language, and that slurs as a particular category of taboo language are unlikely to affect all participants in the same way. As I discussed in section 2.2. of my literature review and section 5.1. of this chapter, slurs by definition target particular communities and demographic groups (Chen, 1998; Croom, 2011; Herbert, 2015; Tirrell, 2012). As can be seen in Tables 5.1. and 5.3., there was a split among participants in terms of whether or not they might be the target of LGBTQ+ slurs, so valid comparisons of LGBTQ+ slurs to non-taboo language across all participants might never have been possible. However, participant demographic data did enable more specific demographic comparisons to be made, which supported this second explanation and lent valuable insight into what most affected response times in the experiment.

5.5.1.3. Gender Differences

As indicated by my by-subjects and by-items analyses, and clarified by the results outlined in section 5.4.2.1., results indicated a significant effect of gender on response times across all words in the experiment. Specifically, women's response times were found to be slower than those of men. This finding was consistent when response times for each individual condition: women provided slower response times to negative low-arousal words, negative high-arousal words, and LGBTQ+ slurs. The greatest difference in response times between women and men was found for the negative high-arousal words, followed by the LGBTQ+ slurs, and finally the negative low-arousal words.

It is important to note that the experiment sample was less balanced for gender than it was for sexual identity; as Table 5.3. indicates, there were nearly three women for every man in the experiment, and as such these results should be considered more cautiously. However, the gender difference in response time was consistent across all conditions in the experiment, and typically reached high levels of significance ($p \leq .01$). Furthermore, there is some precedent in the literature for a gender difference in response to taboo and emotionally arousing stimuli. Geer and Bellard (1996) found that in an unprimed lexical decision experiment, women displayed significant delays in response times to both sexual (taboo) and romantic (non-taboo, emotionally arousing) verbal stimuli. A similar delay was observed among men, but it was smaller and was not significant.

For LGBTQ+ slurs in particular, there is also some precedent for a gender difference in my earlier studies. When discussing the results of my LGBTQ+ norming study (see sections 4.4.2.2., 4.5.3. and 4.5.4. for full details), I observed that women consistently reported lower familiarity, personal use, and reclamation behaviour ratings for LGBTQ+ slurs than men. Conversely, women

reported higher arousal, negativity, offensiveness, and tabooness ratings for LGBTQ+ slurs than men. A combination of all of these word properties could have caused differences in response times. However, as I mention in section 5.2., the most important of these is perhaps the higher tabooness ratings among women; increased perceptions of tabooness have often been associated with impaired lexical accessibility (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005; Williams & Evans, 1980).

However, these findings do not as readily explain the gender difference for the two non-taboo conditions. While Geer and Bellard's (1996) study did find a delay in women's response times to a category of non-taboo, emotionally-arousing language, other studies have suggested that among non-taboo language, increased emotional arousal should improve lexical accessibility regardless of the gender of the participant (Graves, Landis, & Goodglass, 1981; Strauss, 1983; Williamson, Harpur, & Hare, 1991).

To attempt to resolve this contradiction, I consider two related possibilities. The first explanation is one which I outlined in section 4.5.4., when discussing the gender differences in word ratings for LGBTQ+ slurs. I pointed to a body of sociological and particularly sociolinguistic research which has identified a particular burden on women to use forms of language which are felt to be more polite and indirect (e.g. Brown, 1980; Coates, 1988; Fishman, 1980; Holmes, 1995), a gendered speech norm developing from pervasive, patriarchal and heteronormative socialisation from an early age (Cameron, 2006). This socialisation is not something experienced in the same way by all women, but I suggested that as a broader phenomenon, it may explain why women more consistently rated all LGBTQ+ slurs more negatively than men did, regardless of whether the LGBTQ+ slur in question was clearly gendered. In this regard, this explanation might help explain the response time delay among women to the LGBTQ+ slurs in this experiment: if women generally respond with greater shock and aversion to slurs, such reactions might have contributed to the delay to women's response times. It may also help explain women's slower response times to the negative high-arousal words. Referring to the negative high-arousal words (examples in Table 5.2, full list in Appendix 5), it is clear that these would not be likely to be considered polite, indirect or emotionally neutral; indeed, this category was formed on the basis that they are not. If it is more common that women rather than men are taught to avoid language like this, then it may explain why women's lexical access for negative, high-arousal words was more impeded. It may also suggest that the perceptual defence mechanism (Williams & Evans, 1980; Madan et al., 2017) proposed for why tabooness is associated with delayed lexical access might respond to certain non-taboo stimuli as well, if social factors like these are also considered. Finally, it might explain why the difference in response times between women and men was greater for negative high-arousal words than it was for LGBTQ+ slurs, despite the slurs being the examples of taboo language in the experiment. Both men and women are likely to have adverse reactions to the tabooness of a slur. However, if there is reason to believe that women might have an aversion to language which is not taboo but is both negative and shocking, and that the cause of this is not as widely experienced by men, then it makes sense that a gender difference in response times might be most pronounced for these words.

However, these possibilities still do not account for the women's slower response times to words which were neither taboo nor highly emotionally arousing. For this, I consider a second explanation. If there is good reason to expect that women would have delayed response times to LGBTQ+ slurs – and, possibly, to the negative high-arousal words as well – then it is possible that the simple presence of such words in the experiment caused increased delays in women's response times to all stimuli. The potential of such a 'knock-on' effect on the negative low-arousal words in

particular is supported by the fact that the gender difference in response times for these words was the smallest.

5.5.1.4. Sexual Identity Differences

As indicated by my by-subjects and by-items analyses, and clarified by the results outlined in section 5.4.2.1., results indicated a significant effect of sexual identity on response times across all words in the experiment. Specifically, non-heterosexual participants' response times were found to be slower than those of heterosexual participants. As with the gender differences, this finding remained consistent when response times for each condition were isolated: non-heterosexual participants provided slower response times to negative low-arousal words, negative high-arousal words and LGBTQ+ slurs. The greatest difference in response times between non-heterosexual and heterosexual participants was found for the LGBTQ+ slurs, followed by the negative low-arousal words, and finally the negative high-arousal words. All sexual identity differences reached a high level of significance (all *p*s ≤ .01).

The second hypothesis I set out in section 5.2. predicted that increasing reclamation of LGBTQ+ slurs would reduce any impact of arousal and tabooness on response times, causing them to be less delayed. For the same reasons, my third hypothesis predicted that any such effect of reclamation on response times would be strongest in the demographic group(s) likelier to engage in it, particularly the LGBTQ+ participants. While these results alone could not establish that the difference found between sexual identity groups was related to an effect of reclamation, they support my expectation that sexual identity might have modulated the processing of the LGBTQ+ slurs, in that they did identify a response time difference between the two sexual identity groups. However, this difference was the opposite of the one I had predicted. Instead of being faster, the response times of non-heterosexual participants were significantly more delayed than those of the heterosexual participants they were being compared to.

This is somewhat difficult to explain, as there is no basis in the literature to explain an overall difference in lexical decision performance based on sexual identity, as opposed to the small precedent that does exist for gender differences (e.g. Geer & Bellard, 1996). I suggest that the likeliest explanation is that half of the words in the experiment specifically and negatively target members of the LGBTQ+ community. If the presence of the LGBTQ+ slurs had a different effect on non-heterosexual and heterosexual participants, then their inclusion alone may have been sufficient to drive a difference in response times based on sexual identity.

Initially, it may seem clear why response times would have been slower among non-heterosexual participants than heterosexual ones, particularly to the LGBTQ+ slurs. Similar to the overall gender difference in results, this difference does suggest greater aversion to the LGBTQ+ slurs among the non-heterosexual participants, particularly since the greatest difference in response times between non-heterosexual and heterosexual participants was found for LGBTQ+ slurs. Given the nature of slurs discussed in section 2.2.1., it makes sense that those more targeted by such derogatory language would display the most aversion to it. In section 5.5.1.3., I suggested reasons why women's responses may have also reflected greater aversion to the LGBTQ+ slurs, so combining these findings may also explain the by-items interaction effect that indicated non-heterosexual women provided the slowest responses.

However, these results are inconsistent with some of the findings from my norming study in Chapter 4. In section 5.5.1.3., I made a connection between the slower response times I found for

women and women's responses to my LGBTQ+ slur norming study. I link the two findings because the broader literature suggests an association between giving lower non-emotional word property ratings, giving higher tabooness ratings and producing slower overall response times (Madan et al., 2017). While I observed a difference based on sexual identity in my norming study results, the lower ratings for non-emotional word properties and higher ratings for tabooness were within the heterosexual rather than non-heterosexual group. On this basis, slower response times should have been expected among the heterosexual group, but these were found to have been faster than the non-heterosexual group in the experiment.

Additionally, it is reasonable to expect that non-heterosexual participants may have been more frequently exposed to LGBTQ+ slurs than heterosexual participants. This is supported by the sexual identity differences in my norming study results, in which familiarity ratings were typically higher in the non-heterosexual group when compared with the heterosexual group (see section 4.4.2.3.). However, the same study identified that higher familiarity ratings were associated with lower arousal, negativity, offensiveness, and tabooness ratings. Because such properties are associated with increased delays in taboo language processing, the group with greater familiarity should be expected to show reduced delays, but here the opposite was true.

Without other findings to refer to, it is difficult to explain these apparent contradictions in my results. It is possible that the non-heterosexual participants in the experiment were more averse to encountering LGBTQ+ slurs than the ratings supplied by the larger sample of non-heterosexual participants in my norming study would suggest. If true, this reaction may have had a greater impact on results than any specific word properties I measured. This possibility is supported by the fact that almost none of these word properties contributed to the models of lexical decision response times displayed in Tables 5.4. and 5.5.

I also consider that the relationship between increased familiarity and reduced emotional responses may not always hold if that level of familiarity is attained through repeatedly being made the object of highly derogatory language. Any such effect of LGBTQ+ slurs on non-heterosexual participants may also have been enhanced by the contrast of the LGBTQ+ slurs with the entirely non-taboo words in the study, a possibility I explore further when discussing comparisons made between LGBTQ+ slurs and other taboo words in the next chapter (see section 6.5.1.4.).

5.5.1.5. Best Model of Lexical Decision Response Times Across Conditions

I now turn to the combination of word properties measured in the experiment which best predicted response times in the lexical decision task, across all words and all participants. As presented in Table 5.7., the best model of factors influencing response times included just participant gender and sexual identity, and predicted that response times would be slower among women and slower among non-heterosexual participants, which is consistent with the by-items interaction effect and broader patterns discussed in sections 5.5.1.1., 5.5.1.3. and 5.5.1.4.

I make two initial observations about these results. First, although significant at the $p < .001$ level, the model was only able to account for 1% of the variation in response times. That this was the effect size of the best model I was able to produce suggests either that response times generally varied quite unpredictably, or that there were one or more variables I did not control for which could have improved the ability of the model to predict results. As I suggested in section 5.5.1.2., this is representative of a broader problem with response time studies like lexical decision tasks, in which response time measurements can be influenced by a wide variety of factors that are difficult

to collectively control for. Considering this, I argue that the model is still useful for having identified a combination of factors which was able to predict response times with success.

This brings me to my second observation, concerning the nature of the predictions the model made. Again, while the findings from Madan et al. (2017) indicated lexical decision times were best predicted by 'non-emotional' factors (e.g. familiarity, personal use, word frequency, number of letters and number of syllables) and tabooness, none of these added significantly to the model for my lexical decision results. Instead, only participant gender and sexual identity added significantly to the model. I suggest that one explanation for this is similar to that which I provided for the lack of difference in response time by condition: the main respect in which this experiment differed from Madan et al.'s was in its specific focus on slurs as a class of taboo stimulus, and it is possible that the factors they described as non-emotional are not as effective at predicting variation in response times when this is the case. This is supported by the fact that gender and sexual identity were the only factors that significantly contributed to a model of overall response times, since both are of central importance to LGBTQ+ slurs in particular. I consider the model to be further evidence in support of the explanations I provided for these differences in sections 5.5.1.3. and 5.5.1.4.: that responses to slurs may be influenced by gendered social norms regarding the use of taboo language, and by ingroup/outgroup dynamics concerning whom particular slurs target. The latter is especially relevant to the last key finding that emerged from the lexical decision results.

5.5.1.6. Best Model of Lexical Decision Response Times for LGBTQ+ Slurs

In order to explore if reclamation behaviour had any effect on performance in the lexical decision task for the LGBTQ+ slurs, I created a model to predict just the LGBTQ+ slur response times. As Table 5.7. shows, gender and sexual identity again predicted response times, in the same manner discussed for all words in section 5.5.1.5. However, participants' own overall reclamation ratings – the mean of each reclamation behaviour rating provided by each participant during the survey portion of the experiment – improved the model. It revealed that increasing participation in and awareness of reclamation of a slur could predict a decrease in response time during the lexical decision task. This model was significant at the $p < .001$ level, and could account for 2% of the variance in response times to LGBTQ+ slurs. Again, I acknowledge that this is limited predictive power for the best model to be produced, and attribute this to the same issues I describe in section 5.5.1.5.

As raised in section 5.3.4.3., multicollinearity between word property ratings was an issue in the data, and so I could not insert participants' own overall reclamation behaviours alongside the normed familiarity, personal use and reclamation behaviour ratings, nor each individual reclamation behaviour rating given by each participant. This meant that it was not immediately clear whether it was participants' overall personal experience of reclamation that was reducing response times, or if this effect was being caused by a related variable, such as increasing familiarity. However, I attempted to insert all such variables into the model in place of participants' own overall reclamation ratings, and found that none added significantly to the model when substituted in.

This is one of the most important findings from the experiment. It is consistent with the findings from my norming study in Chapter 4 that increased overall reclamation ratings were associated with decreased tabooness ratings (see Table 4.6.), which existing research associates with faster response times (Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005; Williams & Evans, 1980). It is also important to note that it was the combination of every reclamation behaviour

I measured – personal reclamations and awareness of other people's reclamations, in both self-reference and reference to others – that improved the model, rather than any particular one of these. Potentially, response times are more likely to be reduced when reclamation is experienced and engaged with more often and in multiple contexts. Most significantly, it appears to confirm the second of my hypotheses set out in section 5.2., in that participants' lexical access was found to be faster when they gave higher overall reclamation ratings, suggesting that reclamation may reduce the impairment to lexical access typically observed for taboo language.

This finding also supports my decision to measure participants' individual engagement with reclamation behaviours, despite having 'general population' measures of this from my norming study. I did this so as not to assume that reclamation was less important at an individual level than it was at a collective one, nor that reclamation ratings across my experiment participants would necessarily match normed ratings generated by different participants, given that attitudes towards and participation in reclamation projects can be extremely individualised (Brontsema, 2004). That the model was improved by a combination of participants' own reclamation behaviour ratings – rather than any of the normed reclamation behaviour ratings – justifies this approach.

While these findings do seem to support my second hypothesis, they are particularly strong evidence against my third: that any effect of reclamation on the processing of LGBTQ+ slurs should be expected to be strongest among the non-heterosexual participants, as the likeliest group to be engaging in and aware of their reclamation. Both sexual identity and a measure of reclamation were in the model for LGBTQ+ slurs specifically, but suggested opposite effects: slower response times were associated with non-heterosexual participants, but faster response times were associated with increased overall reclamation ratings given by the participants themselves. I therefore consider that participants' involvement in and awareness of reclamation may not be neatly attributed to a particular demographic, and that the effects of both reclamation and sexual identity in my results may be experienced separately.

5.5.2. Free Recall Results

I now discuss the key findings from the portion of my experiment involving the free recall task. As with discussion of the lexical decision results, I begin with the by-subjects and by-items analyses, then consider the results of the comparisons I made by experimental condition, gender and sexual identity. I then discuss what is indicated by the models which predicted recall likelihood, first to all words in the experiment, and then specifically to the LGBTQ+ slurs.

5.5.2.1. By-Subjects and By-Items Analyses

As with the lexical decision responses, the results of the by-subjects and by-items analyses for free recall largely showed the same patterns: regardless of whether response times were averaged for individual participants or individual target words, there were significant differences between the three experimental conditions, but no significant differences between groups of gender or sexual identity. I discuss each of these findings in more detail in sections 5.5.2.2. – 5.5.2.5., in relation to the more focused analyses of these variables that I conducted. Again, it is important to note that these significant differences were present in both types of analysis, so again likely represent more generalisable patterns. However, I also observed that while significant differences between experimental conditions were present in both by-subjects and by-items analysis, the nature of these differences was slightly different depending on whether recall likelihood was averaged for each

participant or for each item. In the by-subjects analysis, post-hoc tests revealed all three conditions differed significantly in terms of recall, including the two non-taboo conditions. Here, LGBTQ+ slurs were the likeliest to be recalled, and the non-taboo, low-arousal words were the least likely to be recalled. In the by-items analysis however, the two non-taboo conditions did not significantly differ from each other, but both were significantly less likely to be recalled than LGBTQ+ slurs. I argue this slightly misalignment in by-subjects and by-items differences raises two possibilities. First, that there may have been differences between individual participants affecting their recall likelihood, which in sections 5.5.2.6. and 5.5.2.7. I argue may be linked to personal familiarity with LGBTQ+ slurs, and to reclamation. Second, that the by-items results suggest the broader distinction between taboo and non-taboo language might affect recall likelihood more than distinctions between *specific types* of taboo and non-taboo language. This is a possibility further evidenced in Chapter 6, where results indicated that LGBTQ+ slurs did not categorically differ from other kinds of taboo language in terms of recall likelihood.

Also similar to the lexical decision results was the by-items analysis revealing a two-way interaction effect of gender and sexual identity. Unlike the lexical decision results however, the presence of this interaction effect is more unusual because there were no significant differences between groups of gender nor sexual identity when considered individually. Perhaps because of this, post-hoc tests revealed less uniform differences: the results suggested that heterosexual men were likelier to recall all words than non-heterosexual men and heterosexual women; and non-heterosexual women were likelier to recall all words than non-heterosexual men. The findings relating to heterosexual men may be related to the lexical decision results, in that the same interaction effect for response times suggested heterosexual men had the fastest (and accordingly least adverse) reactions to all words. However, this makes the apparent recall advantage for non-heterosexual women over non-heterosexual men harder to explain, given that non-heterosexual women showed the most impaired response times in the lexical decision task. I discuss some possible reasons for these results further in sections 5.5.2.3. – 5.5.2.5., where I specifically consider the (lack of) gender and sexual identity differences in the data.

Again, it is important to note that this interaction effect was only significant when results were averaged for individual items in the experiment, suggesting it is more closely attributable to differences between the specific items that were included. As with the lexical decision results, I argue that the explanation for this is likely that half of the items in the experiment were gendered, sexuality-based slurs. As such, I do not consider this interaction effect to be as generalisable as the differences between conditions, or the lack of broader gender and sexual identity differences, though I still consider it throughout the rest of this discussion.

5.5.2.2. Differences between Conditions

As indicated by my by-subjects and by-items analyses, and as clarified by the results outlined in section 5.4.2.2., results consistently showed significant differences between experimental conditions, and more typically showed that differences between all three conditions were significant. In particular, results of the more focused analyses showed that the LGBTQ+ slurs were recalled more than twice as frequently as negative high arousal words (34.60% vs 15.90% recall, respectively), which in turn were recalled slightly more frequently than negative low-arousal words (15.90% versus 11.20% recall, respectively). This contrasts with the lexical decision findings discussed in section 5.5.1.2., in which there was no difference between conditions found for response times,

suggesting that the conditions affected free recall in a way that they did not affect lexical decision in my experiment. Therefore, unlike my lexical decision findings, my free recall findings did support the first of my hypotheses outlined in section 5.2.: that the taboo words in my experiment would be more frequently recalled than either class of non-taboo words.

I argue there are two main explanations for why LGBTQ+ slurs might have been more frequently recalled across all responses than the two non-taboo conditions. The first is because the LGBTQ+ slurs represented the only examples of taboo language in the experiment, and the literature has consistently highlighted the increased recallability of taboo versus non-taboo language, as explained in section 5.1. of this chapter. I therefore consider my experiment to have replicated these findings. The importance of tabooness to recall likelihood in this experiment is further reinforced by the results presented in Table 5.9., which suggested tabooness ratings were a key predictor of recall likelihood for the LGBTQ+ slurs in particular, a finding which I discuss further in section 5.5.2.7. The second is the concept of 'semantic relatedness', as I discussed in section 2.3.2. of my literature review chapter. Research has suggested that one of the reasons taboo words are better recalled than non-taboo words is because they tend to share meanings with other taboo words, and this increased relatedness enhances their lexical retrievability (Buchanan et al. 2006; Tulving & Pearlstone, 1966). Indeed, the shared tabooness of taboo words might be considered a semantic relation in its own right (Bertels et al., 2009; Madan et al., 2017). If this is true for taboo language in general, it follows that this effect might be particularly noticeable if the taboo words in question are from a narrower semantic field, all being slurs targeting a common group, such as the LGBTQ+ community. In the next chapter, I report the results of a similar experiment I conducted to compare LGBTQ+ slurs to other taboo words which were not slurs, in order to investigate whether there would still be a difference that could be attributed to greater semantic relatedness for LGBTQ+ slurs compared to taboo words more broadly.

Given the breadth of evidence that tabooness significantly enhances recallability, I consider the first explanation to be the strongest. However, I suggest that the close semantic relatedness of the LGBTQ+ slurs may also have contributed, and argue that comparison to taboo language more broadly is needed to explore this possibility.

Regarding the significant – but less pronounced – difference between negative high-arousal and negative low-arousal words, I argue the explanation for these results is more straightforward. The literature clearly demonstrates that increased emotional arousal leads to better performance in recall tasks for many types of stimuli, including verbal stimuli (Anderson & Phelps, 2002; Kensinger & Corkin, 2003; Ochsner, 2000). Arousal was the only measure on which the negative high-arousal and negative low-arousal words differed, and the high-arousal words were more frequently recalled than the low-arousal words. Though this finding does not concern the impact of tabooness nor linguistic reclamation, I consider it to be further evidence that my experiment was able to replicate well-established findings from existing literature.

5.5.2.3. Gender and Sexual Identity Differences

Another aspect in which free recall results differed from lexical decision results was in there being no overall differences in recall likelihood based on participants' gender or sexual identity (see section 5.4.2.2.), however there was a two-way interaction effect when recall likelihood was averaged across items, which I discussed in section 5.5.2.1. Though this finding was for all words in all

conditions, I considered the lack of a general sexual identity difference in recall likelihood an early indication that the third hypothesis I made in section 5.2. may not be correct.

As I discussed in section 5.5.1.3., there is some basis in literature for a gender difference in taboo language lexical decisions, although this does not exist for sexual identity. Unfortunately, there is not such a clear precedent for either in regards to taboo language recallability, so a precise reason for the lack of demographic differences is difficult to determine, beyond the possibility that these factors simply do not affect this process.

Another possibility is that given that there was an effect of condition on recall likelihood (as I discussed in section 5.5.2.2.) which was not present in the lexical decision results, these conditions – and the reasons I give for their importance – may simply have been more influential on recallability, to the extent that any possible effects of gender or sexual identity were minimised.

This possibility is supported by the next two sets of findings I discuss, which suggest that although there was no effect of gender or sexual identity across the entire set of results, there were some differences when each demographic group was examined individually.

5.5.2.4. Gender Differences for Specific Conditions

The results I described in section 5.4.2.2. indicated that in both women and men, the LGBTQ+ slurs were recalled the most frequently, followed by the negative high-arousal words, then the negative low-arousal words. Furthermore, in both gender groups, the LGBTQ+ slurs were recalled over twice as frequently as the negative high-arousal words. As with the differences between conditions across all responses, I consider these findings to have replicated those of existing research in two key respects: that tabooness significantly enhances recallability, and that increased emotional arousal enhances recallability as well.

However, results also indicated that while differences between all three conditions were significant among women, only the difference between LGBTQ+ slurs and negative high-arousal words was significant among men. For the latter, there was no significant difference in recall likelihood between the two non-taboo conditions in the experiment.

One possible explanation for this result is that it is drive by the imbalance of women and men in the experiment (see Table 5.3.). However, I also wish to return to the possibility of a gendered difference in responses to negative high-arousal and negative low-arousal words. In sections 5.4.2.1. and 5.5.1.3., I noted that the greatest difference in lexical decision response times was for the negative high-arousal words, not for the LGBTQ+ slurs. I also pointed to findings from the literature that women's lexical decision responses to non-taboo, emotionally arousing words were significantly different to neutral words, while men's were not (Geer & Bellard, 1996). While I acknowledge that these findings concern responses to a lexical decision rather than free recall task, I argue they are suggestive of a difference in emotional arousal for women that is not observed in men, which was also observed in the recall data. As in section 5.5.1.3., I argue that this difference may be attributed to an additional social burden placed on women to use forms of language which are felt to be more polite and indirect (Brown, 1980; Coates, 1988; Fishman, 1980; Holmes, 1995), which may in turn make their responses to non-taboo emotionally arousing language stronger than men's. This might partially explain the particular by-items interaction effect which suggested that non-heterosexual women's responses were actually stronger than non-heterosexual men's, as heightened arousal responses should be associated with increased recall.

5.5.2.5. Sexual Identity Differences for Specific Conditions

In section 5.4.2.2., I also outlined the recall differences between conditions within both the non-heterosexual and heterosexual groups. In both non-heterosexual and heterosexual participants, the LGBTQ+ slurs were recalled the most frequently, followed by the negative high-arousal words, then the negative low-arousal words. Again, in both sexual identity groups, the LGBTQ+ slurs were recalled over twice as frequently as the negative high-arousal words. I consider these results to have replicated existing findings in the same manner discussed in section 5.5.2.4.

These results also indicated that for both non-heterosexual and heterosexual participants, the differences between LGBTQ+ slurs and both non-taboo conditions were significant, but the difference between the two non-taboo conditions was not. In this regard, the pattern of differences between conditions was identical for the two sexual identity groups, which is supported by the fact that there was no overall sexual identity difference in recall likelihood.

Given the effects of tabooness on recall likelihood that I have already detailed, it is not surprising that LGBTQ+ slurs – as the only examples of taboo language in this experiment – differed significantly from the non-taboo words in terms of recall likelihood regardless of participant sexual identity. Regardless of their status as slurs and the nature of the groups they target, it was expected that these would be the most frequently recalled. However, given their relationship to sexual identity, it is surprising that the degree to which LGBTQ+ slurs were more frequently recalled was almost identical when comparing non-heterosexual and heterosexual results. I consider this further evidence against my third hypothesis: that reclamation frequency would be lower in the non-heterosexual group because of reclamation modulating effects of tabooness. Instead, as discussed in section 5.5.2.3., there was no difference between sexual identity groups.

As I argued that overall, free recall results may have been more influenced by differences between conditions than by participants' gender identities and sexual identities, so too might the degree of recall likelihood for LGBTQ+ slurs have been more influenced by their tabooness than by participants' sexual identity. I discuss this possibility further in section 5.5.2.7., and explore in more detail in the experiment to follow, in which I compared LGBTQ+ slurs to other taboo words which were not slurs.

5.5.2.6. Best Model of Recall Likelihood Across Conditions

Next, I will discuss which combination of word properties measured in the experiment best predicted recall likelihood times in the free recall task, across all words and all participants. Unlike either of the lexical decision models presented in Tables 5.5. and 5.6., participant gender or sexual identity did not contribute significantly to this model. Instead, as Table 5.8. shows, the best significant model of recall likelihood included the normed familiarity and negativity ratings, and predicted that recall likelihood increased when familiarity was higher and when negativity was lower. The model was significant at the $p < .001$ level, was able to explain 4% of the variance in recall likelihood, and could correctly classify 75.70% of recall results.

As I discuss in sections 5.3.4.3. and 5.5.1.6., the issue of multicollinearity in the data meant that familiarity and personal use ratings could not be entered into the same model, nor could negativity be entered alongside arousal, offensiveness or tabooness. However, as with the case of participants' own overall reclamation ratings in the LGBTQ+ slur response time model, none of the other properties which strongly correlated with familiarity and negativity were significant when

substituted. As such, I determine that familiarity and negativity were indeed the factors affecting recall likelihood.

Arguably, these results do not replicate two key findings from existing taboo language research. In Madan et al.'s (2017) study, familiarity ratings did not contribute significantly to their model of recall likelihood, and valence – of which negativity is an aspect – was the only 'emotional' word property which did not contribute significantly to their model. The opposite was true for both findings in my results. It is also surprising that negativity had an effect on recall likelihood, when the literature has generally stressed the importance of arousal over valence for effects on memory, and my stimuli were designed in such a way as to differ by arousal but not by valence.

I argue that although it was not found to have had an effect in Madan et al.'s study, the finding that increased familiarity generally predicted increased recall likelihood nevertheless makes sense; it is logical that words which are more familiar and more widely used would be the easiest for participants to remember.

In terms of why the effect of negativity was significant in my experiment but valence was not in Madan et al.'s, I suggest the issue is related to the types of stimuli used. Madan et al.'s study included both positive and negative verbal stimuli, and their valence ratings produced values associated with both positive and negative valence. In my experiment, all words were negative, and my ratings represent degree of negativity. In this regard, I argue that my experiment measures a different aspect of valence. Whether a word is positive or negative may still not have a significant effect on recall likelihood, but when all words are negative, how negative they are perceived to be can be shown to affect recall likelihood. Considered in this way, I do not believe that my results are incompatible with those of previous research, but can add to them.

It is more difficult to determine why it was negativity in particular that added significantly to the model, while related properties like arousal and tabooness – both of which there is greater precedent for in the literature – did not. I consider two possible explanations for this. The first is that in an experiment where all stimuli were negative, degree of negativity may have been a greater distinguishing factor than if there had also been a contrast between positively- and negatively-valenced words present. If so, degree of negativity may also have been more influential than other factors like arousal and tabooness.

The second possibility relates to my experiment's specific and unique focus on slurs, as a category of word with the potential for linguistic reclamation. That slurs invite the possibility of reclamation may have had a particular effect on perceptions of valence and negativity, since reclamation is often framed as the positive use of a typically highly negative word (e.g. Bianchi, 2014). Indeed, this is how my own post-test survey framed reclamation for participants, so that it could be easily understood (see Appendix 3). There is also some support for this in the results of my LGBTQ+ slur norming study: Table 4.7. shows that the strongest inverse correlations for all reclamation behaviour ratings were with negativity, as opposed to arousal, offensiveness and tabooness. Although unanticipated, this may help explain why my results emphasised valence over the other emotional word properties that I measured. I also consider that an explanation which considers an impact of linguistic reclamation to be particularly convincing, given the findings of the model which predicted recall likelihood for just the LGBTQ+ slurs, which I now turn to discussing.

5.5.2.7. Best Model of Recall Likelihood for LGBTQ+ Slurs

As I did when modelling my lexical decision results, I created a model to predict recall likelihood for just the LGBTQ+ slurs, in order to explore if any reclamation behaviours had any effect on the words in the study which could be reclaimed. As Table 5.9. shows, recall likelihood for the LGBTQ+ slurs was best predicted by a combination of tabooness and normed self-reclamation ratings, which indicated the likelihood of a slur being used by oneself to refer positively to both oneself and other people. It was significant at the $p < .001$ level, was able to explain 6% of the variance in recall likelihood, and could correctly classify 67.70% of recall results.

Again, the presence of multicollinearity in the data meant that tabooness could not be entered into the same model alongside arousal, negativity or offensiveness, nor could self-reclamation be entered alongside familiarity, personal use or the other reclamation behaviours. However, as before, none of the other properties which strongly correlated with tabooness and self-reclamation were significant when substituted. As such, I again determine that tabooness and self-reclamation were indeed the factors affecting recall likelihood.

It is worth noting here that unlike the lexical decision model for LGBTQ+ slurs (Table 5.7), the model for their recall likelihood did not include sexual identity at all, and therefore these results do not support my third hypothesis that any reclamation effects would be stronger in the non-heterosexual group. As discussed in sections 5.5.2.3. and 5.5.2.5., this is most likely because there were no differences found between sexual identity groups, whether looking across all experimental conditions or just at LGBTQ+ slurs. The model predicted that recall likelihood increased when both tabooness and self-reclamation ratings were higher.

The finding that it was tabooness – rather than any other emotional word property – which helped predict LGBTQ+ slur recall likelihood more closely replicates findings from existing literature, which suggest that taboo words demonstrate the most exaggerated increase in recallability (Buchanan et al., 2006; Jay et al., 2008; Kensinger & Corkin, 2003; Madan et al., 2017). However, I note that this was only the case when all of the recall data entered into the model were for taboo words. It was therefore not a matter of whether a word was taboo or not, but how taboo a particular taboo word was felt to be.

Results again showed that linguistic reclamation had an impact on the LGBTQ+ slurs, this time through increased self-reclamation increasing their recall likelihood. Although this finding confirms my expectation that reclamation would modulate the recallability of LGBTQ+ slurs, it did not do so in the manner I hypothesised. I predicted that because I found reclamation ratings to negatively correlate with the arousal and tabooness judgements responsible for the increased recallability of taboo words, slur recall likelihood would decrease if their reclamation ratings increased. Instead, the opposite was true, meaning that while reclamation affected the lexical accessibility of slurs as I had predicted (see section 5.5.1.6.), this was not the case for their recall likelihood. It is also interesting to note that unlike the findings for lexical decision response times, it was not participants' own overall reclamation ratings which contributed to their recall likelihood, but the more 'general population' self-reclamation ratings provided by my norming study. This suggests that this effect was not so much determined by whether a participant in my experiment personally engaged in reclamation of an LGBTQ+ slur, but whether the reclamation of a slur was being more widely participated in beyond the lab.

Although it was not found to contribute to the recall likelihood model for just the LGBTQ+ slurs, I suggest both of these findings may be related to the importance of word familiarity in

determining recall likelihood across the whole experiment. The results I discuss in sections 5.4.4. and 5.5.2.6. demonstrate that familiarity generally increased recall likelihood, and I argue it is likely that wider personal engagement in reclamation of a slur is likely to increase familiarity as well, as it expands the range of contexts in which that slur can be used and encountered.

5.6. Conclusion

This experiment aimed to compare the lexical accessibility and recallability of LGBTQ+ slurs to two categories of non-taboo language: negative high-arousal words and negative low-arousal words. It also sought to identify which – if any – combinations of participant identity, various word properties and linguistic reclamation behaviours could best predict and explain both types of processing. I consider the experiment to have been successful in achieving both of these aims.

The experiment found that performance in the lexical decision task did not differ by experimental condition, but did differ by both participant gender and sexual identity: response times were slower among women and among non-heterosexual participants. A two-way interaction of gender and sexual identity was also present, but only when data was averaged by-items, which I suggest is explained both by the specific language used in the experiment, and by the reasons I gave for the broader differences between gender and sexual identity groups. I argued that the gender difference may reflect a particular pressure on women to avoid shocking and/or taboo language. If true, I attribute this to socialisation within the framework of a patriarchal and hegemonically masculine society, and suggest that this is a general trend rather than something which all women necessarily experience and respond to in the same way.

I observed that the sexual identity difference is harder to explain; the delayed response times among non-heterosexual participants may suggest greater aversion to slurs among the targets of such language, but is a finding less consistent with the results of my norming study and the wider literature. I suggested these inconsistencies may be related to differences between my experiment sample and the larger, more general-population sample used for my norming study, or potentially to non-heterosexual participants' responses specifically to LGBTQ+ slurs being different to their responses to other kinds of taboo language. Gender and sexual identity were also the only factors able to produce a significant model of all response times in the lexical decision task, which also demonstrated that response times to all words in the experiment could be predicted to be slower among women and non-heterosexual participants.

Regarding performance in the free recall task, the opposite was true; neither participant gender nor sexual identity individually affected recall likelihood, but experimental condition did. There was again a two-way interaction of gender and sexual identity, but only when data was averaged by-items, and the specific differences were less uniform than those for lexical decision. However, results more consistently indicated that across all participants, LGBTQ+ slurs were recalled over twice as frequently as both negative high-arousal words and negative low-arousal words. Furthermore, closer analysis revealed that in women's results, there was a difference in recall likelihood between negative high-arousal and negative low-arousal words not present when results from other demographic groups were isolated, supporting my earlier argument that socialisation drives a general trend of women responding more strongly to negative, highly arousing language. The model which best predicted recall likelihood across all words in the experiment included both word familiarity and word negativity, and indicated that increased recall likelihood was associated with increased familiarity and decreased perceptions of negativity.

Most importantly, both of the models designed to predict response times and recall likelihood for just the LGBTQ+ slurs included some measure of linguistic reclamation. Alongside participant gender and sexual identity, participants' own overall reclamation ratings were found to contribute to the model of LGBTQ+ slur response times, indicating that participants who generally reported greater reclamation of a slur also displayed less delay in their response times to the same slur. Alongside tabooness ratings, normed self-reclamation ratings – indicating the likelihood of a slur being used by oneself to refer positively to both oneself and other people – contributed to the model of LGBTQ+ slur recall likelihood, indicating that participants more frequently recalled slurs which are more widely reclaimed.

Considering all of these findings in terms of whether they support or reject the three hypotheses I outlined at the start of this chapter, my findings were mixed. My first hypothesis - that the LGBTQ+ slurs in my study would generally be associated with reduced lexical accessibility but improved recall likelihood compared to the two non-taboo conditions – was supported by the results of the free recall task, but not by those of the lexical decision task. There were no overall differences in lexical decision response times between experimental conditions, but the LGBTQ+ slurs were recalled over twice as frequently as either of the two non-taboo conditions.

Conversely, my second hypothesis - that any such effects for the LGBTQ+ slurs might be reduced by greater participation in and awareness of their reclamation – was supported by the results of the lexical decision task, but not by those of the free recall task. Some measure of reclamation contributed significantly to the processing of LGBTQ+ slurs in both tasks, but while increased reclamation reduced delay to lexical access, it improved rather than reduced recallability.

My final hypothesis - that any effect of reclamation on the processing of the LGBTQ+ slurs should be expected to be strongest among the non-heterosexual participants – was contradicted by the results of both tasks, though not in the same manner. Sexual identity did affect performance in the lexical decision task, but not as I had predicted: non-heterosexual participants generally displayed slower rather than faster response times to LGBTQ+ slurs. Interestingly, this was still the case even in the same model that predicted response times for LGBTQ+ slurs would be faster when participants' own reclamation ratings were higher, suggesting that demographic group membership should not be assumed to be indicative of reclamation behaviour, and that the two might have generally opposing effects on the processing of relevant slurs. In the free recall task, sexual identity was not found to have affected recallability at all, although reclamation still did.

Because LGBTQ+ slurs were the only examples of taboo language in this study, it was most limited in its ability to make comparisons between slurs and other kinds of taboo language. This meant that the study was not able to identify whether participants were specifically responding to them as slurs, or simply as examples of taboo language. It was also unable to compare examples of taboo language which do undergo reclamation (e.g. slurs) to those which typically do not (e.g. body parts, general insults). For this reason, my next chapter discusses a repeat of this experiment, in which I instead compared LGBTQ+ slurs to taboo words which were not slurs.

Chapter 6 – Comparing LGBTQ+ Slurs to Non-Slur Taboo Words in a Lexical Decision and Free Recall Task

6.1. Introduction

This chapter focuses on the fourth study of my research: a lab-based experiment to compare the processing of LGBTQ+ slurs to other taboo words which were not slurs, in order to determine if there was anything unique about the processing of slurs relative to other kinds of taboo language. Again, given the emphasis of my research on the reclamation of LGBTQ+ slurs, I was particularly interested in participants' responses to these words and the degree to which reclamation affected them.
I begin by revisiting some of the findings from the literature that are relevant to this experiment. I then discuss in detail my participant sample, experiment design, experiment procedure and methods of data analysis, before outlining the results. Then, I present some of the most significant findings to emerge, with particular regard to the differences found between word types and demographic groups, as well as the variables – including measures of reclamation behaviour – which were best at predicting performance in the experiment tasks.

6.2. Research Background

Because this experiment is grounded in largely the same research as the experiment discussed in Chapter 5 of this thesis, I do not revisit the literature relevant to both experiments in this section, which concerns the impacts of emotional arousal, tabooness and linguistic reclamation on taboo language processing. This information is discussed in detail throughout sections 2.2. and 2.3. of my literature review, and its relevance to this experiment is the same as outlined for the previous experiment in Chapter 5 (section 5.2.). A key study relevant to the design and analysis of both the previous chapter and this chapter (Madan et al., 2017) is also described in detail in section 5.2. of the previous chapter.

As I state in section 6.1., where this lexical decision and free recall experiment differs to the previous experiment is in all of its stimuli being examples of taboo language. Because of this, the main comparison in this version of the experiment is not based on differences in emotional arousal or tabooness, which should generally be high for all stimuli. Instead, the focus is on whether the taboo language in question is a slur or not, the effects that this has on taboo language processing, and how linguistic reclamation might affect any differences. Accordingly, in this section I specifically return to the differences between slurs and other kinds of taboo language, and why this distinction is important to explore in an experimental study like this one.

Approaches to defining slurs have tended to stress two common factors: that language constitutes a slur if it is specifically intended to derogate (Chen, 1998; Croom, 2011) and this derogation represents an exercise of systematic and hegemonic social marginalisation and oppression (Tirrell, 2012; Herbert, 2015). Indeed, it has even been suggested that just the utterance of a slur can perform an *exercitive* speech act – one which grants or denies rights to a person or group (McGowan, 2012; Herbert, 2015). In this regard, slurs meet many suggested criteria for being considered taboo language, in that they are likely to be highly emotionally arousing (Jay, Caldwell-Harris, & King, 2008) and a perceived cause of harm (Allan, 2019), and consequently they are of restricted social use (Brandes, 2019).

However, considering that research participants themselves may be targets of particular slurs, or may more strongly object to them because of their particular capacity to cause harm, I

suggest their use as examples of taboo language in experimental research presents an additional question. Namely, do participants respond to slurs in the same way that they respond to any other kind of taboo language, or do slurs produce a unique type of taboo language response? When encountering a slur, do we solely respond to its tabooness, to the fact that it is a slur, or a combination of both? This issue has not been directly addressed in existing literature, and is therefore the primary goal of the experiment described in this chapter.

Furthermore, because literature on linguistic reclamation consistently emphasises that it is a phenomenon particular to slurs and not to other kinds of taboo language (Bianchi, 2014; Brontsema, 2004; Herbert, 2015), I argue that it is also necessary to determine whether slurs are generally processed differently from other kinds of taboo language, before making definite claims about how reclamation might modulate the effects of emotional arousal and tabooness on taboo language processing. If slurs are processed in the same manner as taboo words that are not slurs, then it is likely that any effect of reclamation found is modulating arousal and/or tabooness. If slurs are processed differently, then this may still be the case, but it is also possible that reclamation is modulating perception of the word as being a slur.

Based on the available literature and the findings of my previous studies, I made the following three predictions. First, that the slurs in my study would generally perform worse in terms of lexical accessibility but better in terms of recallability, compared to taboo words that were not slurs. This is a cautious prediction, given that there is no direct evidence that the two may differ at all. Second, that any differences between slurs and other kinds of taboo language might be reduced by greater participation in and awareness of the linguistic reclamation of particular slurs, given that reclamation may lessen arousal responses and tabooness judgements for those slurs. Finally, that any effects of reclamation will be strongest in the demographic group(s) likelier to engage in it, particularly LGBTQ+ participants given the focus of this experiment.

As in the previous experiment, I tested my predictions using a lab experiment, in which lexical accessibility was measured via a lexical decision task, and recallability via a free recall task. I again used all of the normed word property and reclamation behaviour ratings collected and outlined in Chapter 4 to explore how these affected performance in these tasks, and again used a new survey to collect reclamation ratings from the lab participants themselves. This survey also collected demographic information about my participants, to determine if and how these broader factors might have affected performance in the tasks. I now outline the methodological considerations for each stage of designing and running the experiment.

6.3. Methodology

6.3.1. Sample

Forty-four participants took part in the experiment. Data from 2 participants were excluded, both for providing incomplete responses. As such, a final total of 42 participants were entered into the analyses.

Of these, 11 were recruited through a combination of posts on social media (Facebook and Twitter) and research posters placed around the University of Nottingham campus and Nottingham-based community centres, with permission. Participants recruited via these methods were able to take part in a raffle for a £10 shopping voucher as a participation incentive, though this was optional. The remaining 31 participants were undergraduate students at the University of Nottingham, who

were recruited in exchange for course credit. These participants were not eligible to take part in the voucher raffle.

All participants were aged between 16 and 29 years old and were all native or near-native speakers of British English. This age range was selected to control for effects of participant age, while also acknowledging for the differences in LGBTQ+ slur ratings between 16-29 year olds and 30+ year olds that were identified from my norming survey in Chapter 4. I collected additional respondent demographic data in order to enable demographic comparisons (see section 6.3.2.3. for details). Demographic data is presented in Table 6.1. There was a mean participant age of 19.88 (SD: 3.06), the majority of participants were women, all participants were cisgender and a slight majority of participants were heterosexual rather than non-heterosexual. These categories were later refined for the purposes of statistical analysis, as discussed in section 6.3.4. and illustrated in Table 6.3.

Table 6.1. – Overall Participant Demographic Data

Age	***N***	**%**	**Gender**	***N***	**%**	**Cisgender?**	***N***	**%**	**Sexual Identity**	***N***	**%**
18	21	50.00	Woman	30	71.43	Yes	42	100.00	Straight/Heterosexual	23	54.76
19	10	23.81	Man	11	26.19	No	0	0.00	Gay/Lesbian/Homosexual	9	21.43
20	1	2.38	Non-Binary	1	2.38	Total	42	100.00	Bisexual/Pansexual/Polysexual	8	19.05
21	2	4.76	Total	42	100.00				Queer	1	2.38
22	2	4.76							Asexual	1	2.38
25	1	2.38							Total	42	100.00
26	1	2.38									
27	3	7.14									
29	1	2.38									
Total	42	100.00									

6.3.2. Materials

As in Chapter 5, the experiment involved three main stages: a lexical decision task, a free recall task and a survey task to collect participant demographic information and reclamation behaviour ratings. In this section, I outline the materials involved in each stage of the experiment, although to avoid significant repetition I focus mainly on the differences between these materials and those described in the previous chapter (see section 5.3.2.).

6.3.2.1. Lexical Decision Task

For the lexical decision task, stimuli fell into two experimental conditions, based on word type: 25 LGBTQ+ slurs and 25 taboo words that were not slurs. Where I later refer to experimental 'conditions' as an independent variable, I am referring to these 2 word type categories. All stimuli were taboo, to ensure that the main manipulation in the experiment was whether a word was a slur. For the same reason, stimuli were balanced across conditions in terms of character length and number of syllables. All stimuli ranged from 3 to 7 characters in length, and contained a maximum of 2 syllables. The task also included 50 non-words as filler items.

Besides the difference in the experimental conditions, the process for selecting the LGBTQ+ slurs and the words these were being compared to was exactly the same as detailed in section 5.3.2.1., as was the process for producing the non-words. Table 6.2. provides examples of the stimuli used in the experiment, divided by condition. The full list of experimental stimuli can be found in Appendix 6.

Table 6.2. – Examples of Lexical Decision Task Stimuli

LGBTQ+ Slurs	Non-Slur Taboo Words	Non-Words
Bender	Asshole	Bastick
Dyke	Dick	Cancy
Faggot	Dildo	Gabtord
Lesbo	Hooker	Lexpo
Queer	Pussy	Perty
Tranny	Twat	Telker

6.3.2.2. Free Recall Task

As in the previous experiment (see sections 5.3.2.2. and 5.3.3. for details), the goal of the free recall task was to recall as many real words from the lexical decision task as possible within a set time frame. Therefore, the materials for participants to recall were identical to those used in the lexical decision task (see section 6.3.2.1.), excluding the non-words.

6.3.2.3. Survey Task

The surveys used for the survey task were identical to those used in the experiment described in Chapter 5. For the details of their design, please refer to section 5.3.2.3.

6.3.3. Procedure

Besides the number of stimuli included in the experimental stage of the lexical decision task (100 as opposed to 98), the procedure for this experiment was identical to that of the experiment described in Chapter 5. For all details of that procedure, please refer to section 5.3.3.

6.3.4. Data Analysis

The data analysis procedures used to identify the results of the lexical decision and free recall tasks were largely identical to those used in the previous experiment. As with my discussion of the materials for this chapter, my discussion in this section focuses on the differences between this experiment and the one described in the previous chapter, in order to avoid substantial repetition. For full details of all the statistical procedures used in the previous experiment, please refer to section 5.3.4.

Prior to any analysis, I visually inspected the data for outliers, such as abnormally low or high response times for single participants, or high amounts of incorrect responses. Two participants consistently provided response times > 2000ms, which were considerably slower than all other participants, so their responses were removed from all analyses. For the free recall results, I also ran a validity check to ensure that participants were recalling words that had actually been included in the lexical decision task. This was especially important given that the LGBTQ+ slurs belonged to a common semantic field, and some groups of non-slur taboo words also belonged to shared semantic fields; I needed to be confident that participants were not simply reporting words which they thought could have been included in the task. I found that 94.16% of all words recalled were those which had been included in the lexical decision task, so was satisfied that participants were accurately and appropriately completing the recall task. It is interesting that this was slightly higher than in the previous experiment comparing LGBTQ+ slurs to non-taboo words; the validity of recall responses improving slightly when all real words were taboo was an early indication that taboo

words are more recallable, as the literature suggests (Jay et al. 2008; Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Madan et al. 2017).

The main difference between the two studies was that the present study had a 2 x 2 x 2 structure (as opposed to 3 x 2 x 2), with the independent variables being as follows: the two experimental conditions (LGBTQ+ slurs vs. taboo words that were not slurs), gender (women vs. men), and sexual identity (non-heterosexual vs. heterosexual). Accordingly, this affected the number of groups entered into the by-subjects and by-items repeated-measures ANOVAs, and meant that any analyses focusing specifically on differences in participant responses by experimental condition were conducted using independent-samples t-tests (more suitable for an independent variable with two groups) rather than a one-way ANOVA (more suitable for an independent variable with three or more groups).

As in the previous experiment, I recategorized the participant demographic data presented in Table 6.1. to ensure that groups were more balanced, for the same reasons described in section 5.3.4. This produced the two groups for gender and sexual identity each, as detailed above when describing the 2 x 2 x 2 structure. Table 6.3. shows the new distributions of participants in the new groups.

Table 6.3. – Revised Gender and Sexual Identity Group Distributions

Gender	***N***	**%**	**Sexual Identity**	***N***	**%**
Woman	30	73.17	Non-Heterosexual	18	43.90
Man	11	26.83	Heterosexual	23	56.10
Total	41	100.00	Total	41	100.00

For my hierarchical multiple regression models (for lexical decision results) and binomial logistic regression models (for free recall results), all assumption checks described in sections 5.3.4.3. and 5.3.4.4. were again ran, and most were identically satisfied. Some issues with multicollinearity were again identified, but were addressed in the same manner described in the previous experiment (see sections 5.3.4.3. and 5.3.4.4.). The only assumption checks that differed slightly in this experiment were those of the binomial logistic regressions, which produced the free recall models later described in section 6.4.4. As in the previous experiment, I first checked the linearity of any continuous independent variables in the model – with respect to the logit of the dichotomous dependent variable – using the Box-Tidwell (1962) procedure. I applied Bonferroni corrections appropriate to the number of terms in both models, which in this experiment meant significance being accepted when $p < .017$ and $p < .007$ respectively (Tabachnick & Fidell, 2014). In the first model, all continuous independent variables in the final model were linearly related to the logit of the dependent variable, $p > .017$. In the second model, all continuous independent variables except for normed tabooness ratings were linearly related to the logit of the dependent variable, $p > .007$. To correct for non-linearity in the tabooness ratings, I applied an inverse cube root transformation (Menard, 2010; Ryan, 1997; Thompson, Xie, & White, 2003), after which the transformed tabooness ratings were found to be linearly related to the logit of the dependent variable, $p > .007$. Casewise diagnostics for both models revealed that there were no cases with standardised residuals greater than ±3, so no cases needed to be considered outliers and removed (Berry, 1993; Cook & Weisberg, 1982; Fox, 1991).

6.4. Results

6.4.1. Results of By-Subjects and By-Items Analyses

For the by-subjects analysis, Table 6.4. provides the results for both response time and recall likelihood in the 2 x 2 x 2 (condition x gender x sexual identity) structure. Table 6.5. provides the same for the by-items analysis. In both tables, mean recall likelihood values represent a percentage, where a value of 1.00 would indicate a 100% likelihood of recall within a particular group of an independent variable. Significant differences are indicated in both tables.

Table 6.4. – By-Subjects Analysis of Response Time and Recall Accuracy

	Group	Mean RT	SD	Mean Recall Likelihood	SD
Condition	LGBTQ+ Slurs	889.83	202.25	0.30	.46
	Non-Slur Taboo Words	896.01	210.36	0.29	.46
Gender	Women	880.81[a]	202.08	0.31	.46
	Men	918.97[a]	222.71	0.27	.44
Sexual Identity	Non-Heterosexual	859.66[b]	220.76	0.29	.46
	Heterosexual	921.69[b]	188.80	0.30	.46

[a] Difference was significant, $p = .001$.

[b] Difference was significant, $p < .001$.

Table 6.5. – By-Items Analysis of Response Time and Recall Accuracy

	Group	Mean RT	SD	Mean Recall Likelihood	SD
Condition	LGBTQ+ Slurs	901.44	20.75	0.31	0.20
	Non-Slur Taboo Words	903.59	13.39	0.30	0.18
Gender	Women	869.03[a]	23.46	0.32	0.19
	Men	936.00[a]	28.65	0.30	0.23
Sexual Identity	Non-Heterosexual	859.68[b]	27.61	0.29	0.19
	Heterosexual	945.35[b]	30.15	0.33	0.21

[a] Difference was significant, $p < .001$.

[b] Difference was significant, $p < .001$.

For the by-subjects analysis, interactions analyses revealed a significant 2-way interaction of Gender * Sexual Identity on recall likelihood ($F(1, 49) = 5.57$, $p < .05$, partial $\eta^2 = .14$). In this case, however, post-hoc tests were not able to determine significant differences between specific groups of gender and sexual identity, which may have been due to unequal distributions for these specific groups in the sample more strongly affecting the by-subjects analysis, where data was averaged for each participant in the sample. For the by-items analyses, interactions analyses also revealed a significant 2-way interaction of Gender * Sexual Identity on recall likelihood ($F(1, 49) = .09$, $p < .001$, partial $\eta^2 = .25$). This time, post-hoc tests were able to reveal that non-heterosexual women were likelier to recall all words than non-heterosexual men (mean difference 12.20%, SD: 3.70%, $p < .01$); and non-heterosexual men were less likely to recall all words than heterosexual men (mean difference -14.10%, SD: 3.30%, $p < .001$).

6.4.2. Results of Detailed Analyses of Participant Responses

Here, I report the results of the detailed analyses of participant responses, separated by analyses of response time in the lexical decision task, and of likelihood of recall in the free recall task.

6.4.2.1. Lexical Decision Task Reponses

Response accuracy to all target words across all in the lexical decision task reached an acceptably high threshold, with an overall response accuracy of 82.34%. Accordingly, the results reported in this section are not considered to have been affected by the general accuracy of participants' lexical decisions.

I conducted an independent samples t-test to identify whether there were any significant differences in response time for each condition. The analysis found no significant differences in response times between the two conditions, $t(1836) = -.64$, $p = .52$.

I then conducted two independent-samples t-tests to check for response time differences based on both participant gender and sexual identity.

For gender, a Welch t-test was run due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances ($p < .001$). There were 30 women and 11 men. Results indicated that response times were faster among women (880.81ms, SD: 202.08ms) than among men (918.97ms, SD: 222.71ms), a significant difference of -38.16ms (95% CI, -61.16ms to -15.16ms) $t(751.16) = -3.26$, $p = .001$.

For sexual identity, a Welch t-test was run due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances ($p < .001$). There were 18 non-heterosexual participants and 23 heterosexual participants. Results indicated that response times were faster among non-heterosexual participants (859.66ms, SD: 220.76ms) than among heterosexual participants (921.69ms, SD: 188.80ms), a significant difference of -62.03ms (95% CI, -80.99ms to -43.07ms) $t(1679.07) = -6.42$, $p < .001$. The results of both t-tests are visualised in Figure 6.1.

I then examined response times for each condition, and investigated whether there were significant response time differences based on gender and sexual identity within each. This involved running four independent-samples t-tests in total: a gender and sexual identity comparison for both conditions. All were run as Welch's t-tests, due to the assumptions of homogeneity of variances being violated, as assessed by Levene's tests for equality of variances (all ps $< .05$). There were 30 women and 11 men in all gender comparisons, and 18 non-heterosexual participants and 23 heterosexual participants in all sexual identity comparisons.

For the LGBTQ+ slurs, results indicated that response times were faster among women (875.42ms, SD: 193.72ms) than among men (919.75ms, SD: 226.96ms), a significant difference of -44.33ms (95% CI, -77.44ms to -11.23ms) $t(357.41) = -2.63$, $p = .01$. Results also indicated that response times were faster among non-heterosexual participants (860.88ms, SD: 221.07ms) than among heterosexual participants (916.34ms, SD: 179.53ms), a significant difference of -55.46ms (95% CI, -82.17ms to -28.74ms) $t(813.85) = -4.07$, $p < .001$.

For the non-slur taboo words, results indicated that response times were faster among women (885.73ms, SD: 209.43ms) than among men (918.21ms, SD: 218.96ms), a significant difference of -32.48ms (95% CI, -64.57ms to -0.39ms) $t(391.33) = -1.99$, $p < .05$. Results also indicated that response times were faster among non-heterosexuals (858.44ms, SD: 220.70ms) than among heterosexuals (926.38ms, SD: 196.63ms), a significant difference of -67.94ms (95% CI, -94.83ms to -41.05ms) $t(859.45) = -4.96$, $p < .001$. The results of both sets of t-tests are visualised in Figure 6.2.

Figure 6.1. – Response Time Differences for Gender and Sexual Identity (Across Conditions)

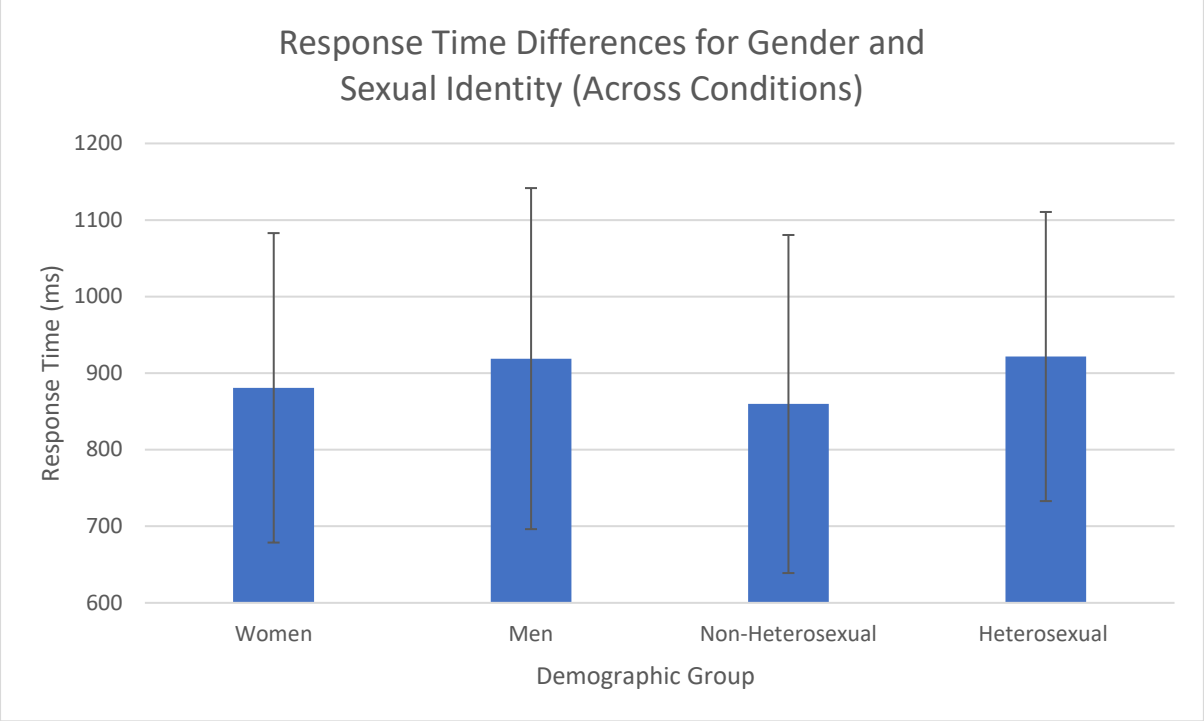
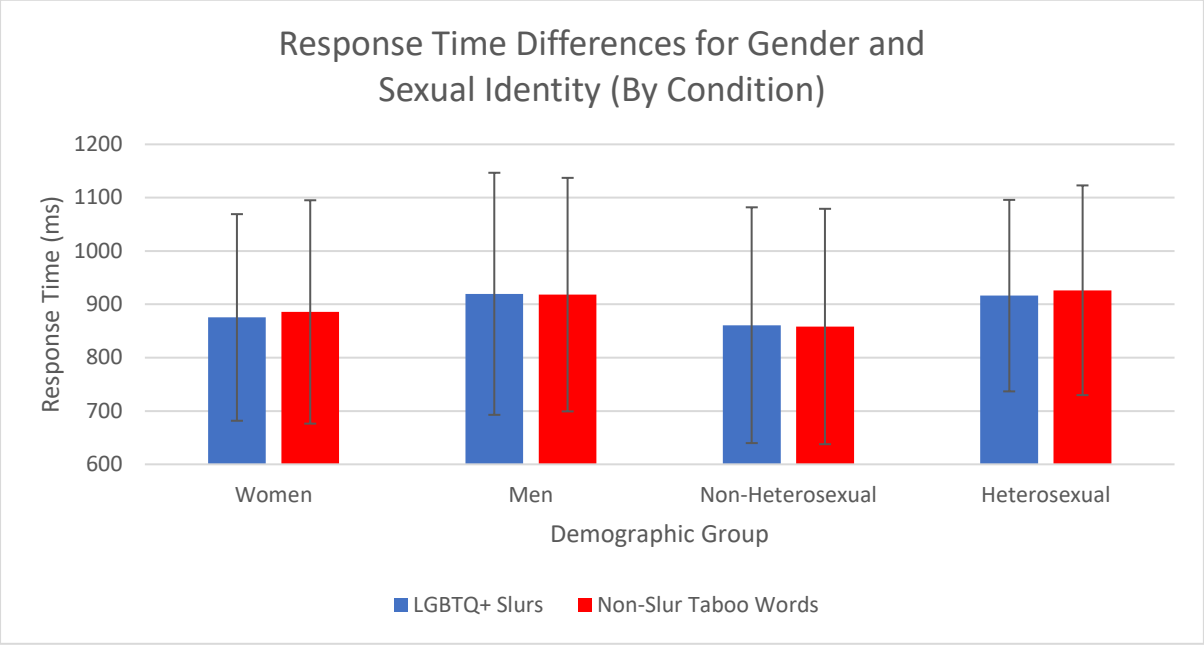


Figure 6.2. – Response Time Differences for Gender and Sexual Identity (By Condition)



Collectively, these results suggest that women consistently provided faster response times in all conditions when compared to men, and the greatest of these differences was in response times for the LGBTQ+ slurs. Non-heterosexual participants also consistently provided faster response times in all conditions when compared to heterosexual participants, and the greatest of these differences was in response times for non-slur taboo words.

6.4.2.2. Free Recall Responses

I conducted a chi-square test of homogeneity to look for differences in recall likelihood based on condition. Results indicated that 30.10% of LGBTQ+ slurs were recalled, compared to 29.00% of non-slur taboo words, a non-significant difference in proportions of .011, $p = .59$.

I then conducted two chi-square tests of homogeneity to check if there were any differences in recall likelihood based on participant gender and sexual identity groups.

Gender did not have a significant effect: women recalled 30.9% of words, while men recalled 26.7% of words ($p = .07$). Sexual identity also did not have a significant effect: non-heterosexual participants recalled 29.3% of words, while heterosexual participants recalled 29.8% of words ($p = .79$).

Finally, I conducted four further chi-square tests of homogeneity to test if there were any recall likelihood differences between conditions, within each individual participant demographic group: women, men, non-heterosexual and heterosexual.

For women, results indicated that 31.10% of LGBTQ+ slurs were recalled, compared to 30.60% of non-slur taboo words, a non-significant difference in proportions, $p = .83$.

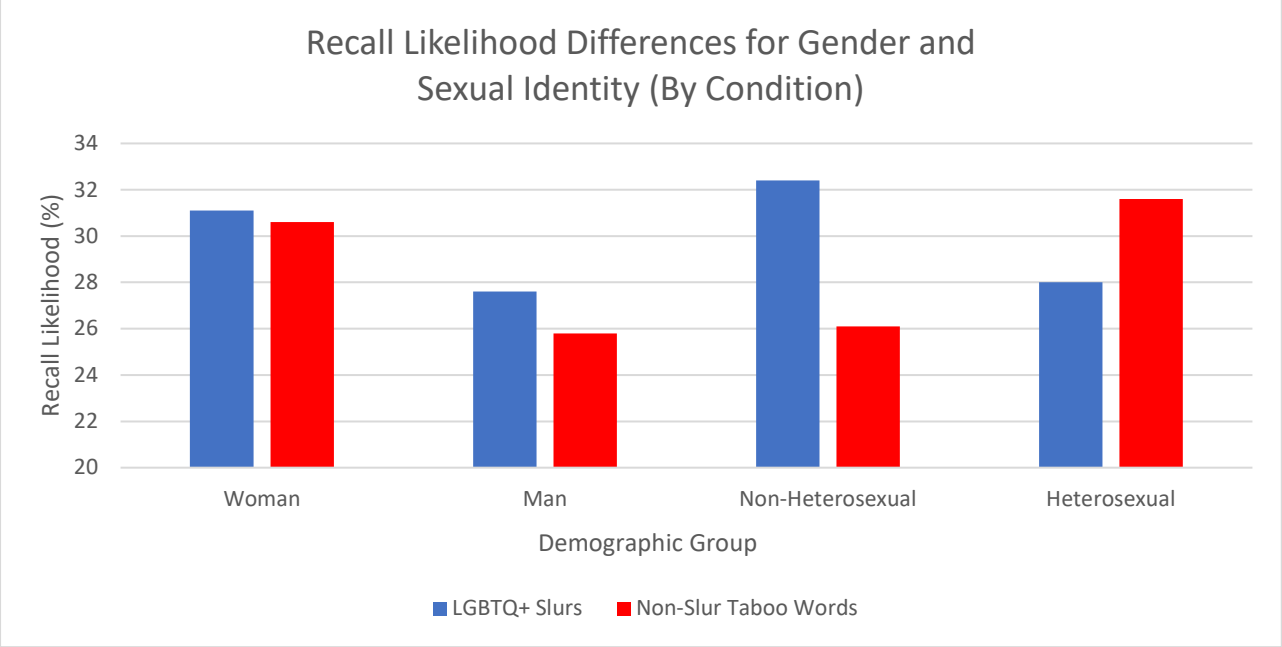
For men, results indicated the same patterns as for women: 27.60% of LGBTQ+ slurs were recalled, compared to 25.80% of non-slur taboo words, a non-significant difference in proportions, $p = .63$.

For non-heterosexual participants, results indicated that 32.40% of LGBTQ+ slurs were recalled, compared to 26.10% of non-slur taboo words, a significant difference in proportions, $p = .03$.

For heterosexual participants, results indicated an opposite but not significantly different pattern compared to non-heterosexual participants: 28.00% of LGBTQ+ slurs were recalled, compared to 31.60% of non-slur taboo words, a non-significant difference in proportions, $p = .20$. The results of all four chi-square tests of homogeneity are visualised in Figure 6.3.

Collectively, these results suggest that across condition, LGBTQ+ slurs were no likelier to be recalled than non-slur taboo words, and this was usually true regardless of gender and sexual identity, for which there were also no significant differences across conditions. However, among non-heterosexual participants exclusively, LGBTQ+ slurs were significantly likelier to be recalled than non-slur taboo words.

Figure 6.3. – Recall Likelihood Differences for Gender and Sexual Identity (By Condition)



6.4.3. Results of Modelling Lexical Decision Responses

As in the previous experiment (see Table 5.6.), I found that the best model of response times across all words in the lexical decision task included just participant gender and sexual identity. The model significantly predicted response times and explained 14% of the variance in response times, $F(2,1679) = 131.89$, $p < .001$, adj. $R^2 = .14$. Both variables added significantly to the prediction, $p < .001$. Where this model differed to that of the previous experiment was in indicating that response times to LGBTQ+ slurs response times to all words were correctly predicted to be slower when participants identified as a man compared to as a woman, and when participants identified as heterosexual compared to non-heterosexual. I provide the regression coefficients and standard errors in Table 6.6.

Table 6.6. – Model of Response Times to All Words in Lexical Decision Task

Response Times (All Words)	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	*UL*				
Model						.14	.14***
Constant	545.85***	507.38	584.33	19.62			
Gender	114.51***	97.04	131.99	8.91	.31		
Sexual Identity	107.16***	91.46	122.85	8.00	.32		

Note. Model = "Enter" method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardised coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2. *** $p \leq .001$

Again, as in the previous experiment (see Table 5.7.), the best model of response times to the LGBTQ+ slurs specifically in the lexical decision task was found to include participant gender, sexual identity and participants' own overall reclamation behaviour ratings. The model significantly predicted response times and explained 13% of the variance in response times, $F(3,785) = 41.22$, $p < .001$, adj. $R^2 = .13$. All three variables added significantly to the prediction, $p < .01$. Again, this model

differed to that of the previous experiment in indicating that response times were correctly predicted to be faster when participants identified as a woman compared to as a man and when participants identified as non-heterosexual compared to heterosexual. However, this model mirrored the previous model in also indicated that response times were faster when participants gave higher personal overall reclamation behaviour ratings for a slur. I provide the regression coefficients and standard errors in Table 6.7.

Table 6.7. – Model of Response Times to LGBTQ+ Slurs in Lexical Decision Task

Response Times (LGBTQ+ slurs)	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	*UL*				
Model						.14	.13***
Constant	597.86***	536.91	658.80	31.05			
Gender	105.64***	80.32	130.96	12.90	.29		
Sexual Identity	97.22***	74.15	120.29	11.75	.29		
Participant Overall Recl	-17.64**	-29.02	-6.27	5.80	-.10		

Note. Model = "Enter" method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardised coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2; Participant Overall Recl = mean of participant's own reclamation behaviour ratings for an LGBTQ+ slur. ** $p < .01$, *** $p < .001$.

6.4.4. Results of Modelling Free Recall Responses

The logistic regression model which best explained overall recall likelihood included just the normed familiarity word ratings. The interactions described by the model were significant, $\chi^2(1) = 22.16$, $p < .001$. A Hosmer and Lemeshow goodness of fit test indicated that the model was not a poor fit, $p > .05$. The model explained 2.0% (Nagelkerke R^2) of the variance in recall likelihood and correctly classified 70.50% of cases. Sensitivity was .00%, specificity was 100.00%, positive predictive value was 100.00% and negative predictive value was 70.45%. Increasing familiarity ratings were found to be associated with an increased likelihood of recall likelihood, as indicated in Table 6.8.

Table 6.8. – Model of Recall Likelihood for All Words in Free Recall Task

Recall likelihood (All Words)	*B*	SE	Wald	*Df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							LL	*UL*
Constant	-1.92	.23	67.30	1	< .001	.15		
Familiarity	.34	.07	21.61	1	< .001	1.40	1.21	1.61

The logistic regression model which best explained recall likelihood for just the LGBTQ+ slurs included number of characters, the normed overall reclamation ratings and the transformed tabooness ratings. The interactions described by the model were significant, $\chi2(3) = 58.92$, $p < .001$. A Hosmer and Lemeshow goodness of fit test indicated that the model was not a poor fit, $p > .05$. The model explained 8.0% (Nagelkerke R2) of the variance in recall likelihood and correctly classified 72.20% of cases. Sensitivity was 17.90%, specificity was 96.30%, positive predictive value was 32.50% and negative predictive value was 73.15%. Fewer characters, increasing normed overall reclamation ratings and increasing tabooness ratings were all found to be associated with an increased likelihood of recall likelihood, as indicated in Table 6.9.

Table 6.9. – Model of Recall Likelihood for LGBTQ+ Slurs in Free Recall Task

Recall likelihood (LGBTQ+ Slurs)	*B*	SE	Wald	*Df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							LL	*UL*
Constant	-13.50	2.26	35.61	1	< .001	<.001		
Characters	-.17	.07	6.52	1	.01	.85	.74	.96
Overall Reclamation	1.26	.20	40.80	1	< .001	3.54	2.40	5.22
Tabooness*	7.77	1.35	33.39	1	< .001	2368.18	169.78	33032.00

*An inverse cube root transformation was applied to the tabooness ratings to correct linearity.

6.5. Discussion

6.5.1. Lexical Decision Results

In this section of my discussion, I have grouped the key findings from the lexical decision task. First, I consider the results of the comparisons I made by experimental condition (LGBTQ+ slurs and taboo words which were not slurs), gender, and sexual identity. I then discuss what is indicated by the models which predicted response time variance, first to all words in the experiment, and then specifically to the LGBTQ+ slurs.

6.5.1.1. By-Subjects and By-Items Analyses

For the lexical decision responses, the results of the by-subjects and by-items analyses largely showed the same patterns: regardless of whether response times were averaged for individual participants or individual target words, there were no significant differences between the two experimental conditions, but the response times of men and heterosexual participants to all target words were slower than those of women and non-heterosexual participants. As I discuss in more detail in sections 6.5.1.3. and 6.5.1.4., this matches the previous experiment in terms of the significant differences in response time being between gender and sexual identity, but the nature of these differences are the opposite of those in the previous experiment. Neither the by-subjects nor by-items analyses revealed any significant interaction effects on lexical decision responses. At this stage, it is important to note that all significant differences that were observed were present in both types of analysis. As in the previous experiment, this suggests that these differences represent more generalisable patterns, and cannot be solely attributed to differences between the individual participants in my sample, nor between the specific items that I selected to include in my experiment. Accordingly, I argue that my subsequent discussions of these results in relation to broader patterns of social behaviour again cannot be dismissed as products of the experiment design.

6.5.1.2. Differences between Conditions

As indicated by my by-subjects and by-items analyses, and clarified by the results in section 6.4.2.1., results showed that there were no significant differences in response times to each word type condition across all participant responses. This is consistent with the lexical decision results from the previous experiment (see section 5.5.1.2.), except where those suggested no overall differences in the lexical accessibility of LGBTQ+ slurs and two classes of non-taboo words, these results suggested no overall difference in lexical accessibility between LGBTQ+ slurs and other taboo words which are not slurs. Accordingly, my lexical decision results do not support my first hypothesis as set out in

section 6.2.: that response times to slurs would be even more delayed than they are to other kinds of taboo language.

This is not a question which has yet been explored in the literature, so the finding that slurs do not appear to be processed differently to other kinds of taboo words - in spite of the particularly negative contexts in which they are typically used – is a novel one. I argue that the explanation for this is straightforward: slurs are already widely recognised as a type of taboo language, and reproduce the same tabooness effects seen in other examples of taboo language.

More importantly, I argue that this finding helps to clarify some of the findings from my previous lexical decision experiment. In section 5.5.1.2., I raised the possibility that the reason there may have been no differences found between the taboo and non-taboo words in my previous study was because all of the taboo words in that study were LGBTQ+ slurs. I suggested that slurs may either be processed differently to other kinds of taboo language, or that responses to slurs may be more affected by individuals' experiences of and relationship to them. I argue that this finding suggests that the first explanation is unlikely, as the slurs in my experiment did not differ in terms of their lexical accessibility. Instead, I argue this finding adds weight to the second explanation. In the previous experiment, I suggested that this explanation was reinforced by the findings that I did find response time differences based on participant gender and sexual identity (see sections 5.5.1.3. and 5.5.1.4.). The same is true for this experiment; I again found gender and sexual identity differences in response times, between the LGBTQ+ slurs and the taboo words which were not slurs.

6.5.1.3. Gender Differences

Though the results I describe in sections 6.4.1. and 6.4.2.1. indicated that there was a significant difference in response time between women and men, this difference contrasted the gender difference found in the previous experiment (see section 5.5.1.3. for discussion). Where in the previous experiment, women's overall response times had been found to be slower than men's, in this experiment they were found to have been faster. This finding was consistent when response times for each condition were isolated: women provided faster response times to both the LGBTQ+ slurs and non-slur taboo words. The greatest difference in response times between women and men was found for the LGBTQ+ slurs. I consider there to be several possible explanations for this finding.

The first is that this difference may again have been a consequence of the sample. As in the previous experiment, the sample was less balanced for gender than it was for sexual identity; as Table 6.3. indicates, there were nearly three women for every man in the experiment, and as such these results should be considered more cautiously. Nevertheless, the overall difference reached a high level of significance, $p \leq .001$. The same was true within both conditions, though not to as high a level, with women's responses times being significantly faster for both the LGBTQ+ slurs ($p \leq .01$) and the non-slur taboo words ($p < .05$).

Another possibility is the difference in stimuli between the two experiments. In the previous experiment, taboo stimuli were being compared to non-taboo stimuli, with the finding that women's responses to all stimuli were slower. Geer & Bellard (1996) – one of the few studies which has looked at gender differences in lexical decision performance – made the same comparison, and also found women's response times to be slower, particularly for words with taboo semantic content. However, in this experiment, all of the stimuli were taboo. Potentially, women's response times might only be slower when the point of comparison is between taboo and non-taboo words. In explaining why women's response times were slower in the previous experiment, I pointed to

sociolinguistic research suggesting a patriarchal pressure on women to avoid language seen as impolite or socially unacceptable (Brown, 1980; Coates, 1988; Fishman, 1980; Holmes, 1995). I suggest it is possible that this pressure may be more strongly felt when non-taboo words are also present in the experiment, as a contrast to the taboo words.

However, this does not explain why women's response times were in fact faster than men's, as opposed to there being no gender difference in results. Because of this, I consider another possibility. Throughout my thesis, I have noted that – despite being identified via responses to a free-response survey – the LGBTQ+ slurs in my studies lean towards targeting non-heterosexual men, be it based on sex acts they are felt to participate in or gender expressions they are felt to display (see sections 3.5.2. and 4.4.2.2. for discussions). Referring to Table 6.2., this was also the case for the stimuli in this experiment. Arguably, the same table also suggests this was true of the non-slur taboo words, more of which referred to body parts and insults typically associated with (cisgender) men. Potentially, women's response times were faster overall because women were less often the subject or target of all of the taboo words in this experiment.

This does raise the question of why the same effect was not observed in the previous experiment, given that the same LGBTQ+ slurs were used, but I suggest two possible explanations for this. First, neither category of non-taboo words in the previous experiment – which the LGBTQ+ slurs were being compared to – contained words likely to be perceived as gendered (see table 5.2. for details). There were therefore fewer total examples of gendered language in the previous experiment, compared to this one. Second, I consider that the presence of both taboo and non-taboo language may have affected response times in the previous experiment, more than the gendering of the stimuli did. Non-taboo stimuli were not present in this experiment, so the gendering of stimuli may have become more relevant to responses.

6.5.1.4. Sexual Identity Differences

As indicated by my by-subjects and by-items analyses, and clarified by the results in section 6.4.2.1., results indicated a significant sexual identity difference in response times to all words in the experiment: non-heterosexual participants' response times were found to be faster than those of heterosexual participants. As with the gender differences, this finding remained consistent when response times for each condition were isolated: non-heterosexual participants provided faster response times to both LGBTQ+ slurs and non-slur taboo words. Unlike the gender differences, these sexual identity differences reached a very high level of significance in all cases, $p < .001$.

The second hypothesis I set out in section 6.2. predicted that increasing reclamation of LGBTQ+ slurs would reduce any impact of arousal and tabooness on response times, causing them to be less delayed. For the same reasons, my third hypothesis predicted that any such effect of reclamation on response times would be strongest in the demographic group(s) likelier to engage in it, particularly the LGBTQ+ participants. While these results alone could not establish that the difference found between sexual identity groups was related to an effect of reclamation, there was nevertheless a significant difference between the two sexual identity groups which had manifested as I predicted. I therefore considered this early support for both hypotheses.

As indicated in Table 6.3., the experiment sample was much more balanced for sexual identity than it was for gender, so these results did not share the same concern. Instead, I consider other explanations concerning the stimuli themselves.

As was the case for the previous experiment, there is no basis in the literature to explain an overall difference in lexical decision performance based on sexual identity, compared to the small precedent that exists for gender differences. As I suggested in section 5.5.1.4., I argue that the likeliest explanation is that half of the words in the experiment specifically and very negatively describe members of the LGBTQ+ community. If the presence of these words had a different effect on non-heterosexual and heterosexual participants, then the inclusion of enough of them may have been sufficient to affect the mean of all response times in both demographic groups.

In making this argument, I must address an apparent contradiction. In section 6.5.1.3., I noted that many of the words in both experimental conditions described or were associated with men, and suggested that women may therefore have been less averse to encountering them, explaining their faster response times. Here, I observe that one of my experimental conditions entirely included words which explicit target non-heterosexual people, yet suggest that this may also explain the faster – not slower – response times found in this group. This may suggest there are separate reasons for one's aversion to taboo language being reduced, both by being or not being the target of that language.

To explain why being the target of the LGBTQ+ slurs could have reduced aversion to these words among non-heterosexual participants, I first consider the finding that response times to just the LGBTQ+ slurs were still faster in the non-heterosexual group. This finding appears consistent with the literature and with the results of my norming study. As I discussed in sections 4.4.2.3. and 4.5.3., non-heterosexual participants almost always provided higher familiarity, personal use and reclamation behaviour ratings, but lower arousal, negativity, offensiveness, tabooness and age of acquisition ratings. Based on studies such as Madan et al.'s (2017), the lower ratings for 'non-emotional' word properties and higher tabooness ratings provided by heterosexual participants for LGBTQ+ slurs may partially account for the slower response times to such words among this group. However, this is not the only explanation for this difference, and further explanation is also required for heterosexual participants' slower response times to the non-slur taboo words.

Given that the slurs in the experiment were all ones which targeted LGBTQ+ individuals, I believe another reason that sexual identity would affect response times to these words is that sexual identity is more often the determining factor in whom they target. Interestingly, though, it was heterosexual participants – as the group generally not targeted by this language – whose response times suggested the most aversion to LGBTQ+ slurs. Importantly for the central question of my research, this also means that non-heterosexual participants – the group able to engage in linguistic reclamation of these slurs – was not the one with the slowest response times, something which I discuss further in section 6.5.1.6.

This suggests a possible ingroup/outgroup effect on response time, in which members of a dominant outgroup are actually the ones who display the most delayed responses. In explaining some of the sexual identity differences in my word ratings for LGBTQ+ slurs (see sections 4.4.2.3. and 4.5.5.), I raised the key issue of ingroup/outgroup membership, and suggested that membership of an LGBTQ+ ingroup might be associated with increased exposure to LGBTQ+ slurs, and critically to non-derogatory uses of such language. If this is true, then I believe this might also help to explain why response times to LGBTQ+ slurs were slower among the heterosexual participants, who are generally not part of such an ingroup (unless they are also not cisgender, which the vast majority of my participants were).

At this stage, it is worth considering why this sexual identity difference aligned much better with the results of my norming study and with the existing literature than that found in the previous

study (see section 5.5.1.4.). In my discussion of those results, I suggested that the manner in which the LGBTQ+ slurs affected non-heterosexual participants may have had more impact than any of the word properties myself and other taboo language researchers have measured. To support this, I pointed also to the fact that almost none of these word properties contributed to the models of lexical decision response times for that experiment, which was also true in this one (see Tables 6.4. and 6.5.). However, if this were true, I argue the same difference in response times should have been found in this experiment, rather than the opposite one.

As another possibility, I suggested that any effect of LGBTQ+ slurs on non-heterosexual participants may have been enhanced in the previous experiment by the contrast of the LGBTQ+ slurs with the entirely non-taboo words in the study. I now consider this explanation more likely, given that such a contrast was not present in this experiment, and a different set of results was found.

6.5.1.5. Best Model of Lexical Decision Response Times Across Conditions

I now turn to the combination of word properties measured in the experiment which best predicted response times in the lexical decision task, across all words and all participants. As in the previous experiment (see Table 5.6., section 5.5.1.5.), the best significant model of response times included just participant gender and sexual identity, and predicted that response times would be faster among women and faster among non-heterosexual participants. This finding is consistent with the results I discussed in sections 6.5.1.3. and 6.5.1.4. As with the sexual identity difference in response times, I also considered this model to provide early support for my second and third hypotheses, though again these results could not determine an effect of reclamation on LGBTQ+ slurs; the model was for response across all conditions, including those which were not slurs and could not undergo reclamation.

Unlike in the previous experiment, in which the overall response time model was only able to account for 1% of variation, gender and sexual identity were found to account for 14% of the variation of all response times in this experiment. Additionally, the model was significant at the $p < .001$ level. This finding suggests that gender and sexual identity were better able to explain variations in response time when LGBTQ+ slurs were being compared to other taboo words, as opposed to non-taboo words. I argue that this support my developing argument that these demographic factors become more relevant when there is not a distinction between taboo and non-taboo language also affecting response times, as discussed in sections 6.5.1.3. and 6.5.1.4.

Again, this model does not replicate the findings from existing studies such as Madan et al.'s (2017), in which lexical decision times were best predicted by 'non-emotional' factors (e.g. familiarity, personal use, word frequency, number of letters and number of syllables) and tabooness. In my experiment, none of these added significantly to the model, despite all having been measured. Instead, only participant gender and sexual identity added significantly to the model. As in the previous experiment, this contrast could have been caused by the specific focus of my experiments on LGBTQ+ slurs as a class of taboo stimulus. Non-emotional word properties may not be as effective as gender and sexual identity at predicting variation in response times when this is the case, as these demographic factors are directly related to the meaning of a significant number of the words being tested. Tabooness may also not have been a predicting variable in this experiment because all the stimuli included were taboo, so there was less variability in tabooness ratings.

As I argued when discussing differences in responses times between groups based on gender and sexual identity, I do not believe that these variables were able to predict response times because they directly affected participants' ability to make fast and accurate lexical decision judgements. Instead, I consider the model to be further evidence in support of the explanations I provided for these differences in sections 6.5.1.3. and 6.5.1.4.: that more aversion and sensitivity to slurs compared to other taboo words may actually be demonstrated by those not part of the targeted group.

6.5.1.6. Best Model of Lexical Decision Response Times for LGBTQ+ Slurs

In order to explore if reclamation behaviour had any effect on performance in the lexical decision task for the LGBTQ+ slurs, I created a model to predict just the LGBTQ+ slur response times. As Tables 5.5. and 6.5. show, these results were the same in this experiment as they were in the previous experiment. Gender and sexual identity again predicted response times, though not in the same manner as the previous experiment, as discussed in section 6.5.1.5. As before, participants' own overall reclamation ratings – the mean of each reclamation behaviour rating provided by each participant during the survey portion of the experiment – further improved the model. It revealed that increasing participation in and awareness of reclamation of a slur could predict a decrease in response time during the lexical decision task (see Table 6.7.). I consider these findings the confirmation of my second and third hypotheses, for which the previously discussed findings were also providing initial support.

This model was significant at the $p < .001$ level, and could account for 13% of the variance in response times to LGBTQ+ slurs. As with the overall lexical decision results discussed in section 6.5.1.5., the model had noticeably higher predictive power in this experiment than it did in the previous one, in which it was only able to account for 2% of the variation in response times. I consider this further evidence that these factors become more relevant when there is not a distinction between taboo and non-taboo language also affecting response times.

As raised in section 6.3.4., multicollinearity between word property ratings was an issue in the data, and so I could not insert participants' own overall reclamation behaviours alongside the normed familiarity, personal use and reclamation behaviour ratings, nor each individual reclamation behaviour rating given by each participant. This meant that it was not immediately clear whether it was participants' overall personal experience of reclamation that was reducing response times, or if this effect was being caused by a related variable, such as increasing familiarity. However, I attempted to insert all such variables into the model in place of participants' own overall reclamation ratings, and found that none added significantly to the model when substituted in.

This finding is consistent with those from my norming study, in which increasing overall reclamation ratings were associated with decreasing tabooness ratings (see Table 4.6.), which existing research in turn associates with faster response times (Geer & Bellard, 1996; Madan et al. 2017; Thomas & LaBar, 2005; Williams & Evans, 1980). It is also important to note that it was the combination of every reclamation behaviour I measured – personal reclamations and awareness of other people's reclamations, in both self-reference and reference to others – that improved the model, rather than a particular one of these. As I did when explaining the importance of this model in the previous experiment, I argue this finding indicates that response times are more likely to be reduced when reclamation is experienced and engaged with more often in multiple ways.

Finally, I reiterate that this finding supports my decision to measure participants' individual engagement with reclamation behaviours, despite having 'general population' measures of this from my norming study, for the same reasons I discuss in section 5.5.1.6. That the model was improved by a combination of participants' own reclamation behaviour ratings – rather than any of the normed reclamation behaviour ratings – justifies having taken this approach.

6.5.2. Free Recall Results

I now proceed to discuss the key findings from the portion of my experiment involving the free recall task. As with discussion of the lexical decision results, I begin by consider the results of the comparisons I made by experimental condition, gender, and sexual identity. I then discuss what is indicated by the models which predicted recall likelihood, first to all words in the experiment, and then specifically to the LGBTQ+ slurs.

6.5.2.1. By-Subjects and By-Items Analyses

As indicated in Tables 6.4. and 6.5., neither the by-subjects nor by-items analyses found any significant differences in recall likelihood, in terms of any of the independent variables on their own. I discuss this lack of differences between groups of any one of the independent variables in sections 6.5.2.2. and 6.5.2.3. However, further analysis did reveal a difference between experimental conditions among just the non-heterosexual participants, which I discuss further in section 6.5.2.4.

Interestingly, both by-subjects and by-items interaction analyses revealed a significant two-way interaction of gender and sexual identity on recall likelihood. Post-hoc tests for the by-subjects analysis were not able to identify where combined gender and sexual identity groups differed significantly, which I attributed to possible issues with the distributions of these groups within the sample. However, post-hoc tests for the by-items analysis revealed that non-heterosexual men were less likely to recall all words than both non-heterosexual women, and heterosexual men. Because this interaction effect was present in both by-subjects and by-items analysis, this suggests a more generalisable combined effect of gender and sexual identity on recall likelihood, with non-heterosexual men perhaps exhibiting a particular disadvantage when recalling both LGBTQ+ slurs and non-slur taboo words. However, because the specific nature of this interaction effect could only be established for one type of analysis suggests this finding may need further testing, with a sample specifically designed to explore this interaction further. For this reason, I do not discuss this interaction in further detail in this discussion, but I do consider it in relation to the specific finding among non-heterosexual participants discussed in section 6.5.2.4.

6.5.2.2. Differences between Conditions

As indicated by my by-subjects and by-items analyses, and clarified by the results in section 6.4.2.2., results showed that the LGBTQ+ slurs were recalled in 30.10% of cases, compared to 29.00% of non-slur taboo words, a non-significant difference, $p = .59$. This result complements the lack of response time differences between the two conditions in the lexical decision task (see section 6.5.1.2.), but contrasts the free recall findings from the previous experiment, in which LGBTQ+ slurs, negative high-arousal and negative low-arousal words all differed significantly in terms of recallability (see section 5.5.2.2). As with the results of my lexical decision task, then, the free recall results do not support my first hypothesis as set out in section 6.2.: that recalling slurs would be even likelier than recalling other kinds of taboo language.

I argue that these results are not incompatible with the findings of the previous experiment; rather, they suggest that while LGBTQ+ slurs are more recallable than non-taboo language, they are no more or less frequently recalled than other kinds of taboo language. In this regard, they suggest the same conclusion as the lexical decision results in this experiment: in general, slurs are not processed differently to other kinds of taboo language.

This finding clarifies some of the results from my previous experiment. In section 5.5.2.1., I suggested that the close semantic relatedness of the LGBTQ+ slurs might have contributed to their increased recallability in addition to their tabooness, but noted that comparison would have to be made to other taboo language which was less semantically related to test this. I argue this experiment was able to make this comparison, and by finding no difference, suggests that semantic relatedness does not affect recall likelihood for taboo words. Therefore, tabooness by itself may provide the best explanation for the enhanced recallability of LGBTQ+ slurs compared to the non-taboo words in the previous experiment, a finding which would be supported by the literature (Buchanan, Etzel, Adolphs, & Tranel, 2006; Jay et al. 2008; Kensinger & Corkin, 2003; Madan et al., 2017).

6.5.2.3. Gender and Sexual Identity Differences

One aspect in which free recall results for this experiment did align with those of the previous experiment was in finding no overall differences in recall likelihood when gender and sexual identity groups were compared (see sections 6.4.1. and 6.4.2.2.). Although these results were across both conditions, I considered them to provide some evidence against my third hypothesis, in which I predicted that slurs would be less likely to be recalled among LGBTQ+ participants. As I discussed in sections 5.5.1.3. and 5.5.2.3., while there is some basis in literature for a gender (but not sexual) identity difference in taboo language lexical decisions (Geer & Bellard, 1996), there is no such precedent for either regarding taboo language recallability. It is therefore difficult to explain this lack of difference, beyond the possibility that neither gender nor sexual identity generally affect this process.

In discussing the lack of gender and sexual identity differences for recall likelihood in the previous experiment, I suggested that because there was an effect of experiment condition in that experiment (see section 5.5.2.2. for discussion), this may have overridden any possible effects of gender or sexual identity. This is not a possibility for these results, as there was no effect of condition on recallability found in this experiment. Furthermore, this suggests that condition effects overriding gender and sexual identity effects is a less likely explanation than there simply being no overall effect of gender or sexual identity, since there were still no effects of either demographic factor when no differences between conditions were identified.

Despite this overall finding, results did indicate recall differences relating to sexual identity within specific experimental conditions, which I will now discuss.

6.5.2.4. Sexual Identity Differences for Specific Conditions

As part of the group difference analyses I conducted, I investigated whether there were any differences between conditions within individual participant demographic groups: women, men, non-heterosexual participants and heterosexual participants. As I outline in Section 6.4.2.2., there was a significant difference in the recall likelihood of LGBTQ+ slurs versus non-slur taboo words, but only among the non-heterosexual group. Among non-heterosexual participants, LGBTQ+ slurs were

recalled in 32.40% of cases, compared to 26.10% of non-slur taboo words, which was significant at the $p < .05$ level.

This result indicates that while LGBTQ+ slurs may not have differed from other kinds of taboo words across all participants, they did within the one group of participants most consistently targeted by this language. As with the response time differences between non-heterosexual and heterosexual participants (see section 6.5.1.4.), I suggest this result indicates an ingroup/outgroup effect, in which slurs may be processed differently to other kinds of taboo language, but only among the group which they target. Interestingly, this suggests that my first hypothesis – that the LGBTQ+ slurs might perform better in the recall task than the non-slur taboo words – was correct, but only in a specific demographic group and not overall. However, I also consider it evidence against my third hypothesis – that increasing engagement in reclamation might make LGBTQ+ slurs less memorable – since among the non-heterosexual participants, recall was actually likelier for these words.

However, I also note that this result appears to partially contradict some aspects of the interaction effect for gender and sexuality discussed in section 6.5.2.1. Post-hoc tests on the by-items interaction effect found that non-heterosexual men had lower recall likelihood for all words than heterosexual men, but this does not align with the finding that non-heterosexual participants collectively seemed to have a recall advantage for LGBTQ+ slurs. Arguably, non-heterosexual men should have demonstrated a recall advantage over heterosexual men, if there was both a significant interaction of gender and sexual identity, as well as a recall advantage for non-heterosexual participants regarding half of the items in the study. Given some of the issues with the post-hoc tests discussed in section 6.5.2.1., it is possible that the specific comparison between non-heterosexual and heterosexual men is less reliable. However, it is also possible that the apparent recall advantage for LGBTQ+ slurs among non-heterosexual participants may have been largely driven by non-heterosexual *women*, which is also plausible given that women outnumbered men in the sample by a ratio of nearly 3 to 1 (see Table 6.3.).

To try to better explain the recall likelihood results for this experiment, I now turn to the models I was able to create to predict these results, for all words in the experiment and for the LGBTQ+ slurs specifically. I believe that both of these can help explain why a difference was present for just non-heterosexual participants, despite there being no experimental condition, gender, nor sexual identity differences in recall likelihood across all participants.

6.5.2.5. Best Model of Recall Likelihood Across Conditions

The best model of recall likelihood for all words in the experiment was similar to that of the previous experiment (see section 5.5.2.5.). As with those results, the free recall model differed to the lexical decision models presented in Tables 6.4. and 6.5., in that neither gender nor sexual identity contributed significantly to the model. In this regard, this model provided no early support for any of my hypotheses, unlike the overall model of lexical decision results. Instead, as Table 6.8. shows, the best significant model of response times included just the normed familiarity ratings, and predicted that recall likelihood increased when familiarity was higher. The model was significant at the $p < .001$ level, was able to explain 2% of the variance in recall likelihood, and could correctly classify 70.50% of cases. The only difference between this model of free recall results and that of the previous experiment was that this one did not include normed negativity ratings. However, given that – unlike in the previous experiment – all of the words in this experiment were taboo words likely to be

judged as comparably negative, it is not surprising that negativity did not add significantly to the model as it had before.

As I discuss in sections 6.3.4. and 6.5.1.6., the issue of multicollinearity in the data meant that familiarity and personal use ratings could not be entered into the same model. However, personal use did not add significantly to the model when substituted for familiarity, despite the strength of the correlation between the two sets of ratings. As such, I determine that familiarity was indeed the factor affecting recall likelihood.

When discussing the relationship between familiarity and recall likelihood in the previous experiment (see section 5.5.2.6.), I noted that this finding does not replicate those of previous studies like Madan et al.'s (2017), in that non-emotional word properties like familiarity were not found to have contributed significantly to their model of recall likelihood. However, I pointed out that it nevertheless seemed logical to have found that words which are more familiar and more widely used would be the easiest for participants to remember. Having now found this result in both of my lab-based experiments, I am yet more convinced of the relevance of familiarity to recallability, when comparing slurs to both non-taboo language and other kinds of taboo language.

There is another reason I believe this model of recall likelihood to be important: I believe it clarifies the difference between conditions among non-heterosexual participants discussed in section 6.5.2.4. The results of my norming study identified that familiarity ratings for LGBTQ+ slurs were higher among non-heterosexual participants than they were among heterosexual participants. The model outlined in Table 5.7. shows that increased familiarity led to increased recall likelihood overall, in an experiment where 50% of the real words were LGBTQ+ slurs taken from the same norming study. Considering these findings together, it makes sense that an overall difference in recall likelihood between LGBTQ+ slurs and non-slur taboo words manifested, but only among the non-heterosexual participants.

With this in mind, it is perhaps surprising that familiarity was not part of the model which best predicted recall likelihood for just the LGBTQ+ slurs across all participants, which I will now discuss.

6.5.2.6. Best Model of Recall Likelihood for LGBTQ+ Slurs

As with the model of recall likelihood across conditions, the best model of recall likelihood for just the LGBTQ+ slurs in the experiment was similar to that of the previous experiment (see section 5.5.2.7.). As Table 6.9. shows, recall likelihood for the LGBTQ+ slurs was best predicted by a combination of number of characters, tabooness and normed overall reclamation ratings – the mean of each LGBTQ+ slur's reclamation behaviour ratings from the norming survey. The model predicted that recall likelihood increased when number of characters decreased, but both tabooness and overall reclamation ratings increased. In this regard, the model provided partial support for my second hypothesis, in that a measure of reclamation was relevant to recall performance, but reclamation did not appear to have counteracted the effects of tabooness on recall. Instead, reclamation seems to have mirrored the effect of tabooness, in that recall likelihood of the slurs improved as overall reclamation ratings increased. The absence of any measure of gender or sexual identity from this model, despite there being a measure of reclamation present, also seems to disprove my third hypothesis. The model was significant at the $p < .001$ level, was able to explain 8% of the variance in recall likelihood, and could correctly classify 72.20% of cases.

Again, the presence of multicollinearity in the data meant that tabooness could not be entered into the same model alongside arousal, negativity or offensiveness, nor could overall reclamation be entered alongside familiarity, personal use or the other normed reclamation behaviours. However, as before, none of the other properties which strongly correlated with tabooness and overall reclamation were significant when substituted. As such, I again determine that alongside number of characters, tabooness and overall reclamation were indeed the factors affecting recall likelihood.

This model of LGBTQ+ slur recallability was similar to that of the previous experiment (see section 5.5.2.7.) in that is too included tabooness and a measure of reclamation behaviour, but was different in that it also included character length, and the relevant reclamation behaviour was overall reclamation rather than self-reclamation.

As in the previous experiment, the finding that it was tabooness – rather than any other emotional word property – which helped predict LGBTQ+ slur recall likelihood does more closely replicate findings from existing literature, which suggest that taboo words demonstrate the most exaggerated increase in recallability (Buchanan et al., 2006; Jay et al., 2008; Kensinger & Corkin, 2003; Madan et al., 2017). In section 5.5.2.7., I noted that tabooness only contributed to the model in the previous study when all the words being modelled were taboo words, i.e. just the LGBTQ+ slurs. I suggested this meant the concern was not whether a word was taboo, but how taboo the taboo word was felt to be. However, all of the words in this study were taboo words, and tabooness still only contributed to the model of the LGBTQ+ slurs, rather than the overall model discussed in section 5.5.2.5. I therefore reconsider that degree of tabooness mattered specifically for the LGBTQ+ slurs, rather than taboo language in general.

The relevance of character length to the recallability of just the LGBTQ+ slurs was an unexpected finding, for which I consider two possible explanations. The first is simply that shorter slurs were easier for participants to remember by virtue of the fact that they weren't as long. The second considers the impact of familiarity on overall recall likelihood, as discussed in section 6.5.2.5. Referring to table 6.2., the slurs included in this experiment which were only three or four characters long (the two shortest character lengths) were also among those rated most familiar in my norming study (see Appendix 4). It is therefore possible that character length was also reflecting an element of slur familiarity, in addition to any benefit to of a word being shorter to its recallability.

As with all the models of LGBTQ+ slurs produced from both of my lab experiments, results again showed that linguistic reclamation had an impact on the LGBTQ+ slurs, this time through increased normed overall reclamation ratings increasing likelihood of recall likelihood. This replicates the finding from the previous experiment that the normed ratings contributed to the recall likelihood models, while participants' own ratings seem to contribute more to lexical decision models (see sections 5.5.1.6. and 6.5.1.6.). This supports a developing argument that participants' personal engagement in reclamation of an LGBTQ+ slur is more relevant to response times, but for recallability it is more important that reclamation of a slur is being more widely undertaken outside of the lab.

As I mentioned previously, where these results contrast those of the previous experiment is that it was normed overall reclamation – rather than self-reclamation – which contributed to the model. It should be noted that as the mean of all reclamation ratings given in the norming study, overall reclamation ratings do incorporate both of the self-reclamation behaviours, so a specific impact of self-reclamation compared to awareness of others' reclamation behaviour may still be present here. However, after completing this experiment, it was clear that overall reclamation was

generally the measure of reclamation behaviour which made the best individual contribution to models of both response time and recall likelihood (see Tables 5.5., 6.5. and 6.7.). In this regard, it is the model from the previous experiment which only included self-reclamation (see Table 5.9.) that is the more unusual.

6.6. Conclusion

This experiment aimed to compare the lexical accessibility and recallability of LGBTQ+ slurs to other examples of taboo language which were not slurs. It also sought to identify which – if any – combinations of participant identity, various word properties and linguistic reclamation behaviours could best predict and explain both types of processing. I consider the experiment to have been successful in achieving both of these aims.

The experiment found that performance in the lexical decision task did not differ by experimental condition, but did differ by both participant gender and sexual identity: response times were faster overall among women and among non-heterosexual participants. This contrasted the findings of my previous experiment, in which response times were found to be slower overall in the same groups. I suggested that the explanation I gave for these delays in the previous experiment may only apply when taboo words are being contrasted with non-taboo words. When they were instead being compared with other taboo words, as they were in this experiment, I suggested that the lack of taboo/non-taboo contrast may increase the importance of other factors, such as the gendered nature of particular stimuli and the increased exposure to LGBTQ+ slurs experienced by members of the LGBTQ+ community. Again, gender and sexual identity were the only factors able to produce a significant model of all response times in the lexical decision task, which also demonstrated that response times to all words in the experiment could be predicted to be faster among women and non-heterosexual participants.

Regarding performance in the free recall task, there were no overall differences in experimental condition, gender, nor sexual identity, suggesting that LGBTQ+ slurs are not generally recalled any differently to other taboo words. However, there was a difference between LGBTQ+ slurs and non-slur taboo words among the non-heterosexual participants alone, in which the slurs were recalled significantly more often. I argued that this finding suggests that – at least in regards to recallability – slurs might be processed differently by the group which they target. The model which best predicted recall likelihood across all words in the experiment included just word familiarity, indicating that increased recall likelihood was associated with increased familiarity ratings, which I related to the increased recallability of LGBTQ+ slurs found among the non-heterosexual participants.

As in the previous experiment, both of the models designed to predict response times and recall likelihood for just the LGBTQ+ slurs showed that some aspect of linguistic reclamation was relevant, again confirming my hypothesis that it would be. Alongside participant gender and sexual identity, participants' own overall reclamation ratings were found to contribute to the model of LGBTQ+ slur response times, indicating that participants who generally reported greater reclamation of a slur also displayed less delay in their response times to the same slur. Alongside character length and tabooness ratings, normed overall reclamation ratings also contributed to the model of LGBTQ+ slur recall likelihood, indicating that participants more frequently recalled slurs which are more widely reclaimed.

Chapter 7 – Comparing Pupil Dilation Responses to LGBTQ+ Slurs and Non-Slur Taboo Words

7.1. Introduction

This chapter describes the fifth and final study of my research, which was a pupillometry study – one which uses eye-tracking technology to examine changes in the diameter of the pupil as an indicator of various aspects of cognitive language processing (Sirois & Brisson, 2014). Building on the research outlined in the previous chapters, my study sought to compare pupil dilations in response to LGBTQ+ slurs and other taboo words which were not slurs, in order to determine if these categories differed in eliciting emotional arousal. I decided to focus this study on emotional arousal for two reasons. First, given that emotional arousal has been so closely linked to taboo language processing in the existing literature – particularly with regard to its recallability – it was unexpected that arousal did not significantly predict the results of any of my lexical decision or free recall tasks described in Chapters 5 and 6. Therefore, I wanted to further test arousal responses in relation to the LGBTQ+ slurs I used as stimuli in my previous experiments. Second, I wanted to confirm the sensitivity of the conscious ratings of arousal, given that it is an automatic and unconscious emotional response. In my previous lab experiments, arousal was one of many emotional and non-emotional properties I was testing for an effect on slur processing, and collecting conscious ratings of these was an efficient way to measure all properties within a single study (see Chapter 4). However, as the current experiment was to focus only on arousal, I wanted to explore whether a direct, physiological measure might be more sensitive. This would then be able to confirm whether the normed arousal ratings accurately represented actual arousal responses, or else that arousal did not predict participant task performance in my previous experiments because the measure of that particular property was not sensitive enough. I therefore decided to conduct another study which specifically explores arousal responses using a measure that is not consciously controlled. In section 7.2. of this chapter, I explain why I selected pupil dilation in particular for this purpose. As in my previous studies, I was also interested more generally in whether there were any associations between the linguistic reclamation of LGBTQ+ slurs and emotional responses to them.

Data collection was postponed shortly after lab testing began, due to the COVID-19 pandemic. As I detail in section 7.3.1. of this chapter, I was able to gather a small sample of 11 participants before needing to pause the study, but this number did not have the statistical power needed for the analyses I intended to run (t-tests and hierarchical multiple regression modelling, see section 7.3.4. for further discussion). While I could have conducted descriptive analyses of this data – such as the average pupil size for particular conditions and participant groups – these would be limited in their ability to determine any effects of reclamation on arousal responses, and would not provide reliable or representative indications of group differences. Instead, I used the time gained to develop my discussion of the jointly psycholinguistic and sociolinguistic significance of my four completed studies, in particular by returning to all of my data and further theorising why the issues of gender, sexual identity, and ingroup/outgroup membership so consistently interacted with both LGBTQ+ slur processing and reclamation behaviour. This became a key issue in my general discussion (Chapter 8).

For this reason, the focus of this chapter is my finalised experiment design, rather than data analysis. Nevertheless, I consider this an important discussion, in that I will demonstrate how a pupillometry experiment can be designed to test the relationship between linguistic reclamation and the emotional arousal of slurs, and why it is a particularly novel and effective means of doing so. I

begin by outlining the existing research which links pupil dilation to both emotional arousal and emotional language processing. I also discuss the specific advantages of pupillometry for investigating taboo language processing, and the potential effects of linguistic reclamation on the same. I then discuss in detail my intended participant sample, experiment design, experiment procedure and methods of data analysis. Because I did not collect sufficient data to analyse and discuss, I instead devote a section of this chapter to considering possible outcomes of this study, and how each would develop the findings outlined in the previous chapters.

7.2. Research Background

This experiment shares much of its theoretical grounding with the studies conducted in Chapters 5 and 6. For a comprehensive discussion of the impacts of emotional arousal, tabooness and linguistic reclamation on taboo language processing, refer to sections 2.2. and 2.3. of my literature review (Chapter 2). For discussion comparing slurs to other kinds of taboo language, identifying the relationship between slurs and linguistic reclamation, and highlighting the importance of these issues to taboo language processing experiments, refer to section 6.2. of Chapter 6.

Where this experiment differs from those described in previous chapters is in its use of pupil dilations as a physiological measure of emotional arousal, rather than self-reported ratings. In order to explain how this is useful to my research, I refer to Sirois and Brisson (2014), who conducted a comprehensive review of pupillometry research and concluded that pupil dilation is well-evidenced as a dynamic marker of language processing and a reliable measure of cognitive responses to emotional stimuli. Partala and Surakka (2003) explain that this is because pupil size is regulated by the autonomic nervous system – a control system which unconsciously regulates bodily functions, also including heart and respiratory rate – which the results of their own pupillometry experiment proved to respond differently to (positive and negative) emotionally arousing stimuli versus emotionally neutral stimuli. In section 2.3.1. of my literature review, I noted that the release of the hormones epinephrine and norepinephrine have been linked to the effect of emotional arousal on memory (Sharot & Phelps, 2004), and that these hormones are in turn related to the fight-or-flight response produced by the autonomic nervous system in response to arousing stimuli, particularly those perceived to be threatening. Pupil size is therefore a sensitive measure of autonomic nervous system activity, and by extension a sensitive measure of emotional arousal responses. Partala and Surraka also note several advantages of pupil dilation over other physiological responses. They describe it as a truly involuntary index of autonomic nervous system activity, so unlike other arousal indicators such as facial expression, it cannot be demonstrably inhibited, exaggerated, or faked (e.g. Ekman, 1985; Ekman and Friesen, 1982; Surakka and Hietanen, 1998). It is also less obtrusive in that it does not require sensors to be attached to an individual, the associated technology is relatively simple to use, and pupil responses are not as easily disrupted by other bodily movements.

Sirois and Brisson (2014) also identify pupillometry as particularly well-suited to questions of implicit processing, and to processes liable to be affected by social desirability bias. The processing of taboo language is not only a question of implicit processing, but also one for which I argue any conscious assessment of responses is likely to be affected by social desirability. Taboos are socially undesirable behaviours by definition (Allan, 2019), so it is reasonable to assume individuals may report unfavourable responses to taboo language, in order to demonstrate that they acknowledge and abide by its proscription. Accordingly, a major advantage of using a physiological response like pupil dilation to measure responses to slurs would be that it does not rely on

participants being reliable and unbiased judges of their own (unconscious) emotional responses. Based on the previous comparison of pupil dilation to other physiological indicators of arousal, it is arguably the best unconscious measure for this purpose.

However, their review also suggests that is relatively novel to use pupillometry to explore stimuli that are both emotional and linguistic. Language processing studies measuring pupil dilation have tended to focus on issues such as cognitive load during discourse processing (e.g. Zekveld & Kramer, 2014; Zellin, Pannekamp, Toepel, & van der Meer, 2011) speech intelligibility (e.g. Zekveld & Kramer, 2014), word retrieval (e.g. Kuipers & Thierry, 2013; Schmidtke, 2014) and attention while reading (e.g. Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Franklin, Smallwood & Schooler, 2011; Smallwood, McSpadden, & Schooler, 2008). Furthermore, these studies have typically used spoken rather than written language. Similarly, the association between pupil dilation and emotional arousal – in which arousing stimuli are associated with increased pupil dilations – has largely been evidenced by research involving emotional images as opposed to emotional language (e.g. Aboyoun & Dabbs, 1998; Bradley, Miccoli, Escrig, & Lang, 2008; Steinhauer, Boller, Zubin, & Pearlman, 1983).

Where studies have specifically investigated pupil dilation responses to emotional language, results have been less uniform. Some studies have suggested that unlike emotionally arousing images, pupil dilation as a physiological response to emotional language is caused by processing difficulty and not directly by emotional arousal (Siegle, Granholm, Ingram, & Matt, 2001; Vō et al., 2008; Bayer, Sommer, & Schacht, 2011). However, Bayer et al. (2011) acknowledge that ease of language processing is itself modulated by emotional arousal, citing a similar body of evidence that I discuss in Chapter 2 (section 2.3.) and Chapter 5 (section 5.2.). On this basis, even if emotional arousal is not the direct cause of pupil dilation responses to emotional language, pupil dilation is arguably still an indicator of differences in emotional arousal, insofar as it indicates language processing difficulty.

To further support this, a more recent study by Scheepers, Mohr, Fischer, & Roberts (2013) found that pupil responses to detecting rhyme violations while reading limericks were indicative of an emotional response, which could not be explained solely as anomaly detection during written language processing. Several aspects of this study informed the predictions and design of the current study. In it, 40 native speakers of British English (with normal hearing and normal or corrected-to-normal vision) heard spoken limericks belonging to 1 of 5 manipulated conditions, while their pupil dilation responses were recorded using an eye-tracker. In 4 of the 5 conditions, the last word of the limerick violated the familiar pattern of a limerick in some way: semantically (its meaning did not match the rest of the limerick), syntactically (there was a word-order violation immediately before the last word), a rhyme scheme violation (the last word broke the familiar AABBA rhyme scheme of a limerick), or a metric violation (a syllable was added to or subtracted from the last word such that the familiar metre of a limerick was disrupted). As a control measure, the limericks used for each condition were alternated among 5 separate lists, which were in turn alternated among participants, so that each of the 5 lists was seen by a fifth of the total participants. The last condition was a control condition in which there were no violations, and filler limericks were also included which similarly included no violations but were not being measured as part of the task. A repeated measures ANOVA found that the rhyme violation condition elicited significantly different pupil dilations compared to the control condition, while none of the other conditions differed significantly. A further confirmatory experiment revealed that all violations were reliably detected by participants, so the authors argued some aspect of the rhyme violation produced a qualitatively different

response to other kinds of violation, which was being indicated by pupil dilation. Because the rhyme scheme is such a strongly familiar and expected aspect of a limerick, the authors reasons that pupil dilations were reflecting some kind of emotional response to the shock and/or confusion caused by its disruption. Therefore, they concluded that pupil dilations could demonstrably be associated with emotional responses, particularly those concerning surprise and – potentially – emotional arousal.

I therefore concluded that a pupillometry study could support and develop a number of the findings from the studies described in Chapters 4 to 6. Based on my findings on word recallability in Chapter 6, I predict an interaction between sexual identity and taboo word-type: specifically, that solely among non-heterosexual participants, pupil size will be significantly larger (indicating greater emotional arousal) in response to LGBTQ+ slurs compared to non-slur taboo words. Additionally, based on the findings of Chapters 4-6, I predict a main effect of reclamation behaviour in the case of LGBTQ+ slurs: specifically, that pupil size will be significantly smaller (indicating lower emotional arousal) when participants report increased personal and/or general population reclamation of a slur.

7.3. Methodology

7.3.1. Sample

At the time of postponing data collection, 11 participants had taken part in the experiment. Of these, 5 were recruited through a combination of posts on social media (Facebook and Twitter) and research posters placed around the University of Nottingham campus. The remaining 6 participants were undergraduate students at the University of Nottingham, who were recruited in exchange for course credit.

All participants were aged between 16 and 29 years old. As in the experiments described in Chapters 5 and 6, this range was selected to control for participant age, while also acknowledging the differences in LGBTQ+ slur ratings between 16-29 year olds and 30+ year olds that I identified in Chapter 4. I collected additional respondent demographic data in order to enable demographic comparisons during analysis, which are presented in Table 7.1. There was a mean participant age of 19.82 (SD: 2.99), the majority of participants identified as women, all participants were cisgender, and more participants identified as non-heterosexual than heterosexual.

Table 7.1. – Participant Demographic Data

Age	***N***	**%**	**Gender**	***N***	**%**	**Cisgender?**	***N***	**%**	**Sexual Identity**	***N***	**%**
18	7	63.64	Woman	7	63.64	Yes	11	100.00	Straight/Heterosexual	4	36.36
19	1	9.09	Man	4	36.36	No	0	0.00	Gay/Lesbian/Homosexual	3	27.27
23	1	9.09	Total	11	100.00	Total	11	100.00	Bisexual/Pansexual/Polysexual	4	36.36
24	1	9.09							Total	11	100.00
26	1	9.09									
Total	11	100.00									

All participants were native speakers of British English, for three reasons. First, all of the LGBTQ+ slurs in the experiment were taken from those generated by British English speakers in the slur identification survey described in Chapter 3. Second, as discussed in section 2.1.4. of my literature review (Chapter 2), native speakers are likelier than non-native speakers to accurately judge the tabooness of a word (Dewaele, 2019; Jay, 1992, 2000; Thomas, 1983). Finally, unlike the

experiments described in Chapters 5 and 6 – which allowed near-native British English speakers to participate – restricting participation to native speakers only was necessary because of the nature of the experiment. This was because the task involved taking pupil measurements during fixed-pace reading of entire sentences, as opposed to allowing variable response times to individual words. It was therefore necessary to ensure both that participants could read in the target language at the speed sentences were being presented, and that variations in pupil dilations would more likely be caused by the meaning of the words rather than their novelty or unfamiliarity.

All participants were also required to have normal or corrected-to-normal vision, no conditions affecting eye movement (e.g. strabismus), and no conditions that might affect their reading or reading comprehension (e.g. dyslexia).

7.3.2. Materials

There were 60 trials in my experiment, each of which consisted of a single sentence, containing a target word. The sentences varied in length – all between 7 and 13 words long – to avoid the positioning of the target word becoming easy to anticipate, and to enable the creation of more natural sentences. However, target words were never positioned at the beginning or end of a trial sentence. Target words were not positioned at the beginning to ensure that some context could be established before their presentation; results of a study by Gredebäck and Melinder (2011) indicated that pupil dilations are sensitive to contextual constraints. Not positioning them at the end ensured enough time following presentation of the target word to record any pupil dilation responses that might be slightly delayed; Scheepers et al. (2013) indicate that pupil dilations in response to a particular word can occur as much as 2000ms after word onset. Accordingly, I made sure that all target words were always followed by at least 3 other words, to allow a minimum of 2000ms after their presentation (500ms presentation time per word + 500ms fixation points between words).

Target words belonged to one of two word-type conditions: 13 x LGBTQ+ slurs (as examples of reclaimable taboo language) and 13 x taboo words which were not slurs (as examples of non-reclaimable taboo language). The LGBTQ+ slurs were selected from the normed database described in Chapter 4, and the non-slur taboo words were taken from a normed database produced by Janschewitz (2008). Target words from both categories were selected so that they matched for number of characters and number of syllables. Table 7.2. provides examples of the target words used in the experiment, divided by condition.

Table 7.2. – Example Stimuli for Each Target Word Condition

LGBTQ+ Slurs	Non-Slur Taboo Words
Bent	Bastard
Lesbo	Crap
Poof	Dick
Queer	Loser
Shemale	Prick
Sissy	Pussy

The majority of target words were nouns, and were included as part of a simple noun phrase (determiner + noun) within a trial sentence, e.g. ‘a fag’, ‘a moron’. Two of the LGBTQ+ slurs I wished to measure responses to were adjectives (‘bent’ and ‘camp’), so were presented as complements to a preceding noun phrase with the same structure (e.g. “His teacher was bent…”), and were matched with a proportional number of adjectival non-slur taboo words. This ensured grammatical

consistency between conditions, meaning that certain grammatical features (e.g. presence of a determiner) could not be used to discriminate between the two. All experimental trials were structured to ensure as much grammatical consistency as possible up to the presentation of the target word, always beginning with a noun phrase (either a personal name or pronoun), followed immediately by a verb phrase. This was so that no degree of arousal would be prompted by reading a noticeably different sentence to the previous one; a study by Naber, Frässle, Rutishauser, and Einhäuser (2013) suggested that novelty alone can influence pupil dilation. I considered this less important following presentation of the target word, as any effect of target word on pupil size would already have been initiated, but following the findings from Scheepers et al. (2013) discussed earlier, I maintained some grammatical consistency for up to 2000ms following target words.

To ensure that any potential effects of the contextual information on either side of target words were controlled for, I paired each target word in the slur condition with a target word in the non-slur taboo condition, such that both had the same number of characters and syllables. I then paired two sentence contexts, such that a target word pair could be swapped between both and still make sense. I used these pairs to create two versions of the experiment, in which all sentence contexts containing a slur in Version 1 would contain a non-slur taboo word in Version 2. These were alternated between participants such that no participant saw both versions, and that at the end of data collection, target words in both conditions will have been presented in two different contexts an equal number of times. In Table 7.3., I illustrate this with four trials used in the experiment, contained a pair of target words and the pair of contexts they were alternated between in either version.

Table 7.3. – Example Target Word and Context Pairings in Experimental Trials

Version	Context	Paired Target Word	Experimental Trials
1	A	Loser (non-slur taboo word)	Amy would not invite a **loser** to her birthday party.
	B	Lesbo (LGBTQ+ slur)	People thought she was a **lesbo** and wouldn't hang out with her.
2	A	Lesbo (LGBTQ+ slur)	Amy would not invite a **lesbo** to her birthday party.
	B	Loser (non-slur taboo word)	People thought she was a **loser** and wouldn't hang out with her.

Thirty-four filler trials were also included in both versions, consisting of sentences with identical grammatical structure to those containing target words, but containing a filler word unrelated to the experiment. These filler words were matched with target words for word class, number of characters, and number of syllables. The order of all trials used in the experiment was randomised. All trials used in both versions of the experiment are provided in Appendix 9.

7.3.3. Procedure

It is well established that changes in ambient luminance provoke changes in pupil dilation (Sirois & Brisson, 2014; Scheepers et al. 2013), so before participants arrived, I ensured that the lights in the lab were turned on and working properly, and that participants would be given enough time to adjust to the ambient light level in the room before beginning the experiment. Upon their arrival, participants gave their informed consent. They were given time to ask questions and prepare to take part, e.g. by removing eye makeup that could disrupt the ability of the eye-tracker to identify the pupil. They were then sat in front of the computer on which the eye-tracking task was running, and placed their head on a chin rest to minimise head movement. The chin rest was positioned 79cm from the centre of the screen, which – when factoring in screen size and screen resolution (1024 x 768) – ensured that the degree of visual angle was under the recording accuracy threshold for the

eye-tracker (< 0.5 degrees) and all words would be within field of vision when viewing the centre of the screen. Participants were allowed to adjust desk height, chair height and keyboard position for comfort.

The experiment was split into 3 blocks of 20 trials, with a break in between each, to ensure that participants did not lose attention due to fatigue. Before these, I included a practice block of 5 unrecorded trials containing sentences unrelated to the experiment topic, so that participants had an opportunity to get used to how the experiment worked. Before the start of each block, participants were reminded of the task instructions, then the eye-tracker was calibrated.

Trials required participants to read a number of sentences on a computer screen for comprehension, while an SR Research Eyelink 1000 eye-tracker (recording speed: 1000Hz) recorded pupil dilation in response to individual target words. Viewing was binocular, but only one eye was recorded, as both eyes are able to demonstrate the same pupil dilation reactions.

Rather than presenting sentences in full, the experiment used Rapid Serial Visual Presentation (RSVP), in which sentences were presented one word at a time. This is consistent with the design of other word-reading pupillometry studies, such as Franklin et al. (2013), who also suggest that – in combination with consistent presentation durations and the use of fixation points in-between words – RSVP allows for accurate discrimination of pupil dilations for individual words. To ensure that all participants saw the target words for the same amount of time, reading was not self-paced, and the time each word appeared on screen was fixed at 500ms. This also helped ensure that effects of particular words could be identified during analysis, as all words were been presented at set time intervals.

All words were presented in the exact centre of the screen, to ensure that participants' fixations occurred in approximately the same position. All words were presented in Arial font, size 40pt, in black text on a white background. This font size is larger than the 20-30pt range often used in experiments which do not use RSVP, but I made an adjustment following piloting feedback that words appeared too small when presented individually in the centre of the screen. All characters were capitalised so as to have identical height, again to keep fixation position more consistent. I acknowledge that words are not typically written in all capitals, which risked affecting speed of word recognition and introducing unintentional novelty to the sentences. However, I considered that inconsistent pupil position presented a greater threat to accurate pupil size recording, that the 500ms presentation time would accommodate any short recognition delay, and that the total capitalisation of the practice block stimuli would reduce novelty for participants during the experimental block. These formatting decisions are common in pupillometry studies involving RSVP reading (e.g. Franklin et al., 2013; Franklin, Smallwood & Schooler, 2011; Smallwood, McSpadden, & Schooler, 2008), as they collectively minimise disruptions to sentence reading as a result of presenting words one at a time.

At the end of certain trials, participants were presented with a basic yes/no comprehension question, in which they were asked a question about the sentence they had just read and told to provide a keyboard response before continuing ('z' for 'yes', 'm' for 'no'). These were included to encourage participants to pay attention and check task engagement prior to data analysis; the questions were not otherwise relevant to the experiment. They were not included at consistent intervals, so that participants could not predict them, but they were always included after a filler trial. This positioning ensured that reading the question would not excessively cue attention to target trials.

Between all trials, a fixation point (a '+' symbol) was presented in the exact centre of the screen for 500ms, to refocus participants' gazes where words were being presented. This was particularly necessary following comprehension questions, which diverted gaze away from the centre of the screen. Because words were being presented one at a time in the same location, and because neither fixations nor saccades were being measured, it was not necessary to perform drift corrections before the start of a new trial.

Following completion of the sentence-reading task, participants then completed two post-test surveys, the first to gather demographic information (see Table 7.1.), the second to determine each participants' own reclamation behaviours. Both surveys were the same as those used in the experiments described in Chapters 5 & 6 (see Appendix 3), as they were collecting the same information using the same measures. Accordingly, the design and procedural details relevant to these surveys can be found in sections 5.3.2.3. and 5.3.3 of Chapter 5.

Completion of the survey task marked the end of the experiment. Participants were then given a final opportunity to ask questions about the research, and were thanked for their involvement. Overall, completion took between 30-40 minutes per participant.

7.3.4. Data Analysis

In this section, I discuss the approach to data analysis that I intended take upon completing data collection. In my experiment, each individual word in a target sentence represents its own Region of Interest (ROI), and data are recorded as the average pupil diameter value (in millimetres) during each ROI. Following the approach taken by Scheepers et al. (2013), I will use these values to determine the mean pupil diameter in a 2000ms time window from onset of the target word.

I will then use these values from every trial to first conduct by-subjects (averages for each participant) and by-items (averages for each target word) analyses of pupil dilation responses in the 2 x 2 x 2 x 2 structure (condition x gender x sexual identity x experiment version), using repeat-measures ANOVAs. This will enable me to confirm whether I adequately controlled for any impact sentence context (by seeing if pupil dilations did not significantly differ between both versions of the experiment), as well as whether any differences in the other independent variables were present when responses were averaged across both participants and items. It will also enable me to look for interaction effects across the data.

I will then conduct a number of related independent samples t-tests to further investigate differences between participants. For all such analyses, I choose this test because pupil size represented a continuous dependent variable all independent variables had two groups. First, I will determine whether there were any significant differences in pupil dilation between the two word-type conditions (LGBTQ+ slurs and non-slur taboo words) across all words and all participants. Second, I will use pupil size measurements combined with participants' responses to the demographic survey to explore whether there were significant differences in responses to all stimuli between specific participant demographic groups (e.g. women vs. men, non-heterosexual vs. heterosexual participants). Finally, by isolating the data pertaining to each individual demographic group, I will conduct a series of further independent samples t-tests to determine whether there were significant differences between conditions within specific demographic groups, e.g. whether there was a difference in pupil dilation responses to LGBTQ+ slurs and non-slur taboo words among the non-heterosexual participants.

To produce a model which could explore the potential impact of all of the word property and reclamation behaviour ratings described in Chapter 4, as well as participant demographic data, I

will again run a series of hierarchical multiple regressions to find which of these independent variables best explain pupil dilation responses. This procedure is appropriate for generating a model for a continuous dependent variable such as pupil size, while also allowing me to test whether the model could be improved with the addition of further independent variables (Gelman & Hill, 2006). However, this procedure would require checking that the complete dataset meets a number of assumptions, including checking for and removing outliers (Berry, 1993; Cook & Weisberg, 1982; Fox, 1991); checking that all data points had safe leverage values (Huber, 1981); checking that no data points were overly influential (Cook & Weisberg, 1982); checking for linearity (Berry, 1993; Fox, 1991) and checking for multicollinearity between independent variables.

7.4. Discussion of Potential Findings

In this section, I outline several possible outcomes that could emerge from the study once it is run to completion, and discuss the implications of each for my research overall.

7.4.1. Difference Found between Word-Type Conditions

The first possible outcome I will discuss is the one I consider least likely: that across all participants, LGBTQ+ slurs would elicit significantly different pupil dilations to non-slur taboo words. I base this prediction on the findings of the study I describe in Chapter 6, which also compared responses to these two word types in both a lexical decision and free recall task, finding no overall differences between either condition in any of these tasks. If pupil dilation responses to slurs were found to differ to other taboo words in either direction, the findings of this study would be contradictory. This is because they would suggest an overall difference in emotional arousal between the two conditions, which would in turn be expected to produce a difference in processing not identified in any of my previous experiments (Siegle et al., 2001; Võ et al., 2007; Bayer et al., 2011).

One possibility is that this would indicate a difference in emotional arousal that simply did not affect language processing, but this explanation would not be well supported by the literature. I argue it would also not be supported by the results of the experiment I describe in Chapter 5. This experiment used LGBTQ+ slurs as examples of taboo words, and compared their performance in both a lexical decision and free recall task to high- and low-arousal non-taboo words. As discussed in section 2.3.2. of my literature review, taboo language has been shown to demonstrate the most exaggerated effects of emotional arousal on language processing (Jay, Caldwell-Harris, & King, 2008; Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Madan, Schafer, Chan, & Singhal, 2017), and this experiment sought to test if this held true when comparing LGBTQ+ slurs to non-taboo words that differed in terms of arousal. Any difference in pupil dilation between LGBTQ+ slurs and other kinds of taboo language would require me to consider that the results of the study I describe in Chapter 5 did not just represent a comparison of taboo and non-taboo language, but specifically of slurs and non-taboo language.

While that study found no overall differences between the three conditions in terms of lexical decision response times (see sections 5.4.2.1., 5.5.1.2. for discussion), it did find that LGBTQ+ slurs were over twice as likely to be recalled as either of the non-taboo word types across all responses, and that high-arousal non-taboo words were better recalled than low-arousal ones (see sections 5.4.2.2., 5.5.2.2. for discussion). Therefore, although that experiment was not able to make a comparison between different types of highly arousing taboo language, it did demonstrate some overall effects of increasing emotional arousal on language processing. This would not support an

argument that a difference in emotional arousal could have been present without affecting processing.

If there is a difference in arousal responses between conditions which does not affect other aspects of processing, another question concerns what the nature of this difference is. Given that I found no overall differences between LGBTQ+ slurs and non-slur taboo words in the previous experiment, I cannot confidently predict which condition would exhibit greater pupil dilation responses. However, I can consider what the implications of both possibilities might be. If LGBTQ+ slurs generated greater pupil dilation responses, this would suggest that they are more emotionally arousing than other kinds of taboo language. Based on the literature (see sections 2.3.1 and 2.3.2. of Chapter 2), this would in turn suggest that they should perform better in free recall tasks, but worse in lexical decision tasks.

This would contradict my findings from the study described in Chapter 6 that there were no overall processing differences between LGBTQ+ slurs and other kinds of taboo language, but it would have specific implications for findings pertaining to particular demographic groups as well. In particular, the same study found that while there was no overall difference in recall likelihood between the two conditions, LGBTQ+ slurs were significantly better recalled than non-slur taboo words by the non-heterosexual participants (see section 6.4.2.2.). Furthermore, this was the only demographic group to exhibit such a difference. If pupil dilation responses indicated that LGBTQ+ slurs were more emotionally arousing, this could help explain this finding if arousal-enhanced memory is considered. Compared to other kinds of taboo language, LGBTQ+ slurs might be more memorable to non-heterosexual people in particular, because they may be more emotionally arousing to that particular group. If the opposite were true – that non-slur taboo words elicit a greater pupil dilation response – then this difference in recall likelihood among non-heterosexual participants would have to be attributed to something other than emotional arousal, potentially greater familiarity.

An alternative explanation for finding an overall difference between arousal responses to LGBTQ+ slurs and non-slur taboo words in this experiment would be that pupil dilation is simply a more sensitive measure of arousal than the conscious ratings I generated from my norming study described in Chapter 4. By extension, my two lexical decision and free recall experiments (Chapters 5 and 6) may have found no effect of arousal on response time or recall likelihood because the means of measurement – while generally appropriate for those experiments – was not sensitive enough to detect them, as opposed to them not existing. While this would challenge the lack of arousal effects found in the experiments described in Chapters 5 and 6, it would suggest that such an effect could exist, which is a possibility supported by the literature linking arousal to lexical decision and free recall (refer to sections 2.2., 2.3., 5.2. and 6.2. for discussions), and would be supported by the evidence I present in 7.2. of this chapter for the sensitivity of pupil dilation as a measure of arousal.

7.4.2. No Difference Found between Word-Type Conditions

Next, I consider the opposite outcome – that no differences whatsoever are found between pupil dilation responses to LGBTQ+ slurs and to non-slur taboo words. Based on the absence of overall processing differences between both conditions in the previous study (see sections 6.4.2.1, 6.4.2.2.) I consider this a likelier outcome than the one discussed in 7.4.1. Unlike that outcome, finding no general difference between conditions in this experiment would support viewing my normed, self-reported arousal ratings as a reliable measure. However, if no differences between the two

conditions were found within any participant demographic groups, this would mean that any differences in processing identified by previous experiments – such as increased recall likelihood for LGBTQ+ slurs among non-heterosexual participants – could not be attributed to differences in eliciting emotional arousal. Although I do believe this outcome is likelier than finding a general difference across all participants, I do not consider it to be the likeliest outcome, to which I now turn.

7.4.3. No Difference Found Overall, but Differences Found within Specific Demographic Groups

Based on the findings of the studies described in Chapters 5 and 6, and my discussion of whether I would or would not find a difference in pupil dilation between LGBTQ+ slurs and non-slur taboo words, I believe that the likeliest outcome lies between both extremes. Namely, that there would be no overall difference in pupil dilation responses to LGBTQ+ slurs and non-slur taboo words, but such differences may exist within specific demographic groups. As with the outcome discussed in 7.4.2., I would still consider this result to support a view of my normed, self-reported arousal ratings as a reliable measure, as each normed rating represents a general-population mean; there being no overall difference in pupil dilation responses to either condition would suggest no equivalent difference in arousal across all of my participants, but this does not preclude such differences from existing within specific groups. Furthermore, my discussion in sections 4.5.2. to 4.5.5. of Chapter 4 clearly indicates that arousal ratings did differ significantly between specific age, gender, and sexual identity groups, which would also be compatible with this finding.

This is largely due to the relevance of demographic groups to performance in the lexical decision and free recall experiments I have conducted. When also comparing LGBTQ+ slurs to non-slur taboo words, I found faster response times in the lexical decision task among women and non-heterosexual participants, regardless of word type (see section 6.4.2.1.). This experiment would be able to compare pupil dilation responses based on gender and sexual identity, so would be able to determine whether emotional arousal responses to both types of taboo language are affected by either of these factors. If so, then these effects may have superseded any differences in lexical access between slurs and other kinds of taboo language, if indeed such differences exist at all. If not, then a factor other than emotional arousal must underpin gender and sexual identity differences in response times. The latter possibility is especially relevant to instances where I have drawn on sociolinguistic work on gendered demands for and expectations of politeness (e.g. Brown, 1980; Coates, 1988; Fishman, 1980; Holmes, 1995) to explain instances where women's performance in the lexical decision task was worse (see sections 5.5.1.3., 6.5.1.3.). Similarly, the finding that recall likelihood was greater for LGBTQ+ slurs compared to non-slur taboo words among non-heterosexual participants could be further clarified by identifying whether pupil dilations showed the same pattern within this group.

It is worth noting that differences in pupil size could be identified within several different demographic groups, and the findings of my previous studies make it difficult to predict how these might manifest. For example, when comparing LGBTQ+ slurs to non-taboo language, women and non-heterosexual participants exhibited slower response times across all word types (see section 5.4.2.1.). However, as mentioned previously, women and non-heterosexual participants exhibited faster response times across all word types when comparing LGBTQ+ slurs to non-slur taboo language (see section 6.4.2.1.). While these findings make demographic differences hard to predict

for this experiment, they do highlight the usefulness this experiment would have in explaining such points of contrast.

7.4.4. Effect of Linguistic Reclamation on Pupil Dilation Responses to LGBTQ+ Slurs

I now move away from discussing the potential differences in pupil dilation responses to the two word-type conditions in the experiment, and toward the potential effect of linguistic reclamation on the same. In producing a model to explain pupil dilation responses, I would test the contribution of all word properties for which I collected ratings in Chapter 4, not just the reclamation behaviours (see section 4.3.2., Table 4.6.). However, there were 14 of these in total, so for this discussion I focus on reclamation behaviour ratings as the ones most central to addressing the third hypothesis of this experiment: that pupil dilations in response to LGBTQ+ slurs might be reduced when participants indicate increased personal and/or wider reclamation of those words.

If no measures of reclamation behaviour predicted pupil dilation response, this would present a number of issues for interpreting the findings described in previous chapters. First, it would directly contradict a key finding from Chapter 4, in which all reclamation behaviour ratings had a significant inverse correlation ($p < .01$) with arousal ratings (see Table 4.6.), with each reclamation behaviour sharing between 34% and 44% of the variance in its ratings with arousal (see Table 4.7.). Such a finding might provide a useful critique of self-reported measures of arousal; as a more objective measure of arousal, finding contradictory results in a pupil dilation study would suggest individuals do not make reliable conscious judgements of their own unconscious emotional responses. Nevertheless, such results would be unexpected given the strength and consistency of the correlations I found. Second, it would suggest that the effects of linguistic reclamation on LGBTQ+ slur processing described in Chapters 5 & 6 (see Tables 5.5., 5.7., 6.5., 6.7.) are not indicative of an effect of linguistic reclamation on emotional arousal. While such a finding would not challenge the idea that linguistic reclamation modulates LGBTQ+ slur processing, it would challenge my interpretation of *how* it does; rather than affecting emotional arousal, it might suggest that it is the effect of reclamation on familiarity that affects responses to slurs.

However, based on these same findings, I consider it likelier that some effect of linguistic reclamation on pupil dilation would be observed. This would in turn suggest a relationship between linguistic reclamation and emotional arousal, which might modulate the processing of slurs. I would be particularly interested in whether the same measures of reclamation that were found to be relevant in my previous experiments also affected pupil dilations, namely: participants' personal overall reclamation ratings for the lexical decision tasks (see Tables 5.5., 6.5.) and normed self-reclamation and overall reclamation ratings for the free recall tasks (see Tables 5.7., 6.7.).

7.5. Conclusion

In this chapter, I have demonstrated the potential usefulness of a pupillometry study to my research, in its use of an unconscious, physiological measure of emotional arousal that is less subject to individual biases or self-report inaccuracies. Referring to existing literature, I have highlighted why pupillometry studies are appropriate for the investigation of both language processing (Franklin et al., 2011; Franklin et al., 2013; Kuipers & Thierry, 2013; Schmidtke, 2014; Sirois & Brisson, 2014; Smallwood et al., 2008; Zeckveld & Kramer, 2014; Zellin et al., 2011) and emotional effects on cognition (Aboyoun & Dabbs, 1998; Bradley et al., 2008; Sirois & Brisson, 2014; Steinhauer et al., 1983), but also that it still quite novel to use such a study to explore both of these issues

simultaneously (Bayer et al., 2011; Siegle et al. 2001; Scheepers et al., 2013; Vō et al., 2008). I have indicated how I will use such a study to further test whether slurs differ from other kinds of taboo language in terms of eliciting emotional arousal, and explore whether linguistic reclamation behaviours modulate these responses.

Although my study could not be completed due to disruption caused by the COVID-19 pandemic, I have reviewed a number of possible findings that could arise once it is completed. Based on the findings from Chapters 4-6, I have discussed which of these I consider most likely and what its implications for my research would be. Of all outcomes, I consider it likeliest that there will be no overall difference in pupil dilation responses to LGBTQ+ slurs compared to other kinds of taboo language, but that responses may differ within particular demographic groups, most notably non-heterosexual participants. I also consider it likeliest that some measure of reclamation behaviour will be found to modulate pupil dilation responses. If I am correct, such findings will support two emergent narratives. First, that LGBTQ+ slurs – as a type of taboo language – demonstrate a general difference in processing to non-taboo language, but differ in this respect to other kinds of taboo language solely within the specific demographic they target. Second, that the linguistic reclamation of LGBTQ+ slurs does modulate the processing of such words among those who engage in or are more aware of this phenomenon.

I consider one of the main limitations of this study to have been that it only compared LGBTQ+ slurs to other taboo words which were not slurs. Unlike the lexical decision and free recall studies I describe in Chapters 5 and 6, which needed to be conducted as two separate studies due to the amount of time required to complete them, having started to run this study I now believe that this experiment could be expanded. Specifically, an experiment with a similar design could be conducted comparing LGBTQ+ slurs to non-slur taboo words, negative high-arousal words and negative low-arousal words, all within the same study. Not only would this enable more comparisons involving physiological measurements of arousal – including between the categories of words which are not slurs – it would also enable me to better compare arousal responses to the more reclaimed LGBTQ+ slurs to those of other kinds of emotional language. This is therefore a direction I recommend for future study.

Chapter 8 – General Discussion

At the very start of my thesis, I set out three key research questions (RQs). These were:

1. Is the cognitive processing of LGBTQ+ slurs similar to that of other kinds of taboo language?
2. Does the sociolinguistic phenomenon of linguistic reclamation modulate the cognitive processing of LGBTQ+ slurs?
3. What are the sociolinguistic implications of any effects of reclamation on the cognitive processing of LGBTQ+ slurs?

Over the course of my studies, new questions arose which proved essential to providing complete answers to RQ1 and RQ2, further illuminating what distinguishes the cognitive processing of slurs from other kinds of taboo language, and how the phenomenon of linguistic reclamation affects this process. Below, I identify these as sub-questions of RQ1 and RQ2. I do not present them in this way in order to introduce them as new research questions, as they only became clear to me following completion of my studies. Rather, I choose this means of presentation so as to illustrate which of my research questions they have helped to answer, and the order in which my discussion will address them:

1. Is the cognitive processing of LGBTQ+ slurs similar to that of other kinds of taboo language?
    a. What – if any – are the general differences in LGBTQ+ slur processing?
    b. How does gender affect responses to LGBTQ+ slurs in particular?
    c. How does sexual identity affect responses to LGBTQ+ slurs in particular?
2. Does the sociolinguistic phenomenon of linguistic reclamation modulate the cognitive processing of LGBTQ+ slurs?
    a. Which were the specific reclamation behaviours modulating the cognitive processing of slurs, and in what regards?
    b. What factors do these reclamation behaviours relate to, which drive them to modulate the cognitive processing of slurs?

Throughout my discussion, I will also highlight any findings which relate to RQ3, bringing these together to discuss the sociolinguistic significance of my research in Section 8.3. After providing comprehensive answers to all three of my research questions, in Section 8.4. I offer some reflections on the limitations of the research that I have conducted, and how I will develop my future research accordingly.

8.1. Answering RQ1: Comparing the Processing of LGBTQ+ Slurs to Other Kinds of Taboo Language

8.1.1. General Differences in LGBTQ+ Slur Processing

I begin my overall answer to RQ1 by comparing my general findings on the cognitive processing of LGBTQ+ slurs to other kinds of taboo language. This comparison concerns two issues: whether the processing of LGBTQ+ slurs behaves similarly to other taboo language when compared to non-taboo language, and whether there are any significant differences between the processing of LGBTQ+ slurs to other kinds of taboo language. To answer these questions, I focus on my lexical accessibility and recallability findings from Chapters 5 and 6.

Overall, the results of both experiments revealed an intriguing difference, based on the type of processing effects being measured. LGBTQ+ slurs did not exhibit any difference in lexical accessibility compared to non-taboo words; the literature (Geer & Bellard, 1996; Madan, Shafer, Chan, & Singhal, 2017; Thomas & LaBar, 2005; Williams & Evans, 1980) would suggest that lexical accessibility should be impaired for LGBTQ+ slurs compared to non-taboo words, if they behave as other taboo language does. However, there was also no difference in lexical accessibility found between LGBTQ+ slurs and other kinds of taboo words, providing an early indication that there is not a difference in taboo status driving this lack of expected difference.

By contrast, LGBTQ+ slurs demonstrated a significantly increased recallability compared to non-taboo words, which the literature evidences is typical of taboo language (Jay, Caldwell-Harris, & King, 2008; Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Madan et al. 2017). This indication that LGBTQ+ slurs behave similarly to other kinds of taboo language in terms of recallability was further supported by the finding that the two categories did not differ significantly in terms of recallability when directly compared.

It is a novel finding that a specific class of taboo language – in this case slurs – might only demonstrate one of the two most well-documented hallmarks of taboo language processing and not the other, given that most psycholinguistic studies of taboo language have not typically made such distinctions in their stimuli (e.g. Belleza, Greenwald, & Banaji, 1986; Geer & Bellard, 1996; MacKay, Shafto, Taylor, Marian, Abrams, & Dyer, 2004; Madan et al. 2017). This is clearly of note in regards to psycholinguistic research, but – as I argue throughout this chapter in response to RQ3 – a difference in unconscious, emotional responses to slurs being motivated by issues of identity, hegemony and marginalisation should be significant to sociolinguists as well (see section 8.3. for full discussion). Nevertheless, this overall contrast between lexical accessibility and recallability requires an explanation. For this, we ned to look more closely at both sets of results.

It is telling that in both of the lexical decision tasks described in Chapters 5 and 6, only participant gender and sexual identity predicted performance in either (Tables 5.4. and 6.4.). I discuss the specific impacts of both of these in sections 8.1.2. and 8.1.3., when answering RQs 1a. and 1b respectively. However, I wish to make a more general observation about this finding. Madan et al. (2017) suggest that a combination of non-emotional word properties (including word frequency, number of characters, number of syllables, familiarity, personal use, and age of acquisition) and tabooness best explained delayed lexical access for taboo words compared to non-taboo words. These are all factors which I measured in the studies described in Chapters 3 and 4, and subsequently controlled for in my lab experiments. Ultimately, none of these contributed significantly to or improved the lexical accessibility models in either experiment.

I do not argue that Madan et al.'s findings on lexical access are incorrect because my studies did not replicate them, nor doubt the validity of my own findings. Rather, I recall the discussion from the literature that impaired lexical accessibility for taboo language may be caused by a perceptual defence mechanism, and motivation to avoid the processing of particularly negative stimuli (Madan et al. 2017; Williams & Evans, 1980; see section 2.3.2. for full discussion). In the specific case of slurs, because of the harmful social power dynamics that particularly underpin slurs as a class of taboo language, I argue that relevant aspects of participants' own identities might override any effects of the properties that Madan et al.'s study highlighted in relation to taboo language generally. If true, this would help explain why category distinctions made on the basis of variables like tabooness and emotional arousal did not produce any overall differences.

Importantly, because neither a perceptual defence mechanism nor avoidance motivation have been linked to recallability in the literature, if this explanation of my lexical decision results is true then it should not be expected to hold for my free recall results. The results of my free recall analyses in both Chapters 5 and 6 confirm this: neither of my recallability models for all words included any aspect of participant identity.

Instead, when comparing LGBTQ+ slurs to non-taboo language, higher familiarity ratings and lower negativity ratings best explained higher recallability overall (Table 5.7.). When comparing LGBTQ+ slurs to taboo expressions that are not slurs, only higher familiarity ratings best explained higher recallability overall (Table 6.8.). These results do not entirely align with the literature, which suggests that emotional word properties – chiefly emotional arousal – provide the best explanations of the increased recallability of taboo words (e.g. Kensinger & Corkin, 2003; Madan et al. 2017; see sections 2.3.1. and 2.3.2. for full discussion).

However, I note that negativity – a different emotional property – did help to explain recall performance in the experiment described in Chapter 5. Negativity may have been more affected by the effects of reclamation that I discuss in section 8.2. of this chapter; the reclamation of slurs is able to have a particular effect on a slur's valence ratings, explained by its ability to give slurs positive uses and connotations, and evidenced by the inverse correlations between negativity ratings and reclamation behaviour ratings in Chapter 4 (Tables 4.6. and 4.7.). Similarly, in section 8.2.1. I discuss the relationship between reclamation and familiarity in more detail, reclamation in particular appears to have been significant to understanding why familiarity may have generally affected recall.

Briefly, I also note that the emotional arousal of LGBTQ+ slurs – and the impact of linguistic reclamation on the same – might be better assessed by physiological measures than self-reported ratings. This may be truer for arousal than any other emotional properties in my studies, as it is arguably the least conscious of the emotional responses I established normed ratings for. The pupillometry study I outline in Chapter 7 seeks to measure arousal in this way. Upon completion, the results of this experiment could help to explain why arousal might not have significantly contributed to my recallability models.

At this point of my discussion, then, it appears that the processing of LGBTQ+ slurs does not categorically differ from other kinds of taboo language. However, whether it is processed in the manner typically expected of taboo language seems to be task-dependent. For recallability, LGBTQ+ slurs display the typically enhanced recallability of taboo language in comparison to non-taboo language, but no difference in comparison to taboo expressions that are not slurs, although potentially as a result of different causes than those usually expected. Lexical accessibility paints an altogether different picture; in this respect, LGBTQ+ slurs do not appear to categorically differ to other kinds of taboo or non-taboo language at all. Instead, where slurs are concerned, the identity of the person experiencing such language appears to be the most important modulator of lexical accessibility. It is this factor to which I now turn my discussion.

8.1.2. The Effect of Gender on LGBTQ+ Slur Processing

A common theme through all of my studies was the relevance of gender to LGBTQ+ slurs, and to participant responses to the same. This is not altogether surprising, considering the community this language targets. Identities under the transgender umbrella are constructed around relationships between gender and the self, and arguably so too are sexual identities: these are often described in terms of gendered attraction (or lack thereof), some sexual identities and their corresponding labels

are implicitly gendered (e.g. 'lesbian'), and the upholding of what Cameron (2006, p. 178) describes as "institutionalised heterosexuality" – against which non-heterosexual identities are marginalised – is related to the prescription of normative gender relations and behaviours. Therefore, unlike some psycholinguistic studies which have presumed an impact of gender on language processing without providing clear reason, I argue that it is understandable that gender might affect responses to LGBTQ+ slurs.

What is not immediately clear is the manner in which gender affects the processing of LGBTQ+ slurs, and the extent to which it does. To answer these questions, I trace a narrative which developed over the course of my studies, namely the importance of how a participant's gender positions them in relation to LGBTQ+ slurs. I do this by examining what each of my studies revealed about this in turn: how LGBTQ+ slurs are themselves gendered (Chapter 3), how gender seems to shape emotional and non-emotional judgements of LGBTQ+ slurs (Chapter 4), and how gender is found to interact with the cognitive processing of LGBTQ+ slurs and taboo language more broadly (Chapters 5 and 6).

When analysing the results of the survey I conducted to identify LGBTQ+ slurs in British English (Chapter 3), I made the initial observation that there was a greater prevalence of slurs targeting men. This sometimes – but less typically – was part of the slur itself, with gendered words like 'boy', 'man' and 'lord' used as suffixes or as part of a two-word compound. More common was the tendency for slurs to index behaviours which are especially disparaged when performed by men, or those perceived as such. Over a quarter of the 160 slurs I identified referenced sex acts, and of these, just over 75% referenced participating in either anal sex or fellatio. Just under a fifth of all slurs identified referenced non-normative gender expression, with over 80% of these referencing the non-normative expression of femininity.

This finding should be of significant interest to language, gender and sexuality scholars in its own right, and I argued for a dual explanation of this prevalence of masculine-gendered LGBTQ+ slurs: that both hegemonic masculinity and institutionalised heterosexuality harshly punish men who are perceived as 'failing' to uphold them and relinquishing their power (Bersani, 1987; Cameron, 2006; Coates, 2011; Kiesling, 1998; Pascoe, 2005), and that women who do not conform to normative standards of gender or sexual expression are no less present in society, but have been subjected to greater erasure and invisibility (Chirrey, 2007; Turner, 2014). I discussed more detailed findings in relation to these points in section 3.5.2. of Chapter 3, but the most critical point for the present discussion is how this masculine bias in the gendering of LGBTQ+ slurs contextualises the gender differences I found in the three studies that followed.

In my LGBTQ+ slur norming survey (Chapter 4), I found an overall pattern of positive correlations between slur familiarity, personal use, and reclamation behaviours; and between arousal, negativity, offensiveness, tabooness and age of acquisition. Furthermore, all properties within one of these two groups correlated inversely with all those in the other group. Importantly, I found this pattern also manifested as a difference between gender groups: women consistently reported lower familiarity, personal use and reclamation behaviour ratings for LGBTQ+ slurs than men, but reported the same words as being more arousing, more negative, more offensive and more taboo.

I argued one possible explanation for these results may lie with a broader social expectation regarding gender, politeness, and taboo language. Earlier language and gender research suggested a particular burden on women to use forms of language which are felt to be more polite and indirect, which necessarily precludes taboo language like slurs (e.g. Brown, 1980; Coates, 1988; Fishman,

1980; Holmes, 1995). Furthermore, there exists at least one psycholinguistic study of taboo language processing which suggests that women provide slower lexical decision responses to all kinds of taboo language (Geer & Bellard, 1996), suggesting greater aversion and avoidance motivation. This could explain why women respondents to my survey generally reported reduced familiarity with and use of slurs, and more pronounced negative emotional reactions to them.

Wherever I have pointed to this explanation in my studies, I have acknowledged a very fair theoretical criticism, which I now engage with more closely as part of this general discussion. Particularly since the publication of Judith Butler's seminal work *Gender Trouble* (1990), language and gender theory has largely moved away from the suggestion that women's language – insofar as such a thing exists – is necessarily more polite, more supportive, or more submissive (e.g. Lakoff, 1973). This idea has largely been rejected regardless of whether it is attributed to women's 'deficit' (e.g. Jespersen, 1922) or men's social 'dominance' (e.g. Spender, 1985; West & Zimmerman, 1983), to borrow the terms used to describe its associated theoretical positions (Coates, 2016). Indeed, contemporary language and gender theory has generally rejected the entire notion of what Motschenbacher (2007, p. 256) calls 'genderlects' – "female and male speech styles as two stable, clear-cut and opposite gendered varieties" – as they have come to be considered the product (and legitimisation) of a heteronormative gender binary. Instead, it has shifted toward a view of language as being one tool which individuals can use to 'perform' gender moment to moment (Butler, 1990) – purposefully orienting themselves toward or away from particular gendered behaviours, some of which are indexed by language (Baker, 2008). This has more recently been termed the 'dynamic' or 'social constructionist' approach (Coates, 2016).

I entirely support this direction of travel, and do not mean to take a regressive stance by suggesting that there exists a blanket aversion to taboo language among women which can explain my findings. However, even the most contemporary arguments tend to acknowledge that where gendered speech norms do develop, they can be a result of pervasive, patriarchal and heteronormative socialisation from an early age, as suggested by some of the earlier theory (Tannen, 1990; Cameron, 2006). I argue that understanding that hegemonic power structures can exert gendered pressures on language use – which might in turn cause certain patterns of gendered language behaviour to emerge across a broader population – is not incompatible with understanding that these pressures will not be universally experienced and responded to in the same way by all individuals or members of a group. Indeed, I would only advise viewing such patterns as general population norms, and would not advocate assuming any particular response to them at the level of the individual. This is especially relevant to psycholinguistic research, including all of my own, which is often more concerned with establishing broader patterns. I therefore consider it reasonable that there might exist a particular pressure on women to avoid socially unacceptable language, which may have contributed to the trend in their responses to LGBTQ+ slurs.

However, this argument cannot explain another finding from my norming survey by itself. While the general pattern in women's ratings did hold true for some slurs which solely target LGBTQ+ women, the existence of any gender differences disappeared altogether for other slurs targeting the same group. There were no examples of the pattern in men's ratings disappearing for slurs targeting LGBTQ+ men. If a gendered response to tabooness is the only driver of gendered differences, it should not have mattered who the target of a slur was, as all the slurs in my norming survey were similarly taboo. To explain this, I refer back to one of my findings from Chapter 3: that a greater prevalence of LGBTQ+ slurs targeted men. I argue that an individual might judge a slur more harshly if they are not themselves positioned as one of its targets. This possibility was supported by

another finding from my norming survey which I discuss further in section 8.1.3.: the ratings of heterosexual respondents generally differed from those of non-heterosexual participants in the same manner that women's ratings generally differed from men's.

I do not intend to suggest that those not targeted by a slur are more affected by it, but rather that they may have less cause to doubt or challenge its tabooness, an idea which I return to in my later discussions of reclamation (section 8.2.). Those not targeted by particular slurs may also be more motivated to avoid being seen to accept or endorse them, either because they do not wish to be seen as contributing to its oppressive use, or for fear of social repercussion. Importantly, these issues may also motivate the responses of those targeted by a slur, but for these individuals, their association with and experience of it is arguably not a matter of personal choice.

While I support the idea that gendered expectations around taboo language use may well have contributed to women's more adverse ratings of the emotional and non-emotional properties of LGBTQ+ slurs, I consider the tendency toward LGBTQ+ slurs targeting men – and the way that gender positions women in relation to these slurs – a somewhat stronger explanation for my findings up to this point. This position was only reinforced by the findings of my two complete lab experiments, to which I now turn my discussion.

As with the general findings on LGBTQ+ slur processing, the presence of a gender effect in my experiments differed depending on task type: participant gender was a significant predictor of lexical decision response times, but was not associated with recall likelihood. To explain this, I refer back to the ideas of a perceptual defence mechanism and of avoidance motivation discussed in section 8.1.1., which the literature has linked to lexical accessibility but not to free recall. I argue that both explanations I have offered for gender differences in my data – a possible social pressure on women to avoid taboo language, and an effect of not being the target of slurs aimed at LGBTQ+ men – might prompt differences in approach/avoidance motivation, which in turn might produce gendered differences in lexical accessibility of LGBTQ+ slurs, but not recallability.

It is nevertheless surprising that I ultimately found no gender differences for recallability, considering that in my LGBTQ+ slur norming survey (Chapter 4), I did find gendered differences in familiarity (slurs were rated less familiar by women than by men) and negativity (slurs were rated more negatively by women than by men), both of which helped predict recall performance in my experiments. I also found gendered differences in arousal (slurs were rated more arousing by women than by men), which the literature heavily links to recallability, although the results of my experiments did not replicate this particular finding. This is challenging to explain, but since those gender differences manifested across a number of related properties (e.g. familiarity, personal use, and reclamation behaviours; arousal, negativity, offensiveness, tabooness and age of acquisition), but only specific properties (familiarity and negativity) were found to significantly predict recall performance in my experiments, those individual properties may simply have been more statistically reliable predictors of recallability than any effect(s) of participant gender. This is something that the experiment I describe in Chapter 7 may help to clarify once completed, as it will provide an potentially more accurate physiological measure of arousal, which could be compared by participant gender. Given how closely arousal has been linked to recallability, if a gendered difference in arousal is found in this experiment, the absence of both arousal and gender as predictors of recall likelihood may have been a consequence of how arousal was measured in my experiments.

With regard to my lexical decision results, the overall picture is complicated somewhat by the fact that gender did not produce the same effects in both experiments described in Chapters 5 and 6. In the experiment described in Chapter 5, women's lexical decision response times were

significantly slower than men's, while in the experiment described in Chapter 6, they were generally faster than men's. If women's perceptual defence or avoidance motivation was greater for either of the reasons I have suggested so far, women's response times should have been slower in both experiments. This is especially true in regards to being the target of a slur, as the same slurs were used in both experiments, which I have already demonstrated tended to target men more.

Assuming these results are not simply coincidence, or motivated by some other factor affecting either experiment's participant sample, there is only one fundamental difference between the two studies: the language that each was comparing. In the experiment described in Chapter 5, LGBTQ+ slurs were being compared to two categories of negative non-taboo language, which were either high or low in emotional arousal. In the experiment described in Chapter 6, LGBTQ+ slurs were being compared to other examples of taboo language which were not slurs. As discussed in 8.1.1., neither experiment found differences in response times based on these stimulus types, so they must be considered as a whole. However, I note that the greatest differences between women and men's response times were found to be for negative high-arousal words in the first experiment, and for LGBTQ+ slurs in the second.

In my discussions in both chapters, I made two observations in relation to this difference. The first is that if there is a social pressure on women to avoid the processing of taboo language, this might be more strongly felt when a greater number of non-taboo words are presented in contrast, and might be diminished or desensitised if all words are taboo and there is no such contrast. This could explain why women's response times were slower in the experiment described in Chapter 5 (where there was a contrast), but not in Chapter 6 (where there was not).

However, this did not explain why women's response times were actually faster in the experiment described in Chapter 6, as opposed to there being no gender difference at all. For this, I made an observation about my non-slur taboo words category. Although I had taken care to avoid the inclusion of slur words in this category, a number of these taboo words were also gendered, with a greater number of references to body parts associated with (cisgender) men than to (cisgender) women. Arguably, this made the imbalance of gendered references in the experiment described in Chapter 6 greater than the one described in Chapter 5. This could point to women's avoidance motivation having been reduced because they were not the targets of much of the language in the experiment.

This presented another problem: this explanation appears to contradict the evidence I found in Chapter 4: that more adverse ratings for LGBTQ+ slurs were given by the demographic groups who were not their targets. I assumed that because of this, avoidance motivation would be increased among women compared to men, and so should have been increased further rather than reduced when even more stimuli in the experiment were associated with men. The fault here perhaps lies in assuming that conscious versus unconscious responses to taboo language are necessarily the same. Those not targeted by slurs or referenced by taboo language may be more motivated to make conscious, negative judgements of that language. However, they may not be the more unconsciously, emotionally, and automatically affected by such language in terms of processing responses. That reaction may belong instead to the more targeted or referenced group.

To summarise my discussion in this section and provide an answer to RQ1a, I have demonstrated that LGBTQ+ slurs are themselves gendered (and often more so towards men or those perceived to be men), and that gender interacts with both conscious and unconscious responses to LGBTQ+ slurs. With regard to the unconscious processing of LGBTQ+ slurs, lexical accessibility appears to be affected by gender in a way that recallability is not, and I link this predominantly to

the effect of avoidance motivation on lexical decision. How and why gender interacts with these responses is complex and can appear contradictory, but I have suggested two main possibilities. First, that there may exist a socialised, patriarchal pressure on women to avoid taboo language, which may manifest across broader populations, and which may be more strongly experienced when non-taboo language is presented in contrast. Second, that being the target of a slur or otherwise taboo expression seems to affect responses to that language, although the nature of and reasons for these responses may differ depending on whether they are consciously or unconsciously given.

Having now provided a thorough explanation of why participant gender was relevant to the processing of LGBTQ+ slurs, I now move on to the other demographic factor that I found to significantly affect my findings: participant sexual identity.

8.1.3. The Effect of Sexual Identity on LGBTQ+ Slur Processing

Having evidenced that identity is a predictor of responses to slurs, it is not surprising that I found whether or not a participant is heterosexual – the sexual identity comparison that I made throughout my own studies – affects responses to LGBTQ+ slurs. I believe this requires less explanation than the immediate relevance of gender to LGBTQ+ slurs did, although I recall my early discussion of ingroups and outgroups – and their positions within the social intergroup status hierarchy – as being integral to understanding both taboo language and slurs in particular (see Chapter 2, sections 2.1.2. and 2.2.1. for discussion). It is clear that heterosexual and non-heterosexual groups are part of an ingroup/outgroup dynamic relevant to the existence of LGBTQ+ slurs. At this stage, I acknowledge that it is possible to identify as both heterosexual and LGBTQ+; some transgender people identify themselves as heterosexual. However, participants identifying as transgender or non-binary made up a very small portion of any of my study samples, and none also identified themselves as being heterosexual. Because of this, I don't believe this issue will have affected the comparisons I discuss in this section, although I would have preferred my samples to have better included and represented transgender people, particularly to gain a richer understanding of gendered differences and better investigate responses to transphobic slurs. This is something I am keen to develop in future work.

In many respects, my findings on the relevance of sexual identity to LGBTQ+ slurs and their processing mirrored my findings on gender. I consider this to strengthen the validity of the gender differences I found and support the explanations I gave for them; sexual identity is related both to issues of gender and strongly to LGBTQ+ slurs, so it is significant that the patterns of gender differences I found were replicated in the sexual differences in my studies.

Because of this, I consider the explanations for sexual identity differences in my studies to have been largely similar to those I gave for gender differences, so will not focus as much of my discussion on explaining these similarities as I do on explaining differences unique to my sexual identity findings. However, unlike with differences between women and men, I do not make the argument that responses were affected by any specifically heterosexual socialisation regarding norms of linguistic politeness and taboo language use. This is because there is no similar precedent for this in existing theory and studies. Despite this, I still argue that being the target of a slur significantly affects judgements of and responses to LGBTQ+ slurs, that this is particularly important for the lexical accessibility of slurs, and that this applies to differences I found concerning sexual identity.

In my slur norming survey (Chapter 4), I found that differences between heterosexual and non-heterosexual participants matched the differences I found between women and men: heterosexual participants provided lower familiarity, personal use, and reclamation behaviour ratings for LGBTQ+ slurs, but higher arousal, negativity, offensiveness, tabooness, and age of acquisition ratings. I consider this to have supported my argument that conscious judgements of slurs may be more adverse among the group which is not targeted by them, raising an interesting question for sociolinguists in particular about who most reinforces the taboos surrounding marginalised social groups and the language which targets them.

In the experiments described in Chapters 5 and 6, as with gender, I found that there were sexual identity differences in lexical decision response times, but not in performance in the free recall task. Alongside gender, sexual identity was the only other variable that contributed significantly to the best models of response times to all words in both experiments' lexical decision tasks. By contrast, neither gender nor sexual identity contributed significantly to the best models of recall likelihood in either experiment. The way these sexual identity differences in lexical decision manifested also showed a similar pattern to the gender differences: non-heterosexual participants' response times were slower when LGBTQ+ slurs were being compared to non-taboo words, but were faster when being compared to other taboo words which were not slurs. I argued that this supports my belief that being the target of a slur has an effect on avoidance motivation, which affects lexical accessibility in a way that it has not been found to affect recallability.

However, these findings required me to address another point of possible contradiction: women (who I suggested represented a less-targeted group because of the typical gendering of my LGBTQ+ slurs) provided slower responses in the experiment described in Chapter 5, but faster responses in the experiment described in Chapter 6. If these results were a consequence of just being the less- or non-targeted group, then it should have been heterosexual participants rather than non-heterosexual participants who displayed the same pattern of results for sexual identity comparisons, but the opposite was true. To resolve this, I argued that being the target of a slur may still have an effect on unconscious processing responses, but for different reasons for sexual identity compared to gender.

As I have already argued, I do not think there is any potential effect of socialised politeness norms for sexual identity affecting my results, as there is for gender. Additionally, the finding that the non-heterosexual group had faster response times in the experiment described in Chapter 6 (thereby evidencing less avoidance motivation) aligns well with my sexual identity comparison in Chapter 4: non-heterosexual participants provided higher familiarity, personal use, and reclamation behaviour ratings for LGBTQ+ slurs, but lower arousal, negativity, offensiveness, tabooness, and age of acquisition ratings. This finding not only helps explain why non-heterosexual participants displayed less avoidance motivation, but the literature has also linked higher familiarity, higher personal use and an earlier age of acquisition to improved lexical accessibility (Madan et al., 2017). It might have been these consequences of being a member of a slur's target group producing my sexual identity differences, different to those which produced my gender differences.

With respect to these results, I also considered why non-heterosexual participants produced slower response times in comparison to non-taboo words, rather than to other taboo words. Similar to an explanation I gave for my gender differences, I suggested that being members of the group specifically targeted by LGBTQ+ slurs might have had a more significant emotional impact on non-heterosexual participants when entirely non-taboo words were presented in contrast, as opposed to other words which were at least similarly socially unacceptable. Additionally, I did not overlook the

particular link between being non-heterosexual and reclaiming slurs targeting LGBTQ+ people, which I will return to when discussing the impact of reclamation on processing in section 8.2. of this discussion.

Having now discussed similarities between my gender and sexual identity results, there were some interesting differences that were specific to sexual identity. First, of all the demographic group differences I explored in my slur norming survey data (Chapter 4), most were sexual identity differences. Of these, over 70% were differences in one of the measures of reclamation behaviour in my study. This was not the case for either age or gender comparisons, which showed differences across a wider range of emotional and non-emotional properties, thereby strongly linking sexual identity and LGBTQ+ slur reclamation behaviour together. Again, I return to this finding in section 8.2. of this discussion, but argue at this stage that it highlights that this finding highlighted the importance of how an individual's identity positions them in relation to a slur, as unlike other properties I measured, the reclamation of slurs is closely linked to ingroup and outgroup membership.

Perhaps the most interesting sexual identity difference across all of my studies, though, was the finding that participant sexual identity was associated with one difference in recallability. In my discussion thus far and particularly in section 8.1.1., I have made it clear that there were no overall categorical differences between the recallability of LGBTQ+ slurs and other taboo words, nor were there any differences between heterosexual and non-heterosexual participants' performances in the recall task. This is still true, but only looking across my entire sample. In Chapter 6, I found that looking within just the non-heterosexual participant group, there was a statistically significant difference between recall likelihood of LGBTQ+ slurs and non-slur taboo words, with non-heterosexual participants being slightly likelier to successfully recall LGBTQ+ slurs. This was not the case within any other demographic group in the same experiment. Furthermore, in neither experiment were such categorical differences identified exclusively among women or men. This suggests that to the group most targeted by LGBTQ+ slurs, and only to that group, they were more memorable than other taboo words. Given I did not find LGBTQ+ slurs to generally be processed differently to other examples of taboo language in any other respect, this should be of significant interest to both psycholinguists and sociolinguists alike; it evidences that unconscious responses to slurs as a subclass of taboo language might be different to taboo language more generally, but specifically with respect to recallability, and only among their targets. It is this last finding which I argue sociolinguists should take particular note of, in that it appears to be issues of ingroup/outgroup relationships to slurs motivating these particular differences in unconscious responses. As I argued at the start of section 8.1.1., this is not a finding which traditional sociolinguistic approaches could reliably produce, but one which might contribute to the more conscious attitudes and practices which such research can and does reveal.

Identifying the reason for this is more challenging than identifying the difference itself, but I suggested several potential explanations. Both my models of recall likelihood for all words in the experiments described in Chapters 5 and 6 pointed to higher familiarity predicting higher recall likelihood (see Tables 5.6. and 6.6.). My norming results from Chapter 4 suggested that non-heterosexual respondents rated LGBTQ+ slurs higher for familiarity, and it makes sense that those targeted by a slur would have more experience and awareness of it. Similarly, both my models of recall likelihood for LGBTQ+ slurs specifically (see Tables 5.7. and 6.7.) pointed to decreased tabooness and some form of increased general-population reclamation behaviour (self-reclamation in Chapter 5, overall reclamation in Chapter 6) as predictors of increased recall likelihood. I discuss

the reclamation results further in section 8.2., but note here that as with the familiarity ratings, my norming results from Chapter 4 suggested that non-heterosexual respondents rated LGBTQ+ slurs lower for tabooness and higher for all reclamation behaviours. Although it did not appear as a significant predictor in any of my models, I would also like to consider the possibility of a sexuality-based difference in emotional arousal, considering how closely arousal has been linked to emotional language recall in the literature (see section 2.3.1. of Chapter 2 for discussion). Interestingly, if arousal is related then these results would suggest that non-heterosexual participants have greater arousal responses than heterosexual participants, which would not align with my results from Chapter 4. However, my own criticism of those arousal ratings – which I have discussed in this chapter and previously – is that they are conscious and self-reported ratings of an unconscious, automatic emotional response. This is therefore a finding that the experiment I describe in Chapter 7 would be extremely useful for clarifying once completed, by providing physiological measures of emotional arousal in response to LGBTQ+ slurs which could then be subject to demographic comparisons. These findings could help to explain one of the most interesting and unique findings from across my studies, and are therefore an added argument for using eye-tracking and pupillometry technology to explore emotional responses to LGBTQ+ slurs.

As a final point regarding my findings on identity, I raise a possibility which jointly concerns the results outlined in section 8.1.2. and this section. I speculate that if either of the effects of gender and/or sexual identity were sufficiently strong as to affect responses to the entire task, this might explain why there were no general differences between taboo and non-taboo lexical decision responses in Chapter 5, despite previous taboo language experiments suggesting that there should have been. This is not a possibility evidenced in the literature, as previous taboo language experiments have generally found differences in lexical decision responses between taboo and non-taboo language (e.g. Geer & Bellard, 1996; Madan et al., 2017; Thomas & LaBar, 2005; Williams & Evans, 1980), so have not had to contend with this problem. Furthermore, few have focused explicitly as I have on slurs, or considered the impact of relevant identity factors on responses as heavily. Nevertheless, it is worth further investigating whether the presence of LGBTQ+ slurs in the task could have had 'knock-on' effects on the other words in the lexical decision task detailed in Chapter 5, especially for those participants who are targeted by that language. I do not suggest this for the results of the lexical decision task described in Chapter 6, as the lack of stimulus type difference there is likelier due to there being no general difference in the processing of slurs and other kinds of taboo language, particularly since no overall stimulus type effect was found for that experiment's free recall task. Because of the lack of evidence from existing studies for this, I will need to conduct another lexical decision experiment in order to test this.

Rather than summarise of this specific section of my discussion, I will now bring together my general findings on LGBTQ+ slur processing with my findings on gender and sexual identity differences to provide an overall answer to my first research question.

8.1.4. An Answer to RQ1

My first research question asked whether the cognitive processing of LGBTQ+ slurs was similar to that of other kinds of taboo language, and if not, it asked how it differs. Synthesising all of my discussion up to this point, I am confident I can now answer this.

In terms of lexical accessibility, LGBTQ+ slurs generally do not differ from other examples of taboo language. Unusually, I found that when presented as the only examples of taboo language and

compared to non-taboo language, LGBTQ+ slurs did not differ from non-taboo language, as taboo language typically does. Though only a suggestion, this may have been because the presentation of LGBTQ+ slurs to a participant sample which included a large number of individuals targeted by such language was – for a number of potential reasons – disruptive enough to affect responses to all words across the task.

I also found that in contrast to any of the non-emotional word properties which have previously been found to explain taboo language performance in lexical decision tasks (e.g. word frequency, number of characters, number of syllables, familiarity, personal use, age of acquisition), only a combination of participant gender and sexual identity significantly predicted performance in either of my lexical decision tasks. How these differences manifested seemed to depend on what LGBTQ+ slurs were being compared to: women and non-heterosexual participants provided slower responses compared to men and heterosexual participants when LGBTQ+ slurs were being compared to non-taboo words, but faster ones when LGBTQ+ slurs were being compared to other taboo words which were not slurs. To explain the gender difference, I suggested that socialised and patriarchal politeness norms impacting on women, as well as the tendency of LGBTQ+ slurs to target men rather than women, may have produced a gendered difference in avoidance motivation affecting lexical accessibility. To explain the sexual identity difference, I argued that how sexual identity positions an individual in terms of being the target of an LGBTQ+ slur may have also affected avoidance motivation, and consequently performance in my lexical decision tasks. For both kinds of difference, I also pointed to the kind of language LGBTQ+ slurs were being compared to in both experiments as an important factor. This difference appears to have caused some inconsistencies regarding whether performance in the lexical decision task was improved or worsened for particular groups, but I provided explanations for these points of contrast to suggest that these results can be compatible with each other.

The picture for recallability is much clearer. When compared to non-taboo language, LGBTQ+ slurs behaved exactly as taboo language typically does: it was significantly likelier to be recalled. When compared to other examples of taboo language that were not slurs, there was no such categorical difference, again suggesting that slurs are not generally processed differently than taboo language more broadly. Unlike the focus on increased emotional arousal and tabooness in the literature, variables including including increased familiarity and general-population reclamation, and *decreased* negativity, tabooness and number of characters were identified as the best predictors of recall likelihood. Which were relevant depended on whether the analysis identified recall likelihood of all words in the task, or just the LGBTQ+ slurs. Unlike the lexical decision results, there were no recall differences relating to participant gender and sexual identity, and neither significantly contributed to my recall likelihood models. However, I found that non-heterosexual participants were the only group to evidence a statistically significant difference in recall likelihood between LGBTQ+ slurs and taboo words which were not slurs. I argued this could be attributed to differences I found in familiarity, tabooness and/or reclamation behaviour, but also suggested that further study could identify a potential effect of emotional arousal which my studies did not. Most importantly, I argued that this result indicates that for those who are directly targeted by a slur, their processing of such language may not perfectly replicate their processing of taboo language in general.

This is a comprehensive answer to my first research question. However, it is not yet a complete one. To fully address how the cognitive processing of LGBTQ+ slurs compares to taboo language more broadly, I now need to answer my second question, regarding the impact of linguistic reclamation on these processes. It is this question to which I now turn my discussion.

8.2. Answering RQ2: The Impact of Linguistic Reclamation on the Cognitive Processing of LGBTQ+ Slurs

8.2.1. General Insights on Reclamation and Its Effects on LGBTQ+ Slur Processing

I begin this section of my general discussion by exploring what my results revealed about the reclamation of LGBTQ+ slurs as a linguistic phenomenon, and the general impacts that reclamation seemed to have on the cognitive processing of LGBTQ+ slurs. I will then go on to identify the factors I believe underpin the effects of reclamation on slur processing, and note an interesting difference between effects of general-population versus individual reclamation behaviour.

Relevant to all of these discussions is a comment on how I approached the measurement of reclamation in my research, which I now briefly outline. Across all of my studies, I measured four kinds of behaviour relating to reclamation, with an understanding that individuals may experience reclamation as both a speaker and a listener, and in reference to both themselves and others (Baker, 2008; Livia & Hall, 1997; Stephens, 2011). To this end, I measured the extent to which individuals reported using slurs in reference to themselves (*self-self reclamation)* and others (*self-other reclamation),* or experienced other individuals doing the same (*other-self reclamation* and *other-other reclamation*). I also calculated three additional combined reclamation behaviours from these measurements: *self-reclamation* (mean self-self reclamation + mean self-other reclamation / 2), *other-reclamation* (mean other-self reclamation + mean other-other reclamation / 2) and *overall reclamation* (mean self-self reclamation + mean self-other reclamation + mean other-self reclamation + mean other-other reclamation / 4). I measured these combined reclamation behaviours so that I could make broader comparisons between self-referential and other-referential reclamation behaviours, as well as to an 'overall' measure of reclamation behaviour. Finally, in my experiments, I did not presume that the reclamation behaviours of my participants would necessarily match the general-population norms I had established, so made a distinction between these norms and participants' own reclamation ratings that I had asked them to provide. As will become clear throughout this section, making these distinctions and treating reclamation with this degree of complexity proved very valuable to my analyses, to the extent that I would argue this approach should be replicated by future studies. Existing sociolinguistic reclamation research tends to treat slurs as either reclaimed or not reclaimed (see section 2.2.2. for discussion). As reclamation scholars such as Brontsema (2004) and Herbert (2015) argue, this is problematic as it presumes general a consensus and fails to consider context in more detail. My approach therefore presents a novel alternative which may facilitate more sociolinguistic work in this area, given the new evidence it provides about the context of reclamation. Furthermore, I have shown that individuals' own engagement in reclamation does not necessarily depend on or reflect the degree of engagement they see in others, nor across a much broader population; this, too, has potential sociolinguistic implications because – as scholars like Brontsema and Herbert have argued – it emphasises reclamation as an ultimately personal choice that is not strongly influenced by the linguistic behaviour of others, and widespread reclamation as being a product of numerous individual choices, rather than a particular shared effort or group consensus.

The second question of my LGBTQ+ slur identification survey (Chapter 3) invited respondents to provide qualitative insights on the use of the slurs they had identified. This provided two significant insights on reclamation as a social phenomenon, which support and develop existing reclamation research. First, my participants appeared to join contemporary reclamation scholarship in rejecting the idea of slurs being wholly or completely reclaimed: many of their discussions of

particular slurs being reclaimed framed them as "becoming a reclaimed term", or as being "pretty reclaimed", "somewhat reclaimed", or subject to "partial reclaiming" (see Chapter 3, section 3.5.4.3. for additional discussion).

Second, just under 70% of further discussions of LGBTQ+ slur reclamation specifically referred to the reclamation of 'queer' (see Chapter 3, Tables 3.8. and 3.9.), suggesting that it may be a particularly salient and socially recognised example of an LGBTQ+ slur reclamation project. Participant discussions of the uses of 'queer' mirrored many of Bianchi's (2014) suggested purposes of reclamation: responses pointed to its use as a positive umbrella term to demarcate all gender and sexual minorities, and to its particular function as a non-derogatory label of fluid identities, as in the case of the word 'genderqueer'. However, some discussions of 'queer' also reflected the inter-community tensions surrounding reclamation identified by Brontsema (2004): one respondent expressed that they "vehemently disagree" with the reclamation of 'queer', arguing that from their perspective as an older person from a working-class background, the word is "still very much a slur". I argued that this comment particularly stressed the importance of considering multiple aspects of an individual's identity – and the intersections between these – in relation to their decisions to reclaim slurs.

To explore this line of discussion further, I consider findings from my slur norming survey (Chapter 4), which provided statistical evidence of various demographic differences in reclamation behaviour. Predictably, differences in reclamation made up the majority of differences in ratings between non-heterosexual and heterosexual respondents, with reclamation behaviour ratings always being higher in the non-heterosexual group. The reclamation behaviour which most commonly differed between the two sexual identity groups was other-self reclamation, i.e. being aware of other people using a slur positively to describe themselves. I suggested that this result might indicate that greater ingroup access is required to experience others reclaiming language in reference to themselves. While much contemporary sociolinguistic discussion of reclamation has focused on the importance of ingroup membership to an individual's own participation in reclamation projects – to the extent that ingroup membership is considered a prerequisite of reclamation (Bianchi, 2014; Herbert, 2015) – few studies have discussed how ingroup membership might shape exposure to the reclamation behaviour of others. For this reason, I consider this a novel finding of sociolinguistic interest.

I also observed that where there were significant differences between women and men in terms of reclamation behaviours, women consistently gave lower ratings than men. Interestingly, this was true regardless of the gendering of the slur itself; when a slur explicitly targeted LGBTQ+ women, there was either no gender difference in reclamation behaviour, or women still gave lower ratings than men. I attributed this finding to two potential explanations which I have discussed earlier in this chapter. First, as mentioned throughout section 8.1., all of the LGBTQ+ slurs identified in my research more frequently targeted men than women. Consequently, LGBTQ+ women may have felt less able to reclaim them and/or may have been less frequently exposed to their reclamation on account of their gender, and heterosexual women may have felt and experienced something similar on account of both their gender and sexual identity. I view this as further support for my general argument that whether an individual is the specific target of an LGBTQ+ slur affects their responses to it, and that this is particularly true for reclamation behaviour. The finding that women reported lower engagement in and experience of LGBTQ+ slur reclamation may also relate to my earlier argument about a greater social pressure on women to avoid taboo language

altogether, as avoiding ever using a slur is an alternative to reclaiming it, and may be a preferable strategy for women more generally if such a social pressure exists.

Collectively, I consider these findings to have broadly supported an understanding of LGBTQ+ slur reclamation projects as ongoing processes without clear or fixed completion points. In Chapters 2 and 3, I outlined how participation in reclamation is related to ingroup membership and issues of marginalisation, in the sense that being someone targeted by a particular slur is considered necessary to reclaim it *from* that experience, but willingness to engage in reclamation is not universal among all members of these groups. Importantly, the desire to reclaim a slur can be motivated – or indeed diminished – by the impacts of other intersecting aspects of identity, particularly age and social class in the case of LGBTQ+ slurs. I argue that these insights on reclamation from my earlier studies are a vital foundation for explaining the effects my experiments found of linguistic reclamation on the cognitive processing of LGBTQ+ slurs, to which I now turn my discussion.

In addition to the issues of sociolinguistic significance that they highlighted, my results offered a fairly straightforward answer to my second psycholinguistic research question: at least some aspect of reclamation does appear to have modulated the cognitive processing of LGBTQ+ slurs, be it lexical accessibility or recallability. This is evidenced by the four models in Chapters 5 and 6 which identified the best predictors of LGBTQ+ slur lexical accessibility and recallability, all of which included a measure of reclamation behaviour. In the experiment described in Chapter 5, improved LGBTQ+ slur lexical accessibility was best predicted by participants identifying as a man, identifying as heterosexual, and personally reporting higher overall reclamation ratings (see Table 5.7.). Improved LGBTQ+ slur recallability was best predicted by higher tabooness ratings, and higher general-population self-reclamation ratings from the norming study (see Table 5.9.). In the experiment described in Chapter 6, improved LGBTQ+ slur lexical accessibility was best predicted by participants identifying as a woman, identifying as non-heterosexual, and personally reporting higher overall reclamation ratings (see Table 6.7.). Improved LGBTQ+ slur recallability was best predicted by a lower number of characters, higher tabooness ratings, and higher general-population overall reclamation ratings from the norming study.

More complex are the questions of why these particular reclamation behaviours were involved, and most pressingly, why reclamation should have effects on processing at all. To answer this, I consider the second of the additional reclamation questions that emerged over the course of my research, regarding the variables affecting taboo language processing which I believe reclamation interacts with.

8.2.2. Interactions between Reclamation and Other Factors Affecting the Processing of LGBTQ+ Slurs

In this section, I focus less on the results of my lexical decision experiments. This is because both of my experiments suggested that it was demographic factors like gender and sexual identity which significantly predicted lexical decision performance. I have already discussed these results in sections 8.1.2. and 8.1.3., and it seems highly unlikely that the linguistic reclamation of slurs has any direct effect on a participants' gender or sexual identity. However, I acknowledged in 8.2.1. that the reverse is true: gender and sexual identity had effects on the reclamation ratings given in my LGBTQ+ slur norming survey. Specifically, it was participants' own overall reclamation ratings that helped predict lexical decision response times in both experiments, a point I will return to at the end

of this section. For now, I will instead focus on the interactions between reclamation and the factors which I found predicted LGBTQ+ slur recallability.

In Chapter 3, I observed that the slurs most frequently identified in the literature as being the focus of reclamation projects – 'dyke', 'fag', 'faggot', and 'queer' – were 4 of the 5 most frequently identified slurs out of all the 160 identified in my slur identification survey (see Table 3.1.). The fifth was 'tranny', but there is little evidence of this undergoing any reclamation, including from my own norming study which indicated most respondents had indicated never experiencing it being used positively by themselves or others. For the other most frequently-identified slurs though, this was an early indication that a link exists between the familiarity of a slur and its likelihood of undergoing a process of reclamation, or alternatively of successful reclamation projects increasing slur familiarity. I therefore begin this section of my discussion by exploring the relationship between reclamation, familiarity, and LGBTQ+ slur processing.

The connection between reclamation and familiarity was further confirmed by one of the correlation analyses from my slur norming survey data, as illustrated in Tables 4.6. and 4.7. of Chapter 4. Familiarity correlated positively with all of my reclamation behaviour ratings at the $p < .01$ level, and depending on the specific type of reclamation behaviour, the two ratings shared between 49% and 69% of the variance in their responses. The closest correlation was between familiarity and self-reclamation, suggesting that familiarity is especially linked to an individual's own, personal engagement in reclamation projects.

As I described in section 8.1.1., familiarity was a key predictor of the recall likelihood of all words in both of my free recall tasks. Familiarity was not a key predictor of the recall likelihood of just LGBTQ+ slurs in either experiment, but I note that it may have been replaced by reclamation behaviours here, with which it strongly correlates. Because familiarity is a property which is shared by slurs and non-slurs, but reclamation behaviours are not, familiarity may have proven the best means of indicating an effect of reclamation when comparing across all words in my experiments. Additionally, if reclamation strongly affected responses to either a quarter or a third of all experimental stimuli (depending on the experiment), it may have improved the relevance and predictive power of familiarity compared to other studies of taboo language recall. Finally, as I noted at the end of section 8.2.1., it was normed self-reclamation ratings which contributed to the recall likelihood model of LGBTQ+ slurs in Chapter 5, and this was the specific reclamation behaviour with the closest positive correlation with normed familiarity ratings. I consider this to further evidence a strong connection between familiarity and reclamation – particularly personal engagement in reclamation projects – and an important effect of this relationship on LGBTQ+ slur processing, particularly recallability.

My results cannot determine a direction of causality for this relationship, i.e. whether individuals are more likely to reclaim slurs they are more familiar with, or whether familiarity with a slur increases as a result of increased personal and collective reclamation. The latter seems logical, given that reclamation adds to the number of contexts in which a slur can be used and encountered, but both explanations are possible and are not mutually exclusive. My findings provide strong statistical evidence for an individual being more likely to participate in the reclamation of slurs that are more familiar to them, and this could be explained through additional sociolinguistic research. Particularly effective in this regard would be a piece of qualitative research – such as an interview or focus group study – posing the question of whether and how an individual's familiarity with a slur motivates their desire to reclaim it.

In addition to establishing a clear positive correlation between reclamation behaviours and slur familiarity, the results of my LGBTQ+ slur norming survey (Chapter 4) indicated that all measures of reclamation behaviour correlated positively with the personal use of slurs, and with all other reclamation behaviours. However, they all correlated inversely with arousal, negativity, offensiveness, tabooness and age of acquisition. In terms of the relevance of these findings to recallability, I focus specifically on the relationship between reclamation and tabooness.

In the experiment described in Chapter 5, I found that the recall likelihood of LGBTQ+ slurs was improved when both normed self-reclamation ratings and tabooness ratings were higher. The finding that increased tabooness led to increased recall is consistent with findings from the literature (e.g. Bertels, Kolinksy, & Morais, 2009; Madan et al. 2017). Considering this, though, the finding that self-reclamation ratings contributed to the same model was much less expected. This is because the results of my norming study (Chapter 4) indicated that self-reclamation was inversely correlated with properties typically found to boost recall performance, including arousal, offensiveness and tabooness. Similarly, because self-reclamation ratings and tabooness were inversely correlated, it seems unlikely that any individual slur would have received high ratings for both properties, even if such results are not impossible. However, the model did not require that a slur receive high ratings on both measures in order to demonstrate increased recallability; rather, it indicated that increases in these particular measures led to a higher chance of recall. Unlike with familiarity, where it is likely that reclamation did directly interact with that property, I therefore suggest more of an 'either/or' explanation for these results: both self-reclamation and tabooness independently helped to predict recallability of LGBTQ+ slurs, but an increase in one did not cause an increase in the other.

The results of the experiment described in Chapter 6 produced similar findings. I found that the recall likelihood of LGBTQ+ slurs was improved for these words when they contained fewer characters, and when their normed overall-reclamation ratings and tabooness ratings were higher. As with the experiment described in Chapter 5, the finding that increased tabooness led to increased recall was consistent with findings from the literature. The finding that fewer characters were associated with easier recall is unlikely to be related to reclamation, but is simply explained: shorter words use up less working memory to store and recall. While the finding was associated with normed overall reclamation ratings as opposed to just self-reclamation, this study again found that words more widely subject to reclamation showed improved recall performance. Again, this was inconsistent with the correlations identified by my norming study, but I suggest the same explanations for this finding as in the experiment before it.

In terms of why LGBTQ+ slur recall was affected by normed self-reclamation ratings in one experiment and normed overall reclamation ratings in the other, I note that overall reclamation ratings still incorporate both of the self-reclamation behaviours. In this sense, a specific impact of self-reclamation was present in the results of both lab experiments. However, I also noted that across both, overall reclamation – either normed or participants' own – was generally the measure of reclamation behaviour which most often affected both response time and recall likelihood. In this regard, it was more unusual that only self-reclamation was found to contribute to the model of LGBTQ+ slur recall likelihood in Chapter 5. It is possible that making comparisons between LGBTQ+ slurs and other kinds of taboo words prompted greater preoccupation with how *other* people would regard the language in the experiment, given that much more of the stimuli were taboo. This may have caused overall reclamation to be a better predictor of LGBTQ+ slur recallability in the experiment described in Chapter 6, while just self-reclamation was sufficient for predicting the same in the experiment described in Chapter 5.

As a final observation, I note a difference in the reclamation behaviours that predicted lexical accessibility as opposed to recallability. Participants' own reclamation behaviour ratings predicted lexical accessibility, while general-population reclamation norms appear to have predicted recallability. I argue this result aligns very neatly with the main factors that I believe to have to have most affected each kind of processing. If avoidance motivation is the primary factor impairing lexical accessibility, and reclamation lessens the negative emotional responses which drive avoidance motivation, it makes sense that participants' personal reclamation behaviour would have most affected their personal avoidance motivation responses to LGBTQ+ slurs. By contrast, if familiarity was one of the main factors which affected recallability, and widespread reclamation of a slur increases opportunity to become familiar with a slur, then it makes sense that general-population reclamation norms would have most affected participants' ability to recall LGBTQ+ slurs. I therefore stress how valuable it was that my lab experiments collected additional data on participants' own reclamation behaviour, rather than using only my normed ratings, as this observation could not have been made otherwise. Again, I would strongly advise that any future work on reclamation – especially if it is also quantitative and/or experimental – take a similar approach.

8.2.3. An Answer to RQ2

I argue that I have effectively demonstrated that linguistic reclamation modulates the cognitive processing of LGBTQ+ slurs, both with respect to lexical accessibility and to free recall. With this established, I argue that an acknowledgement of the effects of linguistic reclamation also be considered the last part of my answer to m\y first research question, regarding the differences between the cognitive processing of LGBTQ+ slurs and taboo language in general.

My studies have yielded three key insights regarding linguistic reclamation. First, that general population understandings of and debates around the reclamation of LGBTQ+ slurs reflect and support those evidenced in the existing literature. Second, that all measures of reclamation behaviour outlined in my research share a clear pattern of positive and inverse correlations with word properties proven relevant to the processing of taboo language, particularly with familiarity and tabooness in the case of slur recallability. Third, that regardless of whether LGBTQ+ slurs are being compared to other taboo or non-taboo words, an increase in at least one measure of reclamation behaviour predicts improved performance in both kinds of task. Self-reclamation and overall reclamation behaviours appear to have been particularly important, crucially suggesting that it is an individual's personal engagement in reclamation of an LGBTQ+ slur which modulates their processing of it. Finally, depending on the type of processing task being undertaken, either an individual's own reclamation behaviour (lexical access) or general-population reclamation norms (free recall) become more important, and this can be directly related to the other factors I found to determine responses to the task in question.

Having provided answers to my two research questions concerning slur processing, I now turn to my third, in which I discuss the specifically sociolinguistic implications of these psycholinguistic issues, and why making such a disciplinary distinction is perhaps unnecessary when asking questions to which both fields can contribute.

8.3. Answering RQ3: The Sociolinguistic Implications of My Findings

The questions asked here – of how LGBTQ+ slurs are cognitively processed and how linguistic reclamation affects this phenomenon – might appear at first glance to be purely

psycholinguistic ones. Certainly, the quantitative and experimental methods I have used to address these question are standard psycholinguistic approaches (with the possible exception of the slur identification survey in Chapter 3, which included some qualitative, free-response questions about attitudes to particular slurs). However, I conclude this chapter by arguing that my psycholinguistic findings can contribute to answering sociolinguistic questions, and that they provide evidence to support existing sociolinguistic arguments surrounding slurs and reclamation.

To make a case for the former, I first need to identify which sociolinguistic questions concern my research. Before doing so, I observe that all of the issues central to my research – taboos and taboo language, slurs, identity and linguistic reclamation – are issues of society, and the relationship of language to social attitudes and behaviours. Accordingly, it has become my view that none of my psycholinguistic findings could be understood without sociological and sociolinguistic perspectives, and indeed these have formed a significant portion of my literature review (Chapter 2) and have been drawn on throughout my discussion of my findings (Chapters 3 – 6). Because I have consistently drawn on sociolinguistic findings to answer psycholinguistic questions, I am equally motivated to demonstrate that the reverse can also be true.

The sociolinguistic questions which I believe my research helps to answer are as follows: how does engaging with linguistic reclamation – a sociolinguistic phenomenon – affect the way individuals respond to slurs, and what factors underlie participation in these reclamation projects? I argue that these are indeed sociolinguistic questions, because they concern what motivates marginalised individuals to take up and reframe language specifically intended to oppress them, and how doing so might consequentially change those individuals' relationships with and reactions to that language.

Critically, I also argue that these questions cannot be answered solely in regard to those factors which an individual is conscious of and able to articulate. Factors of which an individual might be less conscious or less able to identify reliably – such as their immediate emotional reactions to slurs – likely also contribute to an individual's motivation to reclaim those slurs, and may themselves by affected by their decision to do so. If we are to ask questions about how individuals use and are affected by language which specifically exists to enact a denial of rights and respect (McGowan, 2012; Herbert, 2015), we cannot only be interested in evidence which is able to be consciously articulated or discursively analysed. If there is also evidence of differences in unconscious responses to that language, shaped by our emotional reactions and experiences of it in the real world (and thus broader ideological norms), such responses surely contribute to our attitudes to slurs and our decisions to use them, positively or negatively. This evidence is not supplementary to traditional lines of sociolinguistic inquiry; it is part of the overall picture. However, as I have just suggested, these are difficult for traditional sociolinguistic studies – those involving surveys, interviews, focus groups, and/or discourse analysis, for example – to reliably identify. By contrast, psycholinguistic experimentation is specifically designed to measure largely unconscious responses to language, so can be used to explore these. I do not intend to argue that my psycholinguistic findings provide a complete answer to the sociolinguistic questions I have outlined, nor that either psycholinguistic or sociolinguistic evidence is more or less vital to answering them, but that a complete answer must include both.

Relating this argument to my own findings, I acknowledge that my findings on slur processing – the speed with which they are retrieved from the mental lexicon, and the ease with which they are recalled after being encountered – may not seem of concern to sociolinguistics. However, I have consistently argued and demonstrated throughout my thesis that these two processes are shaped by emotional responses to and social experience of this language. Any differences found in processing therefore reflect social structure and systems which inform those emotional responses and social experiences.

My lab studies (Chapters 5 and 6) found such differences. In both of these studies, I found that LGBTQ+ slur lexical accessibility was best predicted by a combination of participant gender, sexual identity and their own overall awareness of and participation in reclamation of the same slurs. My results not only suggested that reclamation appears to reduce the processing problems associated with greater emotional arousal responses and perceived tabooness, but that these emotional responses to LGBTQ+ slurs are seemingly affected by the relationships particular social groups have to this language. Similarly, both studies found that differences in perceived tabooness and either normed personal (Chapter 5) or overall (Chapter 6) reclamation were affecting the memorability of slurs, a process strongly associated with emotional arousal and tabooness responses in the literature (e.g. Bertels et al. 2009; Madan et al. 2017). This related back to results from my norming study (Chapter 4), which established clear relationships between gender, sexual identity, perceived tabooness, and reclamation behaviours. Compared to men, women reported LGBTQ+ slurs as being more strongly taboo and reported lower engagement in reclamation of them. Compared to non-heterosexual participants, heterosexual participants also reported LGBTQ+ slurs as being more strongly taboo and reported lower engagement in reclamation of them. Once complete, my pupillometry study (Chapter 7) will more directly measure group differences in emotional arousal responses to LGBTQ+ slurs, and how reclamation affects these. I consider all of these results evidence of differences in unconscious judgements of and emotional responses to LGBTQ+ slurs, which may motivate participation in slur reclamation projects, and in turn be altered by that engagement in reclamation. It is in this respect that I argue my findings help to answer the sociolinguistic questions I posed at the start of this section.

Answering these questions is not the only contribution to sociolinguistic work that my research has made. It has also provided a clear evidence base for the strong definitional focus of slurs – and, by extension, linguistic reclamation – on ingroups, outgroups, and intergroup power struggles. The positioning of these in contemporary work (e.g. Bianchi, 2014; Herbert, 2015; Jay, 2019b; Popa-Wyatt, 2016; Tirrell, 2012) as central to understanding slurs and reclamation has largely been based on theoretical and political perspectives, with which I agree, but feel have not been sufficiently evidenced in research as being essential to understanding slurs and reclamation. However, as I discussed throughout this chapter, my results provide reliable, empirical evidence that membership of relevant identity groups measurably changes attitudes to (Chapter 4) and responses to (Chapters 5 and 6) slurs. From these findings, my research has also *raised* new questions which sociolinguistic study could pursue and perhaps better explore than psycholinguistic experimentation, such as why slurs were generally rated as less familiar and more reproachable by the group(s) *not* targeted by them, and the apparent connection between being more familiar with a slur and being more likely to choose to engage in its reclamation.

Relatedly, I argue that any perspective on what motivates engagement in reclamation projects should be informed by the links my research reveals between whether an individual is a member of a group targeted by a slur, their unconscious emotional responses to that slur, and whether they participate in its reclamation. I am not convinced that what motivates individuals to take up particular positions on Brontsema's (2004) model of attitudes to reclamation (see Chapter 2, section 2.2.2.) is solely a conscious and political judgement of the importance and immutability and of a word's history, and of the productiveness of attempting to change this. If it can be demonstrated – as I believe my research has managed to – that an individual's position as the target of a slur (or not) can predict some of their unconscious emotional responses to it, then it is also likely to impact their actual *use* of those slurs. In this way, my findings support existing ones from sociolinguistic research like Brontsema's (2004), concerning attitudes towards reclamation. In turn, it is significant that my research also suggests that these emotional responses can themselves be altered by an individual's subsequent decision to engage in reclamation or not.

Underscoring all of my discussion in this section is a firm conclusion, reached over the course of conducting this interdisciplinary research, that it should be important to sociolinguistics that participation in a fundamentally sociolinguistic phenomenon can transform both conscious *and* unconscious responses to language mired in issues of systemic power, social marginalisation, and hegemony. Similarly, it should be recognised that it is not through sociolinguistic study alone that these responses are revealed, nor questions regarding their nature, impact on attitudes to slurs, and effect on reclamation activity be given answers.

I have now discussed answers to all three of my key research questions. As a conclusion this general discussion of my findings, I will now consider some of the limitations of this piece of research, my ideas for resolving them, and some directions for future work – both psycholinguistic and sociolinguistic – illuminated by my studies.

8.4. Limitations and Future Directions

To begin discussion of the limitations of my research, I first consider the overall reliance of my research on self-reported measures of emotional arousal. This is a standard approach taken in many taboo language processing studies; was suitable for the reporting of more conscious attitudes and behaviours relevant to taboo language processing; and was better suited to the time and resource constraints of the research. However, the literature suggests that emotional arousal is one of the most relevant modulators of taboo language processing, and that arousal responses tend to be automatic, unconscious and physiological. Therefore, physiological measures of arousal may be more suitable for identifying its effects. Accordingly, I will begin addressing this issue by completing data collection and analysis for the pupillometry experiment that I described in Chapter 7. In this chapter, I discussed in detail the advantages of pupillometry over other measures of arousal, both self-reported and physiological. Additionally, throughout Chapters 7 and this chapter, I explained how the results of this experiment could be used to confirm or clarify earlier findings from my research. Primarily, they will help to establish whether arousal responses to LGBTQ+ slurs differ from other kinds of taboo language – both in general and within specific demographic groups – using a physiological rather than self-reported measure. This, in turn, will help clarify why LGBTQ+ slurs did not generally differ from other kinds of taboo language in terms of either recallability or lexical accessibility, but also why there may have been differences between the two conditions within specific demographic groups.

However, I now realise a need to develop the pupillometry study further, to also make comparisons to non-taboo words which differ in degree of arousal. The findings of the lab experiments described in Chapters 5 and 6 suggested that comparing LGBTQ+ slurs to taboo words changed the direction of any differences that had been observed when comparing them to non-taboo words. The pupillometry experiment I describe in Chapter 7 only compares LGBTQ+ slurs to non-slur taboo words, but it could be expanded to make comparisons to non-taboo language within the same study. This may help to explain some of the points of contrast in the results of my two completed language processing experiments.

Regarding a second limitation, the results of the studies I describe in Chapters 4 to 6 consistently highlighted that both participant gender and sexual identity had significant effects on ratings of LGBTQ+ slurs and on both aspects of their processing. However, the results of my slur identification survey (Chapter 3) indicated that significantly more LGBTQ+ slurs predominantly target men, which I considered may have affected both men's and women's responses. Additionally, while all of my studies were generally well-balanced in terms of heterosexual and non-heterosexual participants, all tended to include more women than men. Now that I understand the impact of gender and sexual identity on the processing of LGBTQ+ slurs, I will make two adjustments to my

future research in order to explore these effects more accurately. First, that the gendering of slurs used as stimuli in such experiments is balanced, even if this may come at the cost of controlling for word frequency. Second, that participant samples contain balanced groups of both heterosexual and non-heterosexual women and men, so that these experiments are not only able to make balanced comparisons on both individual dimensions, but can also explore issues when these factors are combined (e.g. differences between non-heterosexual and heterosexual women). Relatedly, I would also prefer any future samples to include a better balance of cisgender and transgender participants, although this may require its own experiment to sample and explore effectively.

I would also like to acknowledge the difficulties presented to my research by attempting to combine lexical decision and free recall tasks into one experiment. While the results of my free recall tasks generally aligned well with existing literature and I was able to provide some clear explanations, the results of my lexical decision tasks were less clear and often appeared to contradict each other across experiments. They did not suggest that slurs demonstrate the impaired lexical accessibility typically seen when comparing taboo language to non-taboo language, and suggested that while response times were slower among women and non-heterosexual participants when slurs were compared to non-taboo language, they were faster in the same groups when slurs were compared to other examples of taboo language.

Due to time and resource constraints, it was necessary to conduct two separate experiments so that both a lexical decision and free recall task could be included in each. However, this meant that in order to gain a full picture of how LGBTQ+ slurs compared to taboo language more generally, I was combining results across two separate experiments with two separate participant samples. Given that response time studies often include significant variation between participants even within the same sample, I suspect that this decision may have particularly affected my lexical decision results, and made them harder to bring together. Given that my free recall results did not evidence the same problem, I could now construct an additional experiment solely involving a lexical decision task, which would include all categories of linguistic stimuli used in this research within one single experiment, and may be able to generate clearer comparisons to further explain my results.

Finally, this research has suggested an intriguing direction for a future piece of qualitative, sociolinguistic work. Results from the studies described in Chapters 4 to 6 suggested that both word familiarity and personal engagement in reclamation have important effects on the cognitive processing of LGBTQ+ slurs, but critically, these two factors appear to be closely related to each other as well. What my research could not discern is the nature of this relationship: whether people are more likely to reclaim slurs they are more familiar with, are more familiar with slurs as a result of their reclamation, or a combination of both. This would be well-suited to an interview or focus-group study, which could use my database of normed word property and reclamation behaviour ratings (Appendix 4) as a starting point. Better understanding this relationship would help elucidate the role which word familiarity and reclamation behaviour both have on the processing of LGBTQ+ slurs.

Chapter 9 – Conclusion

Having now provided answers to the three research questions that I set out in Chapter 1, I conclude my thesis with some final reflections on the research and recommendations for future work. I begin by noting the significance and original contributions of my research to the literature on taboo language processing and linguistic reclamation. Related to this discussion, I follow with my reflections on successfully conducting a piece of interdisciplinary psycholinguistic and sociolinguistic research. Finally, I outline some of the potential applications of my findings with a view to generating real-world impact.

9.1. Original Contributions of the Research

My research has made many original contributions to the field. First, by conducting the only taboo language processing study to focus explicitly on slurs as a type of taboo language, I have applied psycholinguistic study to understanding an especially contentious and socially significant type of language. Throughout my thesis, I have presented a view of slurs as language with a unique purpose of and potential for causing social harm and reinforcing oppressive power structures. Accordingly, I have argued the need for language processing research which explores the cognitive and emotional effects of slurs in particular, as opposed to simply considering them examples of taboo language more generally. While my findings have indicated that slurs are not generally distinct from other kinds of taboo language in terms of their cognitive language processing, there are respects in which they do appear to differ, such as evidencing even greater recallability than other kinds of taboo language within the specific group(s) that they target.

Second, by virtue of being the only taboo language processing research to have focused explicitly on slurs, mine is also the only taboo language processing study to have explored potential effects of linguistic reclamation on this phenomenon. I consider this to be a particularly significant contribution to knowledge, not just in terms of what this teaches us about the effects of reclamation on slur processing, but the new body of evidence this adds to discussions about the motivations for and effects of reclamation as a sociolinguistic phenomenon. My results present clear evidence that reclamation affects slur processing, both in terms of lexical accessibility and recallability; greater engagement with the reclamation of a slur is associated with improved performance in tasks measuring both. Results also suggest that depending on the type of processing effect being investigated, either an individual's own reclamation behaviours (for lexical accessibility) or broader-population reclamation trends (for recallability) are significant predictors of responses.

Third, my research also links reclamation to factors which may explain or contribute to slur processing effects. One of these is a relationship between reclamation, identity and processing: participant gender, sexual identity, and reclamation behaviour were collectively the best predictors of the lexical accessibility of LGBTQ+ slurs, but whether responses were faster or slower for particular groups depended on whether slurs were being compared to other taboo words or non-taboo words. Another set of findings established relationships between reclamation, word familiarity, perceived negativity, and tabooness judgements, all of which contributed to recallability effects. In turn, I argued these results contribute both a clear evidence base for the sociolinguistic focus on hegemony and oppression in defining both slurs and reclamation, as well as indications of unconscious emotional responses relating to reclamation that are currently absent from sociolinguistic discussion of the phenomenon.

Fourth, I have also made very valuable contributions to research in terms of the amount and quality of the data that I have generated. As part of my research, I have identified 160 contemporary British English LGBTQ+ slurs and provided a quantifiable measure of their frequency of use (Chapter

3, Appendix 2). This is over twenty times the number of LGBTQ+ slurs identified in the existing norming data that I was able to draw from in my own research, and I am pleased to have identified a significantly more representative sample of such words. By itself, this data also yielded some interesting findings on the concepts typically indexed by LGBTQ+ slurs, particularly taboo sexual behaviour, perceived abnormality in gender expression, and ideas of deviance and misconduct. These results also further evidenced the importance of community membership, intersectional differences, and importance of intent to attitudes toward slur use and slur reclamation.

Fifth, I have also produced what I understand to be the first database of normed word property and reclamation behaviour ratings for LGBTQ+ slurs (Chapter 4, Appendix 4). Average ratings for every LGBTQ+ slur were generated from just under 35,000 total ratings, making them a highly robust product of a significant amount of research work. I fully intend to publish this data as open-access and freely available, so that it can be used as the foundation for any future research seeking quantitative data on the frequencies, perceptions of and reclamation of LGBTQ+ slurs. My norming data also identified a clear correlation pattern between particular word properties and reclamation behaviours for LGBTQ+ slurs (see Tables 4.6. and 4.7., Appendix 5), patterns which it also found to manifest within particular age, gender, and sexual identity group differences in ratings. These results were essential to explaining the findings of my lab studies, and also helped explain identity-based differences to slurs and reclamation, which should be highly informative for any future experimental research centred on slur processing and/or linguistic reclamation.

There is one final respect in which I consider my research to make a significantly original contribution: its status as a piece of interdisciplinary psycholinguistic and sociolinguistic research. Below, I provide a more dedicated discussion and reflection on this aspect.

9.2. Reflections on Conducting Interdisciplinary Psycholinguistic and Sociolinguistic Research

I consider my demonstration that interdisciplinary psycholinguistic and sociolinguistic research is both possible and valuable to be a significant success of my research as a whole. In Chapter 1, I noted that this has not been frequently attempted, and indeed that the two disciplines may be considered difficult to bring together. This is due to psycholinguistic experimentation often requiring any variability arising from highly personal or contextual uses of language to be controlled or eliminated, while much contemporary sociolinguistic research – especially that which investigates issues of language and identity – rejects the strict categorisations and generalisations which are often required when conducting experimental research.

I made a commitment from the outset of my research to avoid imposing potentially false identity categories and assumptions onto my participants as a consequence – however unintentional – of creating the experimental control required by my methods. I provided theoretical, methodological and political justifications for taking this view, and highlighted that allowing participants to signal their own identities and behaviours in my research – as has been done in variationist sociolinguistic work – was a potential solution to this problem.

While resolving this interdisciplinary tension was initially a hurdle, I now reflect that making these adjustments was relatively straightforward, and indeed strongly benefitted the research. I achieved this in a number of ways: allowing my participants to identify the LGBTQ+ slurs that I would later use in my experiments; allowing participants to entirely self-identify their gender and sexual identity; and measuring several different reclamation behaviours participant-by-participant, rather than assuming a certain experience of reclamation based on participant identity.

Regarding the first of these steps, my research was stronger for identifying a large number of *participant-generated* LGBTQ+ slurs and associated data. This is not well-evidenced in the taboo

language processing literature I had explored, and was motivated instead by my readings of sociolinguistic studies on slurs. Had I simply self-selected LGBTQ+ slurs to collect data for and experiment with, I would have learned far less about the variety of slurs that exist, and likely arrived at far more limited experimental findings.

Regarding the second step, some degree of participant categorisation was a necessary part of my research. However, actively enabling all participants to self-identify and report their own linguistic behaviours for themselves meant that all such categories were established based on how participants had chosen to represent themselves. Similarly, nothing about their language practices was presumed based on those categories.

This last step concerning the measurement of reclamation proved particularly important, considering the effects of individual versus general-population reclamation behaviours that I found on different kinds of processing (see section 9.1.). I would not have been able to identify these at all if a) I had assumed individuals' reclamation behaviour based on their identity, nor b) I had assumed individuals' reclamation behaviour would necessarily align with broader demographic group and general-population norms.

While it is true that all of these steps increased the amount of work I had to do as a researcher, and somewhat lessened my own control over the research design, these sacrifices were worth what my research consequently gained in accuracy, rigour, and representation. Overall, I strongly recommend that more psycholinguistic research take a similar approach, especially if dealing with any aspect of social identity, as it may in fact produce more accurate and representative results. Similarly, I recommend that more sociolinguistic issues – particularly those relating to the emotional effects of oppressive language – be explored using experimental methods, as doing so affords valuable insights and helps to answer these important questions in ways which cannot be achieved via other methods.

9.3. Opportunities for Impact

In sections 9.1. and 9.2., I outlined the significant contributions my research has made to the study of taboo language processing, slurs, and linguistic reclamation, as well as to the broader issue of approaching interdisciplinary psycholinguistic and sociolinguistic research. However, these are not the only ways in which I consider my research to have potential for impact. Here, I end my thesis by briefly considering some of the real-world issues my research can help to address.

One area is organisational policy-making. Having personal experience in carrying out research and policy work in the third sector, I am aware of the emerging desire across many organisations and institutions to allow individuals to self-identify attributes like their gender and sexual identity. This is a progressive policy shift, allowing for the representation of a broader spectrum of identity and the language used to describe it. Such a policy improves inclusion and accessibility, as in the case of someone whose gender is non-binary being asked to name their gender when accessing a healthcare service. It also helps to improve the scope and rigour of organisational research, as in the case of a healthcare organisation seeking to learn more about patients who do not identify their gender using terms like 'man' or 'woman'. It is in this regard that questions about sexuality and gender identity were recently added to the UK National Census in 2021, with an option for self-identification included.

However, a policy problem is presented for organisations when individuals self-identify using language considered in many contexts to be offensive or derogatory, as is the case when reclaiming language. What should a medical practice do if a patient identifies their sexual identity as 'dyke' in a response to a healthcare questionnaire? How should the Office of National Statistics respond to seeing 'fag' among self-identified responses to the national census? Such instances may be cases of

positive and political reclamation, but can provoke doubt and uncertainty about an individual's motivations for describing themselves in this way, often without the possibility of further enquiry. In particular, there can be anxiety about whether allowing these kinds of self-identification would constitute an organisational/institutional endorsement of that language in all contexts in which it can be used. In the case of LGBTQ+ slurs, my research offers clear findings in terms of which slurs are becoming more widely subjected to reclamation efforts, and how engagement in their reclamation is demonstrably linked to less hostile reactions. Both of these results support organisations taking evidence-based, good-faith approaches to the reclamation of slurs by members of the public. I therefore intend to use both of these findings to support wider organisational encouragement and inclusion of reclamation efforts, in support of individuals who respond to oppressive language in this manner.

Another area for impact concerns the public debate surrounding reclamation. As I have discussed throughout my thesis, the literature on reclamation strongly suggests it is a contentious topic even among those able to engage in it, with several competing (and perhaps incompatible) attitudes toward it as a political action. Also prevalent is a false dichotomy of slurs as either being fully reclaimed and universally accepted, or not reclaimed at all and perhaps even unreclaimable. In recent years, social media has further exacerbated divisions and misrepresentations of linguistic reclamation and those who do (and do not) engage in it. What has been somewhat lost in these debates is a recognition that choosing to reclaim a slur is fundamentally a personal choice about how to respond to social harm and exclusion. As my results have shown, when reclamation projects are engaged with, their effects on our responses to the slurs which target us can be transformative, even in ways we might not be conscious of. Evidence that engaging in reclamation can demonstrably change the way targeted individuals react to slurs may aid understanding of why such individuals choose to reclamation as a response to oppression to begin with. Equally, the same evidence would suggest that those who do not engage in reclamation do not experience these effects, but still experience the same harms caused by such language. Among those who do engage in reclamation, this might help to reduce the expectation that all marginalised people are equally comfortable hearing and using the language used to oppress them, even within their own communities. Therefore, I consider my findings to offer some unique insights to reduce tensions and improve understanding on both sides of this important issue.

References

Aboyoun, D. C., & Dabbs, Jr., J. M. (1998). The Hess pupil dilation findings: Sex or novelty? *Social Behavior and Personality: An international journal*, *26*(4), 412-420. doi: 10.2224/sbp.1998.26.4.415

Agresti, A. (2007). *An introduction to categorical data analysis.* Hoboken, NJ: Wiley.

Allan, K. (2019). Taboo Words and Language: An Overview. In K. Allan (Ed.), *The Oxford Book of Taboo Words and Language* (pp.1-27). Oxford: Oxford University Press.

Allan, K., & Burridge, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge: Cambridge University Press.

Anderson, A. K., & Phelps, A. E. (2002). Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *Journal of Cognitive Neuroscience, 14*(5), 709-720. doi: 10.1162/08989290260138618

Antonsen, A. N., Zdaniuk, B., Yule, M., & Brotto, L. A. (2020). Ace and Aro: Understanding Differences in Romantic Attractions Among Persons Identifying as Asexual. *Archives of Sexual Behavior*, 49*,* 1615-1630. doi: 10.1007/s10508-019-01600-1

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, 38(11), 1145–1160. doi: 10.1037/0003-066X.38.11.1145

Baker, P. (2002). *Fantabulosa: A Dictionary of Polari and Gay Slang.* London: Continuum.

Baker, P. (2005). *Public Discourses of Gay Men*. London: Routledge.

Baker, P. (2008). *Sexed texts: Language, Gender and Sexuality*. London: Equinox.

Baker, P. (2019). *Fabulosa! The Story of Polari, Britain's Secret Language*. London: Reaktion Books.

Bar-On, B. (1993). Marginality and epistemic privilege. In L. Alcoff & E. Potter (Eds.), *Feminist Epistemolo*gies (pp. 83-100). London: Routledge.

Bayer, M., Sommer, W., & Schacht, A. (2010). Reading emotional words within sentences: The impact of arousal and valence on event-related potentials. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology, 78*, 299-307. doi: 10.1016/j.ijpsycho.2010.09.004

Bayer, M., Sommer, W., & Schacht, A. (2011) Emotional words impact the mind but not the body: evidence from pupillary responses. Psychophysiology, *48*(11) ,1554-1562. doi: 10.1111/j.1469-8986.2011.01219.x.

Bayer, M., Sommer, W., & Schacht, A. (2012). Font Size Matters—Emotion and Attention in Cortical Responses to Written Words. *PloS one, 7*(5), e36042. doi: 10.1371/journal.pone.0036042

Becker, H. (1963). *Outsiders: Studies in the Sociology of Deviance*. New York: The Free Press.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments & Computers, 18*(3), 299-303. doi: 10.3758/BF03204403

Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA.: Sage Publications.

Bersani, L. (1987). Is the Rectum a Grave? *October,* 43, 197. doi: 10.2307/3397574

Bertels, J., Kolinsky, R., & Morais, J. (2009). Norms of Emotional Valence, Arousal, Threat Value and Shock Value for 80 Spoken French Words: Comparison Between Neutral and Emotional Tones of Voice. *Psychologica Belgica, 49*, 19-40. doi: 10.5334/pb-49-1-19

Besnier, N. (1990). Language and affect. *Annual Review of Anthropology, 19,* 419-451. doi: https://doi.org/10.1146/annurev.an.19.100190.002223

Besnier, N., & Philips, S. U. (2014). Ethnographic Methods for Language and Gender Research. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The Handbook of Language, Gender and Sexuality* (2nd ed.) (pp.123-140). Chichester: John Wiley and Sons.

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics, 66*, 35-44. doi: 10.1016/j.pragma.2014.02.009

Box, G. E. P., & Tidwell, P. W. (1962) Transformation of the Independent Variables. *Technometrics, 4*(4), 531-550. doi: 10.1080/00401706.1962.10490038

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms For English Words (ANEW): Instruction Manual And Affective Ratings (Technical Report C-1)*. Gainesville: University of Florida.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology.* 45(4), 602-7. doi: 10.1111/j.1469-8986.2008.00654.x

Brandes, S.H. (2019). An anthropological approach to taboo words and language. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.372-390). Oxford: Oxford University Press.

Braun, N., Goudbeek, M., & Krahmer, E. (2020). Affective Words and the Company They Keep: Studying the accuracy of affective word lists in determining sentence and word valence in a domain-specific corpus. *IEEE Transactions on Affective Computing.* Early access, 1-12. doi: 10.1109/TAFFC.2020.3005613

Brontsema, R. (2004). A Queer Revolution: Reconceptualizing the Debate Over Linguistic Reclamation. *Colorado Research in Linguistics, 17*. doi: 10.25810/dky3-zq57

Brooke, J., Wang, T., & Hirst, G. (2011). Predicting Word Clipping with Latent Semantic Analysis. *Proceedings of the 5th International Joint Conference on Natural Language Processing*. 1392-1396.

Brown, P. (1980). How and why are women more polite: Some evidence from a Mayan community. In S. McConnell-Ginet, R. Borker & N. Furman (Eds.), *Women and language in literature and society,* 111-136. Santa Barbara, CA: Praeger Publishers.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*(1), 73 - 99. doi: 10.1016/0010-0277(77)90018-X

Brysbaert, M., & Cortese, M. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly journal of experimental psychology, 64*, 545-559. doi: 10.1080/17470218.2010.503374

Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *International Journal of Psychophysiology, 61*(1), 26-33. doi: 10.1016/j.ijpsycho.2005.10.022

Bucholtz, M. (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics, 7*(3), 398-416. doi: 10.1111/1467-9481.00232

Burnard, L. (2007). *Reference Guide for the British National Corpus (XML Edition)*. Retrieved from http://www.natcorp.ox.ac.uk/XMLedition/URG

Burridge, K., & Benczes, R. (2019). Taboo as a driver of language change. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.180-198). Oxford: Oxford University Press.

Burridge, K. (2010). Linguistic cleanliness is next to godliness: Taboo and purism. *English Today, 26*, 3-13. doi: 10.1017/S0266078410000027

Butler, J. (1990). *Gender Trouble*. London: Routledge.

Bzdok, D., Laird, A. R., Zilles, K., Fox, P. T., & Eickhoff, S. B. (2013). An investigation of the structural, connectional, and functional subspecialization in the human amygdala. *Human Brain Mapping, 34*(12), 3247-3266. doi: 10.1002/hbm.22138

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature, 371*(6499), 702-704. doi: 10.1038/371702a0

Callahan, M. P. & Zukowski, K. T. (2019). Reactions to transgender women and men in public restrooms: Correlates and gender differences. *Journal of Homosexuality*, 66(1), 117-138. doi: 10.1080/00918369.2017.1395661

Cameron, D. (1990). De-mythologizing sociolinguistics: Why language does not reflect society. In J. Joseph and T. Taylor (Eds.), *Ideologies of Language* (pp. 79-93). London: Routledge.

Cameron, D. (1992). Naming of Parts: Gender, Culture and Terms of the Penis among American College Students. *American Speech, 67(*4), 367-382. doi: 10.2307/455846

Cameron, D. (2006). *On Language and Sexual Politics*: London: Routledge.

Chambers, J. K. (2004). Studying Language Variation: An Informal Epistemology. In J.K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The Handbook of Language, Variation and Change* (pp.3-14). Oxford: Blackwell Publishing.

Chen, A. H. Y. (1998). The philosophy of language rights. *Language Sciences, 20*(1), 45-54. doi: 10.1016/S0388-0001(97)00010-7

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics, 15*(2), 151-196. doi: 10.1111/j.1467-9841.2011.00478.x

Chetail, F. (2014). Effect of number of syllables in visual word recognition: new insights from the lexical decision task. *Language, Cognition and Neuroscience*, *29*(10), 1249-1256. doi: 10.1080/23273798.2013.876504

Chirrey, D. (2007). Women like us: Mediating and contesting identity in lesbian advice literature. In H. Sauntson & S. Kyratzis (Eds.), *Language, Sexualities and Desires: Cross-Cultural Perspectives* (pp. 223-244). Basingstoke: Palgrave Macmillan.

Clyne, M. (2000). Lingua Franca and Ethnolects in Europe and Beyond. *Sociolinguistica, 14*(1), 83-89. doi: 10.1515/9783110245196.83

Coates, J. (1988) Gossip revisited: Language in all female groups. In J. Coates and D. Cameron (Eds.) *Women in their speech communitie*s (pp. 94-122). London: Longman.

Coates, J. (2011). Pushing at the Boundaries: The Expression of Alternative Masculinities. In J. Coates & P. Pichler (Eds.), *Language & Gender: A Reader* (2nd ed.) (pp.263-274). Chichester: Wiley-Blackwell.

Coates, J. (2016). *Women, Men and Language: The sociolinguistic evidence.* London: Routledge.

Coleman-Fountain, E. (2014). Lesbian and gay youth and the question of labels. *Sexualities, 17*(7), 802-817. doi: 10.1177/1363460714531432

Collett, D. (2003). *Modelling Binary Data*. (2nd ed.). Boca Raton, FL.: CRC Press.

Cook, J. (1967). In J. C. Beaglehole (Ed). *The Journals of Captain James Cook on his Voyages of Discovery: Volume III, Part I: The Voyage of the Resolution and Discovery 1776-1780*. New York, NY: Cambridge University Press for the Hakluyt Society.

Cook, R. D. & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Crespo-Fernández, E. (2019). Taboos in speaking of sex and sexuality. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.41-60). Oxford: Oxford University Press.

Crissman, H. P., Czuhajewski, C., Moniz, M. H., Plegue, M., & Chang, T. (2019). Youth perspectives regarding the regulating of bathroom use by transgender individuals. *Journal of Homosexuality*, No Pagination Specified. doi: 10.1080/00918369.2019.1618646

Croom, A. M. (2011). Slurs. *Language Sciences, 33*(3), 343-358. doi: 10.1016/j.langsci.2010.11.005

Culpeper, J. (2019). Taboo language and impoliteness. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.28-40). Oxford: Oxford University Press.

Culpeper, J. (2011). *Impoliteness: Using Language to Cause Offence (Studies in Interactional Sociolinguistics).* Cambridge: Cambridge University Press.

Damasio, A. R., & Geschwind, N. (1984). The Neural Basis of Language. *Annual Review of Neuroscience, 7*(1), 127-147. doi: 10.1146/annurev.ne.07.030184.001015

Dewaele, J-M. (2019). Linguistic taboos in a second or foreign language. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.218-232). Oxford: Oxford University Press.

Devinsky, O., & Schachter, S. C. (1997). *Norman Geschwind: Selected publications on language, behavior, and epilepsy*. Newton, MA: Butterworth-Heinemann.

Dinneen, L. C., & Blakesley, B. C. (1973). Algorithm AS 62: A Generator for the Sampling Distribution of the Mann- Whitney U Statistic. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 22*(2), 269-273. doi: 10.2307/2346934

Dundes, A. (1994). Towards a Metaphorical Reading of 'Break a Leg': A Note on Folklore of the Stage. *Western Folklore,* 53(1), 85-89. doi: 10.2307/1499654

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Eckert, P., & Wenger, É. (2005). Communities of practice in sociolinguistics. *Journal of Sociolinguistics, 9*(4), 582-589. doi: 10.1111/j.1360-6441.2005.00307

Edelman, L. (2004). *No Future: Queer Theory and the Death Drive*. Durham, NC, US: Duke University Press.

Ekman, P. (1985). *Telling Lies*. New York: W. W. Norton.

Ekman, P., & Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6, 238-252*. doi: 10.1007/BF00987191

Elischberger, H. B., Glazier, J. J., Hill, E. D., & Verduzco-Baker, L. (2016). “Boys don’t cry”—or do they? Adult attitudes toward and beliefs about transgender youth. *Sex Roles: A Journal of Research, 75*(5-6), 197-214. doi: 10.1007/s11199-016-0609-y

Ericson, R. V., Baranek, P. M., & Chan, J. B. L. (1991). *Representing Order: Crime, Law and Justice in the News Media*. Toronto: University of Toronto Press.

Elliot, A. J. (2006). The Hierarchical model of approach avoidance motivation. *Motivation and Emotion, 30*, 111-116. doi: 10.1007/s11031-006-9028-7

Elliot. A. J., Gable, S. L., & Mapes, R. R. (2006). Approach and Avoidance Motivation in the Social Domain. *Personality and Social Psychology Bulletin, 32*(3), 378-391. doi: 10.1177/0146167205282153

Field, A. (2013). *Discovering Statistics: Using IBM SPSS Statistics*. London: Sage.

Fielding, N. G., & Lee, R. M. (1991). (Eds.) *Using Computers in Qualitative Research.* London: Sage.

Finkelstein, S. R. (2019). Swearing and the brain. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.108-139). Oxford: Oxford University Press.

Fishman, P. M. (1980). Conversational insecurity. *Language.* 127-132. doi: 10.1016/B978-0-08-024696-3.50024-7

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: Wiley.

Fox, J. (1991). *Regression diagnostics*. Newbury Park, CA.: Sage.

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: pupillometry of spontaneous thought while reading. *Quarterly Journal of Experimental Psychology, 66*(12), 2289-94. doi: 10.1080/17470218.2013.858170

Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: using behavioral indices to detect mindless reading in real time. *Psychon Bull Rev*., *18*(5), 992-7. doi: 10.3758/s13423-011-0109-6

Gamson, J. (1998). Must identity movements self-destruct?: A queer dilemma. In P. M. Nardi and B. E. Schneider (Eds.), *Social perspectives in lesbian and gay studies* (pp. 589–604). London: Routledge.

Gaucher, D., Hunt, B., & Sinclair, L. (2015). Can pejorative terms ever lead to positive social consequences? The case of SlutWalk. *Language Sciences, 52*. doi: 10.1016/j.langsci.2015.03.005

Geer, J. H., & Bellard, H. S. (1996). Sexual content induced delays in unprimed lexical decisions: Gender and context effects. *Archives of Sexual Behaviour, 25*, 379-395. doi: 10.1007/BF02437581

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Giles, H. & Giles, J. (2013). Ingroups and outgroups. In A. Kurylo (Ed.) *Inter/cultural communication* (pp. 140-162). London: Sage.

Gill R. (2008) Empowerment/Sexism: Figuring Female Sexual Agency in Contemporary Advertising. *Feminism & Psychology*. *18* (1), 35-60. doi: 10.1177/0959353507084950

Gomm, R. (2004). *Social Research Methodology: A Critical Introduction*. Basingstoke: Palgrave Macmillan.

Gordon, Z. (2019). *Who Can Say Faggot? A Two-Part Study on Online Slur Reclamation.* Masters thesis, University of Amsterdam, Amsterdam, Holland. Retrieved from: https://scripties.uba.uva.nl/search?id=696515

Gramsci, A. (1971). *Selections from Prison Notebooks*. Trans. Q. Hoare, & G. Nowell-Smith. London: Lawrence and Wishart.

Gramsci, A. (1985). *Selections from Cultural Writings*. Eds. D. Forgacs, & G. Nowell-Smith, trans. W. Boelhower. London: Lawrence and Wishart.

Graves, R., Landis, T., & Goodglass, H. (1981). Laterality and sex differences for visual recognition of emotional and non-emotional words. *Neuropsychologia, 19*(1), 95-102. doi: 10.1016/0028-3932(81)90049-X

Gredebäck, G., & Melinder, A. (2011). Teleological reasoning in 4-month-old infants: pupil dilations and contextual constraints. *PLoS One, 6*(10), 1-5. doi: 10.1371/journal.pone.0026487

Grose, F. (1785). *A Classical Dictionary of the Vulgar Tongue.* London: S. Hooper.

Hall, K. (2014). Exceptional Speakers: Contested and Problematized Gender Identities. In S. Ehrlich, J. Holmes and M. Meyerhoff (Eds.) *The Handbook of Language, Gender, and Sexuality* (2nd ed.) (pp.220-239). Oxford: Blackwell.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles and Practice* (2nd ed.). London: Routledge.

Hankinson-Nelson, L. (1993). Epistemological Communities. In L. Alcoff & E. Potter (Eds.) *Feminist Epistemologies* (pp.121-159). New York: Routledge.

Hansen, A. (1995). Using information technology to analyze newspaper content. In R. M. Lee (Ed.), *Information Technology for the Social Scientist* (pp.147-168). London: UCL.

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*(4), 561-579. doi: 10.1017/S0142716403000286

Hart, A. (2001). Mann-Whitney test is not just a test of medians: Differences in spread can be important. *BMJ,* 323, 391-393. doi: 10.1136/bmj.323.7309.391

Haugh, M., & Culpeper, J. (2018). Integrative pragmatics and (im)politeness theory. In C. Ilie, & N. R. Norrick (Eds.), *Pragmatics and its Interfaces 294* (pp.213-239). Amsterdam: John Benjamins Publishing Company.

Hayes, J. (1976). Gayspeak. *The Quarterly Journal of Speech, 62*, 256-266.

Henningsen, H. (1957). Taboo-words among Seamen and Fishermen. *Mariners Mirror, 43*(4), 336–7.

Herbert, C. (2015). Precarious projects: The performative structure of reclamation. *Language Sciences*, *52*. 131-138. doi: 10.1016/j.langsci.2015.05.002

Hewitt, R. (1986). *White Talk, Black Talk: Inter-racial Friendship and Communication Amongst Adolescents.* Cambridge: Cambridge University Press.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300. doi: 10.1037/0003-066X.52.12.1280

Hilbe, J. M. (2016). *Practical Guide to Logistic Regression*. Boca Raton, FL: CRC Press.

Holmes, J. (1995). *Women, Men, and Politeness*. London: Longman.

Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Huber, G. P. (1981). The nature of organizational decision making and the design of decision support systems. *MIS Quarterly*, 1-10.

Huggett, R. (1975). *Supernatural on stage: Ghosts and superstitions of the theatre*. New York, NY: Taplinger.

Hughes, G. (2006). *An encyclopedia of swearing: The social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*. Armonk, N.Y: M.E. Sharpe.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hutton, C. (2019), Speech or conduct?: Law, censorship, and taboo language. In K. Allen (Ed.) *The Oxford Handbook of Taboo Words and Language* (pp. 264-284). Oxford: Oxford University Press.

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods, 40*(4), 1065-1074. doi: 10.3758/BRM.40.4.1065

Jansen, A. S. P., Van Nguyen, X., Karpitskiy, V., Mettenleiter, T. C., & Loewy, A. D. (1995). Central Command Neurons of the Sympathetic Nervous System: Basis of the Fight-or-Flight Response. *Science, 270* (5236), 644-646. doi: 10.1126/science.270.5236.644

Jay, T. (1980). A Frequency Count of College and Elementary School Students' Colloquial English. *Catalogue of Selected Documents in Psychology 10*(1).

Jay, T. (2000). *Why We Curse: A neuro-psycho-social theory of speech.* Amsterdam: John Benjamins Publishing Company.

Jay, T. B. (1992). *Cursing in America: A Psycholinguistic Study of Dirty Language in the Courts, in the Movies, in the Schoolyards, and on the Streets*. Amsterdam: John Benjamins Publishing Company.

Jay, T. B. (2019a). The Psychology of Expressing and Interpreting Linguistic Taboos. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp. 77-95). Oxford: Oxford University Press.

Jay, T.B. (2019b). Taboo language awareness in early childhood. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.96-107). Oxford: Oxford University Press.

Jay, K., & Jay, T. (2013). A Child's Garden of Curses: A Gender, Historical, and Age-Related Evaluation of the Taboo Lexicon. *The American journal of psychology, 126*, 459-475. doi: 10.5406/amerjpsyc.126.4.0459

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *American Journal of Psychology, 121*, 83-103. doi: 10.2307/20445445

Jay, T. B., & Danks, J. H. (1977). Ordering of taboo adjectives. *Bulletin of the Psychonomic Society, 9*(6), 405-408. doi: 10.3758/BF03337038

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research*, *4*(2), 267-288. doi: 10.1515/JPLR.2008.013

Jespersen, O. (1922). *Language: Its Nature, Development and Origin*. New York: W.W. Norton.

Johansson, W. (1981). The Etymology of the Word “Faggot." *Gay Books Bulletin*, *6,* 16-18.

Johnson, R.B. & Christensen, L. (2004). *Educational Research: Quantitative, Qualitative, and Mixed Approaches.* (2nd ed.). Boston, MA: Allyn and Bacon.

Johnson, S. (1755). *A Dictionary of the English Language*. London: Strahan.

Jones, L. (2012). *Dyke/Girl: Language and Identities in a Lesbian Group*. Palgrave Macmillan, London.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *The Quarterly Journal of Experimental Psychology, 64*(9), 1683-1691. doi: 10.1080/17470218.2011.605150

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. doi: 10.3758/BF03195800

Kerswell, P., & Sebba, M. (2011). *From London Jamaican to British youth language: The transformation of a Caribbean post-creole repertoise into a new Multicultural London English.* Paper presented 2011 Summer Conference of the Society for Pidgin and Creole Linguistics, July-August, University of Ghana, Accra.

Kerswill, P. (2014). The objectification of ‘Jafaican: the discoursal embedding of Multicultural London English in the British media. In: J., Androutsopoulos, (Ed.) *The Media and Sociolinguistic Change* (pp.428–455). Berlin: Walter de Gruyter.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behaviour Research Methods, 42*(3), 627-633. doi: 10.3758/BRM.42.3.627

Kiesling, S. F. (2001), "Now I Gotta Watch What I Say": Shifting Constructions of Masculinity in Discourse. *Journal of Linguistic Anthropology*, 11: 250-273. doi: 10.1525/jlin.2001.11.2.250

Kiesling, S. F. (1998). Men's identities and sociolinguistic variation: the case of fraternity men. *Journal of Sociolinguistics, 2*, 69-99. doi: 10.1111/1467/9481.00031

Kim, Y. Y. (2001). *Becoming intercultural: An integrative theory of cross-cultural adaptation*. Thousand Oaks: Sage

Kuipers J.R., Thierry G. (2013) ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex. 49*(10), 2853-60. doi: 10.1016/j.cortex.2013.01.012.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-Mediated Memory Consolidation: Role of the Medial Temporal Lobe in Humans. *Psychological Science, 9*(6), 490-493. doi: 10.1111/1467-9280.00090

Lakoff, R. (1973). Language and woman's place. *Language in Society, 2*(1), 45-79. doi:10.1017/S0047404500000051

Lalonde, R. N. (1992). The Dynamics of Group Differentiation in the Face of Defeat. *Personality and Social Psychology Bulletin, 18*, 336-342. doi: 10.1177/0146167292183010

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273. doi: 10.1111/j.1469-8986.1993.tb03352.x

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Startvik (Ed.), *Directions in Corpus Linguistics* (pp.105-122). Berlin: Mouton de Gruyter.

Leech, G. N. (2014). *The pragmatics of politeness.* New York: Oxford University Press.

Lehmann, E. L. (2006). *Nonparametrics: Statistical methods based on ranks*. New York: Springer.

Lemert, E. M. (1967). *Human deviance, social problems and social control*. Englewood Cliffs: Prentice-Hall.

Livia, A., & Hall, K. (1997). *Queerly Phrased: Language, Gender, and Sexuality*. Oxford: Oxford University Press.

Louw, B. (1993). Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In honour of John Sinclair* (pp.157-176). Amsterdam: John Benjamins Publishing Company.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D., Abrams, L., & Dyer, J. (2004). Relations between emotion, memory and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32*, 474-488. doi: 10.3758/BF03195840

Madan, C. A., Shafer, A. T., Chan, M., & Singhal, A. (2017). Shock and awe: Distinct effects of taboo words on lexical decision and free recall. *Quarterly Journal of Experimental Psychology, 70*(4), 793-810. doi: 10.1080/17470218.2016.1167925

Madan, C. R., Caplan, J. B., Lau, C. S. M., & Fujiwara, E. (2012). Emotional arousal does not enhance association-memory. *Journal of Memory and Language, 66*(4), 695-716. doi: 10.1016/j.jml.2012.04.001

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Sciences, 18*, 50-60. doi: 10.1214/aoms/1177730491

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behaviour Research Methods, 44*, 314-324. doi: 10.3758/s13428-011-0168-7

May, T. (1997). *Social Research: Issues, Methods and Process* (2nd. ed.). Buckingham: Open University Press.

McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus based language studies: An advanced resource book*. London: Routledge.

McEnery, T. (2006). *Swearing in English: Bad language, Purity, and Power From 1586 to the Present*. Abingdon/New York: Routledge.

McGowan, M. K. (2012). On 'Whites Only' Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination. In M. K. McGowan & I. Maitra (Eds.), *Speech and harm: Controversies over Free Speech* (pp. 121-147). Oxford: Oxford University Press.

Mehl, M., & Pennebaker, J. (2003). The Sounds of Social Life: A Psychometric Analysis of Students' Daily Social Environments and Natural Conversations. *Journal of Personality and Social Psychology, 84*, 857-870. doi: 10.1037/0022-3514.84.4.857

Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications.* Thousand Oaks, CA: Sage.

Mendoza-Denton, N. (2004). Language and Identity. In J.K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The Handbook of Language, Variation and Change* (pp.475-499). Oxford: Blackwell Publishing. doi: 10.1002/9780470756591.ch19

Motschenbacher, H. (2007). Can the term 'genderlect' be saved?: A postmodernist re-definition. *Gender & Language, 1*(2), 255-278.

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision, 13*(2), 1-20. doi: 10.1167/13.2.11

Ochsner, K. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of experimental psychology. General, 129*, 242-261. doi: 10.1037//0096-3445.129.2.242

Packard, M. G., & Cahill, L. (2001). Affective modulation of multiple memory systems. *Current Opinion in Neurobiology, 11*(6), 752-756. doi: 10.1016/s0959-4388(01)00280-x

Parent, M. C., & Silva, K. (2018). Critical consciousness moderates the relationship between transphobia and "bathroom bill" voting. *Journal of Counseling Psychology, 65*(4), 403-412. doi:10.1037/cou0000270

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59, 185-198*. doi: 10.1016/S1071-5819(03)00017-X

Pascoe, C. J. (2005). Dude, You're a Fag: Adolescent Masculinity and the Fag Discourse. *Sexualities, 8*(3), 329-346. doi: 10.1177/1363460705053337

Pascoe, C. J. (2007). *Dude, You're a fag: Masculinity and Sexuality in High School*. Berkeley: University of California Press.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*(6), 405-428. doi:10.1080/00207594.2012.743665

Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science, 17*(2), 147-152. doi:10.1111/j.1467-8721.2008.00565.x

Plummer, K. (1995). *Telling Sexual Stories: Power, Change and Social Worlds*. London: Routledge.

Popa-Wyatt, M. (2016). Not all slurs are equal. *Phenomenology and Mind, 11,*150-156*.* doi: 10.13128/Phe_Mi-20115.

Potter, J. (2000). Realism and sociolinguistics. *Journal of Sociolinguistics*, *4*(1), 21-24.

Raban, J. (1973). Giggling in code. *Lunch, 20*, 16-17.

Rivers, I. (2001). The bullying of sexual minorities at school: Its nature and long-term correlates. *Educational and Child Psychiatry, 18*(1), 32-46.

Romaine, S. (1984). The status of sociological models and categories in explaining linguistic variation. *Linguistische Berichte*, *90,* 25-38.

Russell, J. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39*, 1161-1178. doi:10.1037/h0077714

Russell, S. T., Clarke, T. J., & Clary, J. (2009). Are Teens "Post Gay"? Contemporary Adolescents' Sexual Identity Labels. *Journal of Youth and Adolescence, 38*(7), 884-890. doi: 10.1007/s10964-008-9388-2

Ryan, T.P. (1997). *Modern regression methods.* Hoboken, NJ: Wiley.

Sarantakos, S. (2005). *Social Research.* (3rd ed.). Basingstoke: Palgrave Macmillan.

Schegloff, E. (1997). Whose text? Whose context?. *Discourse and Society*, *8*, 165-187. doi: 10.1177/0957926597008002002

Scheepers, C., Mohr, S., Fischer, M.H., & Roberts, A.M. (2013) Listening to Limericks: a pupillometry investigation of perceivers' expectancy. *PLoS One*, *8*(9), e74986. doi: 10.1371/journal.pone.0074986

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology, 5*, 137. doi: 10.3389/fpsyg.2014.00137

Schulz, M. R. (1975). The semantic derogation of woman. In B.Thorne & N.Hanley (Ed.) *Language and sex: difference and dominance*. Rowley, MA: Newbury House.

Scott, J. (1990). *A Matter of Record: Documentary Sources in Social Research*. Cambridge: Polity.

Scott, M. (2012). *WordSmith Tools (Version 6)*. Stroud: Lexical Analysis Software.

Sebba, M. (1993). *London Jamaican*. London: Arnold.

Sharot, T., & Phelps, E. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, affective & behavioral neuroscience, 4*, 294-306. doi: 10.3758/CABN.4.3.294

Siegle, G.J., Granholm, E., Ingram, R.E., & Matt, G.E. (2001). Pupillary and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry. 49*(7), 624-36. doi: 10.1016/s0006-3223(00)01024-6.

Sinfield, A. (1994). *The Wilde century: effeminacy, Oscar Wilde, and the queer moment.* New London: Cassell.

Sirois, S. and Brisson, J. (2014), Pupillometry. *WIREs Cognitive Science, 5.* 679-692. doi: 10.1002/wcs.1323.

Smallwood, J., McSpadden, M. & Schooler, J.W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition 36,* 1144–1150. doi: 10.3758/MC.36.6.1144

Smith, G. W., & Smith, D. E. (1998). The Ideology of "Fag": The School Experience of Gay Students. *The Sociological Quarterly, 39*(2), 309-335. doi: 10.1111/j.1533/8525.1998.tb00506.x

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology 15*(1), 72-101. doi: 10.2307/1412159

Spears, R. A. (2000). *NTC's dictionary of American slang and colloquial expressions*. Lincolnwood, Ill.: National Textbook Company.

Spender, D. (1985). *Man Made Language*. London, Routledge.

Stanley, J. P. (1977). Paradigmatic woman: The prostitute. In D. L. Shores & C. P. Hines (Eds.), *Papers in language variation* (pp. 303-321). Tuscaloosa: University of Alabama Press.

Steinhauer, S. R., Boller, F., Zubin, J., & Pearlman, S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology, 20*, 472.

Stephens, R., Atkins, J., & Kingston, A. (2009). Swearing as a response to pain. *Neuroreport, 20*, 1056-1060. doi: 10.1097/WNR.0b013e32832e64b1

Stephens, R. & Umlaud, C. (2011). Swearing as a response to pain-effect of daily swearing frequency. *Journal of Pain, 12,* 1274-81. doi: 10.1016/j.jpain.2011.09.004.

Stevens, H. (2011). Normality and queerness in gay fiction. In H. Stevens (Ed.), *The Cambridge companion to gay and lesbian writing* (pp. 81-96). Cambridge: Cambridge University Press.

Strauss, E. (1983). Perception of emotional words. *Neuropsychologia,* 21(1), 99–103. doi: 10.1016/0028-3932(83)90104-5

Stubbs, M. (1995). Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies. *Functions of Language*, *2*(1), 23-55. doi: 10.1075/fol.2.1.03stu

Surakka, V., & Hietanen, J.K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology, 29*(1), *23-33.* doi: 10.1016/s0167-8760(97)00088-3

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. (6th ed.). Harlow, England: Pearson.

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior. In S.Worchel & W.G.Austin (Eds.), *Psychology of Intergroup Relation* (pp.7-24). Chicago: Nelson-Hall.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14*, 455-461. doi: 10.1111/1467/9280.02453

Tannen, D. (1990). *You Just Don't Understand: Men and Women in Conversation*. New York: Morrow.

Thelwall, M. (2008). Fk Yea I Swear: Cursing and Gender in a Corpus of MySpace Pages. *Corpora, 3*, 83-107. doi: 10.3366/E1749503208000087

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, *4*(2), 91-112. doi: 10.1093/applin/4.2.91

Thomas, L., & LaBar, K. (2005). Emotional arousal enhances word repetition priming. *Cognition and Emotion, 19*(7), 1027-1047. doi: 10.1080/02699930500172440

Thompson, W.K., Xie, M., & White, H.R. (2003). Transformations of covariates for longitudinal data. *Biostatistics, 4*(3), 353-364. doi:

Tirrell, L. (2012). Genocidal Language Games. In I. Maitra & M. K. McGowan (Eds.), *Speech and Harm: Controversies Over Free Speech* (pp.174-221). Oxford: Oxford University Press.

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding Language Science*. Chichester: Wiley Blackwell

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior, 5*(4), 381-391. doi: 10.1016/S0022-5371(66)80048-8

Turner, G. (2008). The road to the lesbian nation is not an easy one: "Us" and "Them" in Diva magazine. *Social Semiotics, 18*(3), 377-388. doi: 10.1080/10350330802217147

Turner, G. (2014). 'A real lesbian wouldn't touch a bisexual with a bargepole': contesting boundaries in the construction of collective identity. *Critical Discourse Studies, 12*(2), 139-162. doi: 10.1080/17405904.2014.974634

van Berkum, J.J.A. (2019). Language comprehension and emotion: where are the interfaces, and who cares?' In G.I. de Zubicaray & N.O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp.736-768). Oxford: Oxford University Press.

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2007). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, *45*(1), 130-140. doi: 10.1111/j.1469-8986.2007.00606.x

Wald, E. (2019). Taboo language used as banter. In K. Allan (Ed.), *The Oxford Handbook of Taboo Words and Language* (pp.334-352). Oxford: Oxford University Press.

Walters, M. A. Peterson, J.L., McDonnell, L. & Brown, R. (2019). Group identity, empathy and shared suffering: Understanding the 'community' impacts of anti-LGBT and Islamophobic hate crimes. *International Review of Victimology*, *26*(2), 143-162. doi: 10.1177/0269758019833284

West, C. & Zimmerman, D.H. (1983). Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. In B. Thorne, C. Kramarae, & N. Henley (Eds.) *Language, gender and society* (pp.102-117). Cambridge, MA: Newbury House.

White, E. (1980). The Political Vocabulary of Homosexuality. In L. Burke, T. Crowley, & A. Girvin (Eds.), (2000) *The Routledge Language and Cultural Theory Reader* (pp.189-196). London: Routledge.

Wierzbicka, A. (1986). What's in a Noun? (Or: How Do Nouns Differ in Meaning from Adjectives?). *Studies in Language*, *10*(2), 353-389. doi: 10.1075/sl.l0.2.05w

Williams, L.J. & Evans, J.R. (1980). Evidence for perceptual defense using a lexical decision task. *Perceptual and Motor Skills*, *50*(1), 195-198. doi: 10.2466/pms.1980.50.1.195.

Williamson, S., Harpur, T.J., & Hare, R.D. (1991), Abnormal Processing of Affective Words by Psychopaths. *Psychophysiology,* 28: 260-273. doi: 10.1111/j.1469-8986.1991.tb02192.x

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology, 51*(3), 277-284. doi: 10.1111/psyp.12151

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology, 81*(3), 133–141. doi: 10.1016/j.ijpsycho.2011.05.009

Zwicky, A. M. (1997). Two Lavender Issues for Linguists. In A. Livia & K. Hall (Eds.), *Queerly Phrased: Language, Gender and Sexuality* (pp.21-34). Oxford: Oxford University Press.

Appendix 1 – Annotated Second Question Responses (Chapter 3)

Key to Coding of Responses:

Red underline: References to specific LGBTQ+ terms.

Orange underline: References to reclamation.

Yellow underline: Opinions/experiences of LGBTQ+ slurs.

Green underline: Relations between identity and LGBTQ+ slur use.

Blue underline: Multiple/context-dependent uses of LGBTQ+ slurs.

Purple underline: Temporal/generational differences regarding LGBTQ+ slurs.

I'm Scottish and I think Jessie may be a Scottish term.

Queer is becoming a reclaimed term, so not everyone will find it offensive. It will depend on age, location, and who's using the word and their intention. Butch is also now a sub-culture and so not always a slur.

I listed those which sprung to mind. I'm not sure if all would be considered slurs.

Feels uncomfortable writing them down.

Pansy is old fashioned.

Some, like queer and dyke, have been somewhat reclaimed and are therefore not always slurs.

Many of the words are abbreviated versions of words used by the LGBTQIA+ community to self-identify - thus, context, performance and delivery are all-important in coding them as insults. Additionally, my experience varies in different parts of the country – 'gay' is much more commonly used as a slur in some communities than in others, for instance.

Whether or not the words are slurs often depends on who uses them, when and why.

I am quite glad that I could not remember any more.

I am straight and would use none of these terms. I think there has been partial reclaiming of queer but it makes many gay men I know very unhappy.

Most of the above were slurs I heard used back when I was at secondary school. Arse bandit was something I heard from the older generation (grandparents) and I assumed was a generational/regional thing.

Reclaimed words can still be used as slurs.

I was unsure about whether or not you were interested in slurs which are used currently, or whether you were interested in slurs which historically referred to the LGBTQ+ community, but which are now just slurs in general. I assumed the former.

Some of these I would see as pretty reclaimed by now and I'd be unlikely to be offended by them (e.g. I use 'queer' as a pretty standard/academic term and 'dyke' to describe myself). It really depends on the identity of the speaker and the context in which it takes place.

I don't feel Queer is a slur but I know it has been used by cishets to be derogatory.

'Correct' terms e.g. "gay", aren't slurs in themselves but can be used as such, similarly use of incorrect pronouns, such as "it" as I listed, can be used as slurs.

Chichi/cheechee/chichiman I heard from members of London's African Caribbean community about 10-15 years ago (also read in a book).

I am sure there are plenty more but these are the ones that come to mind.

"Queer" is the only word I hesitated to add because my understanding is that queer is not necessarily a pejorative and that terms like "queer folk" and "gender queer" etc can be legitimate words for people and behaviours and orientations in the LGBTQ community. However, I think that it is sometimes still used as a slur so that's why I decided to include it. But it's probably the only word I felt conflicted about including in the list.

For the ones in some situations, it's often the tone used behind them. Even though they're technically correct words they're often not wholly accurate or especially sensitive, and for that reason I see them as slurs (when intended as such). Similar for gay/queer/trans when used as a noun, it's often not especially kindly intended. e.g. 'a queer' is much less appropriate than 'a queer person'.

They're all horrible and offensive.

I forgot Chi chi man/batty man/bwoi. I think that both have African origins. Personally, I think queer as a term isn't pejorative.

Important to know that some of these (e.g. queer, dyke, tranny) are also reclaimed slurs that get used by LGBTQ people to describe themselves.

Some of these words have been reclaimed by their communities - e.g. dyke or queer. However, that doesn't mean that it's ok for people outside the community to use it, and it can still be seen as a slur by some members of the community affected by it.

I listed "warped"/"invert" which are archaic and obviously not in common usage, but I've come across them occasionally in very religious material. I also debated including "TERF" because it's often used to target lesbians for their sexual orientation/political stance, but, as it's also applied to heterosexual women, it doesn't quite fit the criteria here.

Despite all the slurs that I've listed are true, I'm still proud of who I am.

Of course, some of these words depend on context and tone of voice (e.g. "gay" and "queer").

I vehemently disagree with the reclamation of the slur 'queer'. For us older lesbians and especially gay men, it's still very much a slur. I'm from a working class community and it's a slur used frequently.

Queer, gay, and dyke, are words that have been and sometimes still are used as slurs, but they have been reclaimed as identities by the queer community, and I would not consider them slurs unless they were clearly used for that purpose.

Some of them are in Mexican Spanish and in French.

Some of the slurs I identified have been reclaimed and used as identities, such as I am a lesbian who would sometimes call myself a dyke.

Queer has been 'reclaimed' by some who use it as an identity rather than a slur.

Personally reclaim queer as an umbrella term as my sexuality is fluid and don't feel like I can isolate it to one specific term.

Interestingly most of these insults are really seen as beyond the pale now and few would use them openly in a non-ironic way. The homophobia these words represent really hasn't gone away though, just people who may be uncomfortable with gay lifestyles have learned to hide it away.

I think most of these are only really used by older generations now.

Some of these slurs are used positively to self-identify (e.g. queer) but they're still also used as slurs by bigots...

I don't think slurs themselves as language have changed much in a long time but the way they are used has changed subtly. I can't particular evidence this but I don't see people saying 'puffter' anymore but gay men are still obviously discriminated against. If my point means anything it is that I still feel prejudice exists at a troubling level but is more disguised.

Some of these we use ourselves so I'm not 100% if they are slurs.

I think several of these are used 'ironically' by (in my experience, particularly) straight middle class men to refer to gay people in a way which they would deny was a slur but which I recognise as distinctly othering.

With terms like "dyke", "faggot", to some extent they have been reclaimed so their status as a slur is debatable. The term "queer" has been reclaimed to such an extent now that its use as a slur is (in my experience) rare. I would also say that if the terms are used by someone outside the LGBTQ+ community, they are more likely to be interpreted as a slur.

Was really hard to think of some as they're not in my vocabulary. They are also quite old fashioned I think? In the 80's calling a straight person a gay or a lesbian was meant to be insulting.

I honestly don't think that many of them specifically target us beyond that.

Essentially any word used to attack someone purely based on identity.

Some I wouldn't consider offensive in many contexts e.g. Gay, Queer, Dyke.

Queer has been reclaimed to an extent and can be used by members of the LGBTQI+ community, but not by others.

Firstly, I'm a straight person! Also many of these feel very high school/90s or 00s. I remember queer being used as an insult but less so now.

These are comments I heard growing up in the 1970s

Not sure about queer - I think of it as a term that's been reclaimed now but when I was a child it was definitely used as a slur and could still be a slur depending on context.

I don't know which ones are offensive in which context. I've heard some people say they don't like "tranny" while others embrace it, so I avoid using it as I assume it is like the n word which can only be said by some people (which is perfectly fine). Growing up I thought queer was a slur but it is obviously being widely used now. As someone who hasn't always embraces her bisexuality, a lot of these words feel bad to use because I was aware that other people didn't know about my sexuality and so would assume that I do not have the right to use these words. Now that I've come out I'm still predisposed to avoid these words.

These are not always slurs and I use some of them about myself or others in the same lesbian community as me.

I think there are many LGBTQ slurs that relate to (actual or assumed) sexual activity between people. In particular gay and bi men. Other may relate to aspects of a persons (actual or assumed) behaviour eg a man, regarless of sexuality, is effeminate and gets called 'a fairy' or a woman who is loud and assertive might be labelled a 'dyke'. I also feel like a slur is implied through conversations such as 's/he's one of 'them' ' which is harder to pin point. As a an LBT woman you might be called a bitch in both a sexist and homophobic manner.

I don't really find the word 'queer' offensive, but I know that some people do.

Appendix 2 – All LGBTQ+ Slurs and Frequencies (Chapter 3)

LGBTQ+ Slur	Frequency	LGBTQ+ Slur (cont.)	Frequency (cont.)
Dyke	86	Lez	4
Faggot	83	Lezzie	4
Tranny	78	Nancy	4
Fag	74	Ponce	4
Queer	74	Uphill Gardener	4
Poof	69	Chick with a Dick	3
Homo	38	Nancy Boy	2
Lesbo	38	Paedo	2
Lezzer	37	She-He	2
Gay (Pejorative)	25	Shim	2
Bender	23	Woofter	2
Gayboy	23	Arthur	2
Batty Boy	22	Batty Man	2
Poofter	20	Butt-Fucker	2
Fairy	19	Deviant	2
Shemale	19	Diesel Dyke	2
He-She	18	Invert	2
Rug-Muncher	16	Mangina	2
Fudge Packer	15	Martha	2
Pansy	14	Mincer	2
Arse Bandit	12	Muff-Muncher	2
Shirt-Lifter	12	Paedophile	2
Bent	10	Perv	2
Batty	8	Pillow-Biter	2
Bummer	8	Pussy Boy	2
Butch	8	She-Man	2
Gender-Bender	8	A Gay (as noun)	1
Ladyboy	8	AIDS Boy	1
Trap	8	Anal Architect	1
Cock-Sucker	7	Arse Goblin	1
Gaylord	7	Ass Pirate	1
Muff-Diver	7	Augustus	1
Bull-Dyke	6	Backdoor Warrior	1
Bum Boy	6	Bitch	1
Pervert	6	Brown-Hatter	1
Camp	5	Bulldagger	1
Hermaphrodite	5	Bum Basher	1
Queen	5	Bumder	1
Shit-Stabber	5	Butt Boy	1
Sissy	5	Butt Pirate	1
Sodomite	5	Butt-Muncher	1
Bugger	4	Cheater	1
Fruit	4	Chichi	1
Greedy	4	Chutney Ferret	1
It	4	Cock Jockey	1
Lemon	4	Cock-Muncher	1

LGBTQ+ Slur	Frequency	LGBTQ+ Slur (cont.)	Frequency (cont.)
Confirmed Bachelor	1	Spoon-Bender	1
Confused	1	The Gays (collective noun)	1
Cunt-Licker	1	TIM/Trans-Identified Male	1
Cushion-Biter	1	Tomboy	1
Dag	1	Tran	1
Deliberate Use of Incorrect Pronouns	1	Transgenderism	1
Dirty Homo	1	Transgenderist	1
Dog	1	Transsexual	1
Effeminate	1	Transtrender	1
Faggy	1	Trixie	1
Fanny-Fiddler	1	Trysexual	1
Fay	1	Tunnel Tester	1
Fish fingers	1	Turd Burglar	1
Freak	1	Twink	1
Friend of Dorothy	1	Twist	1
Fruitcake	1	Unnatural	1
Fruity	1	Warped	1
Ginger	1	Weirdo	1
Ill	1	What Is It	1
Jessie	1	Whore	1
Knobjockey	1	Willy Woofter	1
Lencha	1		
Lesbian	1		
Lettuce Licker	1		
Lezzo	1		
Lickalotapuss	1		
Limp-Wristed	1		
Man-Hating Dyke	1		
Mannish	1		
Mary	1		
Mentally Ill	1		
Nonce	1		
Noun Uses of Gay/Queer/Transgender etc	1		
One of Them	1		
Pillow Princess	1		
Pussy	1		
Pussy-Muncher	1		
Queenie	1		
Ratchet	1		
Rug-Diver	1		
Rump Roaster	1		
Self-Confessed Player of the Pink Oboe	1		
Sick	1		
Slut	1		
Sod	1		
Special Snowflake	1		
Sperm-Rejector	1		

Appendix 3 – Example LGBTQ+ Slur Norming Survey (Chapter 4)

Properties and Use of Potentially Offensive LGBTQ+ Words

Page 1: Participant Information

Before responding to this survey, it is important that you understand why it is being carried out and what you will be asked to do, so that you can decide whether you want to take part. Please read through the following information and then tick the box at the bottom of this screen to proceed.

What is the purpose of the survey?

The results of this survey will be used to inform the design of future experiments investigating the different ways people respond to potentially offensive language and how this might be linked to issues of identity. This research focuses on offensive language applying to the LGBTQ+ community in particular.

Who can take part?

This survey is open to you if you are over the age of 16 and a native or near-native speaker of British English. You do not have to identify as LGBTQ+ to take part in this survey.

Do I have to take part?

It is up to you to decide whether or not to take part. If at any point during completion you no longer wish to take part, you are free to close the survey without submitting it. Any responses you have given up to that point will not be recorded. Please note that due to the anonymity of the survey, **you cannot withdraw your responses after you have finished and submitted**, as it will not be possible to identify which set of responses belong to you.

What will I be asked to do?

First, you will be asked to provide some brief demographic information. The main body of the survey is split into two sections. Together, these consist of 11 rating-scale questions, which aim to gather information about the properties (Section 1) and uses (Section 2) of a number of words which could be considered offensive to members of the LGBTQ+ community. Specific instructions are given at the start of each section, but throughout the survey you will be asked to give ratings for a list of words in response to a particular question. The survey typically takes between 15 and 20 minutes to complete.

Prize Draw

Upon completing the survey, you will be given the option of providing your email address to enter a prize draw for a £10 Amazon voucher. A winner will be chosen at random, and your email address will not be retained after the prize draw has closed.

Privacy

This survey is anonymous, and at no point during this survey will you be asked for information that could be used to identify you. As stated above, if you choose to provide an email address for the prize draw, this information will only be accessible to the researcher and will not be retained after the prize draw has closed. Any data you provide in this survey will be securely kept for the minimum seven years recommended by the University of Nottingham, and will then be destroyed.

Who can I contact if I have any further questions?

1 / 19

If you have any further questions about the research which have not been answered by this information sheet, please feel free to contact me (Daniel Edmondson) via email at: daniel.edmondson@nottingham.ac.uk. You may also contact my School's Ethics Officer (Dominic Thompson) at: dominic.thompson@nottingham.ac.uk.

Have you read and understood the above information? * *Required*

- Yes

Page 2: Statement of Informed Consent

Please tick the box below to confirm that you:

(a) are over 16, (b) understand the purpose of this study, (c) understand that all data are anonymous and that there will not be any connection between the personal information provided and the data, (d) understand that there are no known risks or hazards associated with participating in this study, (e) understand that by submitting this questionnaire, you agree that your answers, which you have given voluntarily, can be used anonymously for research purposes.

I have read the above requirements and confirm that I comply with them. ✱ *Required*

- ◦ Yes

Page 3: Demographic Information

Please indicate which of the following age brackets you fit into. * *Required*

- 16-19
- 20-29
- 30-39
- 40-49
- 50-59
- 60-69
- 70+

What is your gender identity? * *Required*

Does your gender identity match the one you were assigned at birth? * *Required*

- Yes
- No

What is your sexuality? * *Required*

Page 4: Section 1: Word Properties

In this section, you will be asked to rate a list of words which could be considered derogatory or offensive to members of the LGBTQ+ community, according to a number of word properties. You might not personally consider all of the listed words to be offensive, but each has been identified as such in responses to a previous survey. A statement at the top of the page will tell you which property you need to rate each word for. Each word will be listed alongside a rating scale ranging from 1 to 7, with an explanation for what these numbers represent at either end. For each word, simply indicate your rating on the scale.The example below shows how each rating task will be set out.

Example

Please indicate on the scale how pleasant you think this colour is:

	1 (Not At All Pleasant)	2 3 4 5 (Extremely Pleasant)
Red	□	□□□□ □
Blue	□	□□□□ □
Green	□	□□□□ □

Page 5: Section 1: Word Properties

Please indicate on the scale how frequently you have encountered this word. This can be in any context, not just in person.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 6: Section 1: Word Properties

Please indicate on the scale how frequently you use this word in reference to LGBTQ+ people, either positively or negatively. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 7: Section 1: Word Properties

Please indicate on the scale how shocking you consider each word to be. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Not At All Shocking)	2	3	4	5 (Extremely Shocking)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 8: Section 1: Word Properties

Please indicate on the scale how negative you consider each word to be. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Not At All Negative)	2	3	4	5 (Extremely Negative)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 9: Section 1: Word Properties

Please indicate on the scale how offensive you consider each word to be. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Not At All Offensive)	2	3	4	5 (Extremely Offensive)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 10: Section 1: Word Properties

Please indicate on the scale how frequently you would anticipate each word to be met with disapproval. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 11: Section 1: Word Properties

Please indicate, to the best of your knowledge, the approximate age you think you first heard each word.

	0-5	6-10	11-15	16-19	20-29	30-39	40-49	50-59	60-69	70+
Gaylord	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Page 12: Section 2: Word Use

In this section, you will be asked to consider the same words presented in Section 1, but will be asked to rate them based on some questions about their use. The format of the tasks in this section is the same. For the first question, if you do not think that a word could be applied to yourself, you are asked to tick the box which indicates this. Otherwise, as in the previous section, you are asked to give a rating for each word using the scale provided.

Page 13: Section 2: Word Use

Please indicate on the scale how frequently **you** would use each word **positively** to describe **yourself.** If you do not think this word could be applied to you, please tick the box to indicate this. If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	This word does not apply to me	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐	☐

Page 14: Section 2: Word Use

Please indicate on the scale how frequently **you** would use this word **positively** to describe **other people.** If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 15: Section 2: Word Use

Please indicate on the scale how frequently you have heard **other people** use this word **positively** to describe **themselves.** If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 16: Section 2: Word Use

Please indicate on the scale how frequently you have heard **other people** use this word **positively** to describe **other people.** If you have not encountered a word before, you are asked to leave it blank rather than provide a rating.

	1 (Never)	2	3	4	5 (Extremely Frequently)
Gaylord	☐	☐	☐	☐	☐
Camp	☐	☐	☐	☐	☐
Shit-Stabber	☐	☐	☐	☐	☐
Gayboy	☐	☐	☐	☐	☐
Cock-Sucker	☐	☐	☐	☐	☐
Rug-Muncher	☐	☐	☐	☐	☐
Fudge-Packer	☐	☐	☐	☐	☐
Pervert	☐	☐	☐	☐	☐
Sissy	☐	☐	☐	☐	☐
Queer	☐	☐	☐	☐	☐
Butch	☐	☐	☐	☐	☐
He-She	☐	☐	☐	☐	☐
Muff-Diver	☐	☐	☐	☐	☐
Tranny	☐	☐	☐	☐	☐
Trap	☐	☐	☐	☐	☐
Shirt-Lifter	☐	☐	☐	☐	☐
Lezzer	☐	☐	☐	☐	☐
Bum Boy	☐	☐	☐	☐	☐
Gender-Bender	☐	☐	☐	☐	☐
Bent	☐	☐	☐	☐	☐

Page 17: Prize Draw

You now have the option of entering the prize draw to win a £10 Amazon voucher. Whether you decide to take part or not, please make sure you click the 'Finish' button below to ensure that your responses to the survey are recorded.

If you would like to enter the optional prize draw, please provide a contact email address in the space below.

Email address:

18 / 19

Page 18: Thank you!

Thank you very much for giving your time and your input, the survey is now complete! Your responses have been recorded and you are now free to close this page.

If you have any further questions about the research, please feel free to contact me (Daniel Edmondson) at: daniel.edmondson@nottingham.ac.uk.

Appendix 4 – All Word Property and Reclamation Behaviour Norms for LGBTQ+ Slurs (Chapter 4)

Slur	Words	Characters	Syllables	Familiarity	*N*	*SD*	Personal Use	*N*	*SD*	Arousal	*N*	*SD*	Negativity	*N*	*SD*
Arse-Bandit	2	11	3	2.03	77	1.09	1.15	62	0.40	3.39	59	1.34	3.97	58	1.18
Batty	1	5	2	2.33	75	1.14	1.20	60	0.63	2.56	57	1.23	3.79	56	1.16
Batty-Boy	2	9	3	2.35	77	1.21	1.22	65	0.65	2.95	60	1.31	3.81	54	1.18
Bender	1	6	2	2.58	76	1.00	1.25	69	0.60	2.61	70	1.08	3.50	70	1.21
Bent	1	4	1	3.47	78	1.32	1.60	70	1.00	2.56	66	1.17	3.31	67	1.20
Bull-Dyke	2	9	2	1.60	77	0.83	1.20	55	0.56	2.75	48	1.21	3.42	50	1.23
Bum-Boy	2	7	2	2.61	76	1.30	1.16	63	0.48	3.39	59	1.10	3.92	60	0.98
Bummer	1	6	2	2.53	76	1.25	1.22	67	0.60	2.78	67	0.60	3.76	62	1.13
Butch	1	5	1	3.67	78	1.29	2.45	75	1.28	1.76	73	0.95	2.22	74	1.14
Camp	1	4	1	4.15	78	1.05	2.88	75	1.21	1.44	75	0.72	1.95	73	0.97
Cock-Sucker	2	11	3	3.33	78	1.10	1.32	73	0.76	3.27	74	1.21	4.01	74	1.09
Dyke	1	4	1	3.40	77	1.24	1.96	73	1.15	2.55	71	1.32	3.23	71	1.30
Fag	1	3	1	3.81	77	0.96	1.66	73	1.10	3.24	76	1.35	4.01	76	1.09
Faggot	1	6	2	3.66	77	1.06	1.55	73	1.09	3.58	77	1.34	4.43	76	0.87
Fairy	1	5	2	2.61	76	1.10	1.27	67	0.59	2.14	70	1.15	3.10	68	1.15
Fudge-Packer	2	11	3	1.97	78	1.09	1.09	55	0.35	3.61	52	1.11	4.28	53	0.79
Gay (pejorative)	1	3	1	4.29	77	0.86	2.59	74	1.58	2.05	77	1.11	3.00	75	1.28
Gayboy	1	6	2	3.14	78	1.05	1.45	73	0.92	2.49	69	0.99	3.42	71	1.13
Gaylord	1	7	2	2.49	78	0.91	1.29	73	0.68	2.30	70	0.98	3.44	71	1.04
Gender-Bender	2	13	4	2.62	78	1.14	1.25	69	0.72	2.79	67	1.30	3.31	65	1.32
Hermaphrodite	1	13	4	2.08	77	0.94	1.16	67	0.45	2.33	64	1.18	2.70	63	1.33
He-She	2	6	2	2.64	78	1.17	1.15	68	0.47	3.54	69	1.22	4.21	68	1.10
Homo	1	4	2	3.92	77	0.96	2.08	74	1.21	2.61	76	1.34	3.09	76	1.28
Ladyboy	1	7	2	2.48	77	1.07	1.39	70	0.75	2.70	69	1.12	3.37	68	1.18
Lesbo	1	5	2	3.22	77	1.19	1.71	73	1.02	2.32	74	1.14	3.10	73	1.19
Lezzer	1	6	2	2.92	78	1.35	1.51	65	1.03	2.65	62	1.06	3.27	63	1.07
Muff-Diver	2	10	3	1.97	78	1.16	1.28	57	0.77	3.08	51	1.06	3.63	51	1.09
Pansy	1	5	2	2.50	78	1.03	1.13	69	0.34	2.81	70	1.20	3.64	67	1.03
Pervert	1	7	2	3.76	78	1.19	1.20	75	0.66	3.65	77	1.42	4.60	77	0.83
Poof	1	4	1	3.21	77	1.20	1.49	70	0.83	2.75	68	1.14	3.45	67	1.16
Poofter	1	7	2	2.66	76	1.21	1.28	68	0.67	2.88	67	1.21	3.58	66	1.14
Queen	1	5	1	3.66	76	1.20	2.51	73	1.31	1.45	75	0.70	1.79	72	0.99
Queer	1	5	1	4.38	78	0.87	3.22	77	1.45	1.76	78	1.11	1.99	78	1.24
Rug-Muncher	2	11	3	2.26	78	1.29	1.29	58	0.73	3.15	54	1.07	3.95	55	1.04
Shemale	1	7	2	2.35	77	1.07	1.27	67	0.69	2.97	68	1.23	3.66	68	1.28
Shirt-Lifter	2	12	3	1.92	77	1.30	1.31	52	0.83	3.02	46	1.31	3.60	43	1.22
Shit-Stabber	2	12	3	1.38	78	0.65	1.06	50	0.24	3.85	48	1.07	4.44	50	0.86
Sissy	1	5	2	3.06	78	1.17	1.27	73	0.79	2.57	76	1.16	3.59	76	1.11
Sodomite	1	8	3	1.91	77	0.89	1.24	67	0.61	2.95	65	1.40	3.97	64	1.37
Tranny	1	6	2	3.63	78	1.14	1.44	73	1.03	3.51	77	1.37	3.97	76	1.22
Trap	1	4	1	1.55	78	1.06	1.00	49	0.00	2.90	30	1.54	3.74	27	1.46

Slur	Offensiveness	*N*	*SD*	Tabooness	*N*	*SD*	Age of Acquisition	*N*	*SD*	Self-Self Reclamation	*N*	*SD*	Self-Other Reclamation	*N*	*SD*
Arse-Bandit	4.00	59	1.31	3.77	57	1.20	3.91	56	1.42	1.22	23	0.85	1.11	64	0.40
Batty	3.61	56	1.20	3.13	56	1.16	4.12	60	1.80	1.31	26	0.97	1.08	62	0.42
Batty-Boy	3.82	55	1.19	3.47	55	1.17	4.16	58	1.88	1.19	21	0.68	1.13	61	0.56
Bender	3.51	70	1.18	3.26	70	1.14	3.26	70	1.25	1.46	26	0.81	1.17	70	0.56
Bent	3.40	68	1.24	3.16	68	1.36	2.96	69	0.98	1.68	41	1.08	1.46	69	0.98
Bull-Dyke	3.56	50	1.15	3.33	48	1.12	4.67	49	1.48	1.19	16	0.75	1.15	60	0.52
Bum-Boy	3.84	62	1.16	3.81	62	1.01	3.23	64	1.05	1.33	24	0.87	1.11	65	0.40
Bummer	3.69	64	1.26	3.37	63	1.26	3.53	62	1.45	1.22	27	0.80	1.15	67	0.56
Butch	2.01	70	1.04	2.22	72	0.92	3.16	75	0.85	1.57	28	0.74	2.27	73	1.22
Camp	1.93	75	1.00	2.00	74	0.91	2.95	75	1.16	2.35	34	1.18	2.53	74	1.23
Cock-Sucker	3.85	74	1.20	3.88	75	1.19	3.31	75	0.72	1.60	40	0.93	1.28	76	0.74
Dyke	3.28	71	1.30	3.20	71	1.24	3.37	71	0.90	1.67	24	1.09	1.77	74	1.19
Fag	4.04	76	1.11	3.63	76	1.26	2.99	76	0.86	1.55	33	0.90	1.32	76	0.70
Faggot	4.37	76	0.94	3.99	76	1.09	3.05	76	0.86	1.68	34	1.12	1.29	75	0.80
Fairy	3.06	69	1.17	2.75	71	1.12	2.70	71	1.03	1.69	32	1.06	1.43	72	0.80
Fudge-Packer	4.18	50	0.94	4.18	50	1.04	3.91	57	0.91	1.07	28	0.26	1.02	56	0.23
Gay (pejorative)	2.87	76	1.32	2.59	76	1.28	3.07	76	1.58	2.95	43	1.53	2.47	75	1.51
Gayboy	3.29	72	1.05	3.36	70	0.96	2.84	70	0.99	1.88	32	1.13	1.54	71	1.00
Gaylord	3.26	73	1.12	3.40	72	1.00	3.00	72	1.22	1.62	29	0.94	1.30	73	0.78
Gender-Bender	3.33	66	1.37	3.30	67	1.33	3.72	67	1.26	1.43	21	1.12	1.28	67	0.81
Hermaphrodite	2.95	62	1.45	2.57	61	1.20	3.64	66	1.25	1.15	13	0.55	1.07	70	0.31
He-She	4.16	67	0.93	3.94	68	1.13	3.60	70	1.51	1.07	14	0.27	1.11	72	0.43
Homo	3.22	76	1.29	3.00	75	1.20	2.70	76	0.88	2.53	40	1.41	1.79	75	1.23
Ladyboy	3.47	68	1.22	3.10	67	1.16	3.82	66	1.31	1.23	13	0.83	1.17	72	0.53
Lesbo	3.16	75	1.20	2.93	74	1.23	3.09	75	1.21	1.42	24	0.88	1.51	76	1.04
Lezzer	3.20	61	1.03	3.09	64	1.14	3.04	67	1.21	1.15	20	0.37	1.46	65	0.97
Muff-Diver	3.67	48	1.21	3.98	50	1.02	3.91	55	0.99	1.32	22	0.65	1.15	55	0.45
Pansy	3.54	68	1.16	3.35	69	1.17	3.00	69	0.94	1.14	36	0.35	1.08	72	0.33
Pervert	4.40	77	1.10	4.17	77	1.17	2.59	76	0.70	1.34	38	0.78	1.17	77	0.50
Poof	3.67	67	1.04	3.30	66	1.10	2.82	67	0.80	1.66	29	1.08	1.34	70	0.80
Poofter	3.58	66	1.08	3.35	66	1.10	2.99	67	0.95	1.31	29	0.85	1.20	69	0.65
Queen	1.97	72	1.03	1.88	72	0.90	3.18	71	1.17	2.58	43	1.33	2.80	74	1.34
Queer	2.01	78	1.24	2.37	78	1.12	2.88	78	0.82	3.39	46	1.47	3.08	78	1.40
Rug-Muncher	3.85	54	1.14	4.05	55	1.03	3.85	60	0.99	1.38	24	0.77	1.14	58	0.48
Shemale	3.85	68	1.24	3.34	68	1.22	3.84	68	1.38	1.13	16	0.50	1.06	68	0.29
Shirt-Lifter	3.55	44	1.32	3.36	44	1.31	4.04	50	1.51	1.26	19	0.73	1.12	51	0.48
Shit-Stabber	4.31	49	0.89	4.34	50	0.85	4.30	53	1.19	1.13	24	0.45	1.02	52	0.14
Sissy	3.36	76	1.22	3.20	75	1.26	2.32	75	0.84	1.31	29	0.81	1.14	76	0.48
Sodomite	3.80	64	1.36	3.48	64	1.36	3.55	65	0.92	1.30	30	0.65	1.12	69	0.40
Tranny	4.09	76	1.18	3.73	75	1.26	3.34	77	1.14	1.07	14	0.27	1.22	77	0.70
Trap	3.96	26	1.28	3.42	24	1.35	4.94	33	1.77	1.00	10	0.00	1.03	38	0.16

Slur	Other-Self Reclamation	*N*	*SD*	Other-Other Reclamation	*N*	*SD*	Self-Reclamation	*N*	*SD*	Other Reclamation	*N*	*SD*	Overall Reclamation	*N*	*SD*
Arse-Bandit	1.13	64	0.42	1.07	61	0.31	1.12	65	0.46	1.09	65	0.33	1.10	67	0.36
Batty	1.15	61	0.40	1.11	62	0.37	1.10	63	0.47	1.13	64	0.36	1.12	65	0.36
Batty-Boy	1.10	62	0.35	1.08	60	0.28	1.10	62	0.48	1.09	62	0.28	1.09	65	0.34
Bender	1.34	71	0.61	1.29	70	0.62	1.17	71	0.50	1.32	72	0.55	1.26	73	0.44
Bent	1.62	68	1.15	1.65	68	1.09	1.47	69	0.89	1.63	68	1.04	1.56	69	0.91
Bull-Dyke	1.38	61	0.78	1.26	57	0.61	1.13	60	0.48	1.33	61	0.68	1.24	62	0.54
Bum-Boy	1.14	65	0.53	1.19	64	0.47	1.12	65	0.43	1.16	65	0.46	1.14	66	0.43
Bummer	1.16	68	0.48	1.09	66	0.29	1.12	68	0.52	1.12	69	0.34	1.12	70	0.39
Butch	2.77	73	1.32	2.69	71	1.28	2.07	74	1.04	2.73	73	1.24	2.43	75	1.06
Camp	2.82	74	1.22	2.86	74	1.26	2.39	74	1.05	2.84	74	1.12	2.64	74	0.99
Cock-Sucker	1.59	75	0.92	1.28	74	0.59	1.31	76	0.70	1.43	75	0.65	1.38	76	0.63
Dyke	2.27	75	1.32	1.99	73	1.29	1.66	74	1.03	2.14	75	1.25	1.95	75	1.05
Fag	1.80	76	1.02	1.68	74	1.01	1.32	76	0.71	1.74	76	0.94	1.55	77	0.73
Faggot	1.57	76	1.01	1.50	72	0.98	1.31	75	0.74	1.53	76	0.97	1.42	77	0.78
Fairy	1.81	74	0.96	1.70	71	0.88	1.42	72	0.76	1.75	74	0.84	1.63	74	0.73
Fudge-Packer	1.05	55	0.23	1.02	54	0.14	1.03	57	0.15	1.04	55	0.16	1.03	57	0.14
Gay (pejorative)	3.00	75	1.44	2.76	74	1.52	2.40	76	1.42	2.86	76	1.40	2.68	76	1.34
Gayboy	1.89	72	1.18	1.72	72	1.01	1.52	72	0.93	1.81	72	1.00	1.70	73	0.87
Gaylord	1.44	73	0.82	1.37	73	0.75	1.33	73	0.75	1.40	73	0.69	1.39	74	0.67
Gender-Bender	1.52	67	0.97	1.45	67	0.89	1.29	68	0.81	1.49	67	0.87	1.40	68	0.78
Hermaphrodite	1.12	69	0.40	1.15	68	0.43	1.08	70	0.34	1.13	69	0.39	1.10	71	0.33
He-She	1.12	69	0.56	1.16	70	0.58	1.12	72	0.43	1.14	70	0.52	1.13	72	0.47
Homo	2.33	75	1.40	2.09	74	1.26	1.83	76	1.21	2.24	76	1.29	2.06	77	1.16
Ladyboy	1.29	73	0.72	1.34	70	0.74	1.16	73	0.53	1.32	73	0.68	1.26	74	0.59
Lesbo	2.03	74	1.28	1.73	74	1.02	1.46	76	0.91	1.88	76	1.08	1.71	77	0.90
Lezzer	1.80	65	1.02	1.69	65	1.01	1.42	66	0.90	1.75	65	0.91	1.60	66	0.74
Muff-Diver	1.28	53	0.66	1.17	52	0.43	1.14	57	0.41	1.22	54	0.51	1.18	58	0.44
Pansy	1.21	71	0.50	1.14	69	0.49	1.08	72	0.25	1.17	73	0.47	1.13	73	0.31
Pervert	1.36	74	0.85	1.23	75	0.63	1.18	77	0.48	1.30	76	0.69	1.24	77	0.56
Poof	1.59	70	0.92	1.47	68	0.74	1.33	70	0.77	1.52	71	0.74	1.45	71	0.68
Poofter	1.47	70	0.83	1.37	67	0.76	1.19	69	0.66	1.44	70	0.77	1.32	70	0.62
Queen	3.49	74	1.25	3.29	72	1.33	2.58	74	1.22	3.41	74	1.22	3.03	75	1.10
Queer	3.96	77	1.13	3.65	77	1.20	3.06	78	1.34	3.81	77	1.10	3.47	78	1.12
Rug-Muncher	1.28	57	0.65	1.16	56	0.46	1.15	59	0.49	1.22	57	0.51	1.19	59	0.48
Shemale	1.20	71	0.65	1.24	68	0.69	1.07	68	0.31	1.07	71	0.62	1.15	71	0.43
Shirt-Lifter	1.12	49	0.39	1.13	48	0.49	1.12	51	0.46	1.12	50	0.40	1.12	52	0.41
Shit-Stabber	1.08	53	0.27	1.02	53	0.14	1.04	54	0.19	1.04	55	0.17	1.04	57	0.17
Sissy	1.36	75	0.78	1.27	75	0.58	1.15	76	0.50	1.15	75	0.60	1.25	76	0.50
Sodomite	1.14	70	0.49	1.11	66	0.47	1.12	70	0.41	1.12	72	0.44	1.12	72	0.41
Tranny	1.37	76	0.59	1.39	75	0.77	1.23	77	0.70	1.23	76	0.58	1.31	77	0.55
Trap	1.35	34	0.73	1.24	34	0.55	1.01	38	0.08	1.01	35	0.61	1.16	39	0.35

Appendix 5 – All Words in Each Stimulus Condition (Chapter 5)

LGBTQ+ Slurs	Negative High Arousal Words	Negative Low Arousal Words	Non-Words
Batty	Anger	Alone	Bebber
Bender	Assault	Broken	Bech
Bent	Brutal	Carcass	Berfer
Bummer	Chaos	Crime	Bews
Butch	Debt	Cut	Bleen
Camp	Guilty	Foul	Bleer
Dyke	Gun	Illness	Buked
Fag	Pest	Lonely	Bunts
Faggot	Sinful	Mildew	Cade
Fairy	Tense	Rigid	Calsy
Gay	Torture	Rusty	Calt
Gayboy	Toxic	Sick	Cawpy
Gaylord			Chemace
Homo			Chemart
Lesbo			Cunder
Lezzer			Dag
Pansy			Deke
Pervert			Dinsy
Poof			Donder
Poofter			Dyle
Queen			Faw
Queer			Ferty
Shemale			Filmot
Sissy			Fiquot
Tranny			Gabsoy
			Gabtord
			Gacroy
			Gaglops
			Gak
			Gaw
			Hodi
			Hova
			Lexpo
			Loof
			Peagter
			Peef
			Perdort
			Pernart
			Queeg
			Rassy
			Reefter
			Sangy
			Sempy
			Skeen
			Tarmy
			Tesker
			Traspy
			Tusbo
			Tuzzer

Appendix 6 – All Words in Each Stimulus Condition (Chapter 6)

LGBTQ+ Slurs	Non-Slur Taboo Words	Non-Words
Batty	Asshole	Aff
Bender	Anus	Bamps
Bent	Ass	Bastick
Bummer	Balls	Baves
Butch	Bastard	Bebter
Camp	Blowjob	Beevy
Dyke	Boner	Belfer
Fag	Boobs	Bemp
Faggot	Booty	Blit
Fairy	Climax	Booch
Gay	Coitus	Boove
Gayboy	Cum	Bunser
Gaylord	Cunt	Cancy
Homo	Dick	Cass
Lesbo	Dildo	Chauly
Lezzer	Hooker	Chemave
Pansy	Horny	Cug
Pervert	Nutsack	Cung
Poof	Pecker	Dack
Poofter	Penis	Didso
Queen	Pussy	Doder
Queer	Rimjob	Fick
Shemale	Shit	Filmot
Sissy	Stupid	Fizzot
Tranny	Twat	Fum
		Gabtord
		Gagborm
		Gagpoy
		Gax
		Glat
		Heeper
		Hobi
		Hova
		Hurby
		Lexpo
		Minsy
		Nubsand
		Percort
		Perty
		Prupod
		Purveld
		Reefter
		Rucker
		Sneen
		Spimag
		Tarmy
		Telker
		Tusbo
		Wheer
		Wrat

Appendix 7 – All Sentence Stimuli in Each Version of Eye-Tracking Study (Chapter 7)

Version 1 Sentence Stimuli

He was being a flirt and David was flattered.
Lucy was a massive help to me last week.
They thought Tim was a dick so he was not very popular.
Jessie was being a pest so he started ignoring her.
Steph was a nerd and spent all her time revising.
She was a distraction so they failed the module.
Chris thought only a bastard would act like that.
He thinks Joe is a star in his classes.
Gordon was acting like a prick and Michelle was annoyed by it.
Cassidy thought only a nitwit would do something so stupid.
Robin thought that only a poof would behave in that way.
They suspected Emma was a spy because she was always leaving.
People thought she was a lesbo and wouldn't hang out with her.
I heard he slept with a whore who lives on his street.
Johnny would be a legend if he did that.
Ruth is an asset to her community.
She seems like a genius when you speak to her.
Josie had become a bigot and she no longer liked her.
His lecturer was a bore and he stopped listening.
He looked like a poofter in those clothes.
Frazer is a geek and loves playing games.
Abigail is often a downer and people find her draining.
Liam looks like a model in his new outfit.
She felt Brad was a joker and found him annoying.
She thought he was a pussy and teased him for it.
People think Ben is lovely since he is so polite.
She thought he was a friend and enjoyed his company.
He thought Chad was a muppet but he liked him anyway.
People will think you're a queer if you go out like that.
His teacher was crap and Dave didn't like him.
He shouldn't be such a fag if he wants to be liked.
She thought the actor was bent and wouldn't watch his films.
They said James was a batty because of how he acted.
He always avoided that shemale whenever he saw them.
Some girls called Jess a lezzer because of her hobbies.
I wouldn't have thought he was an ass before you told me.
The dress looked like a tent because it didn't fit her.
They thought Martin was fun and liked when he visited.
They thought Will was a gossip and didn't trust him.
Everyone thought they were a laugh at the party yesterday.
Ella thought his fashion was shit and did not like it.
Robert was being a bully so they avoided him.
Rachel looks like a rockstar in her leather jacket.
Don't be such a queen around other people.
Trish is a collector of many board games.
She didn't seem like a twat when I spoke to her.
You'd have to be a dyke to have that haircut.
Josh is a liar and you shouldn't believe him.
Karen was a nuisance to the retail worker.
Lewis is really talented at writing new songs.
Matt thought his waiter was camp and didn't give him a tip.
They thought he was a liability so they fired him.
He thought she looked like a hooker in the outfit she had chosen.
He called Joe a moron and started avoiding him.
Jake was acting like an asshole at the gym yesterday.
They are an annoyance because they are so needy.
His boss is a pain and he doesn't like him.
Amy would not invite a loser to her birthday party.
He knew Hannah was an expert in her field.

Version 2 Sentence Stimuli

Alex called Luke a pussy in order to annoy him.
He was being a flirt and David was flattered.
Lucy was a massive help to me last week.
They thought Tim was a poof so he was not very popular.
Jessie was being a pest so he started ignoring her.
Steph was a nerd and spent all her time revising.
She was a distraction so they failed the module.
Chris thought only a shemale would act like that.
He thinks Joe is a star in his classes.
Gordon was acting like a queen and Michelle was annoyed by it.
Cassidy thought only a nitwit would do something so stupid.
Robin thought that only a dick would behave in that way.
They suspected Emma was a spy because she was always leaving.
People thought she was a loser and wouldn't hang out with her.
I heard he slept with a queer who lives on his street.
Johnny would be a legend if he did that.
Ruth is an asset to her community.
She seems like a genius when you speak to her.
Josie had become a bigot and she no longer liked her.
His lecturer was a bore and he stopped listening.
He looked like an asshole in those clothes.
Frazer is a geek and loves playing games.
Abigail is often a downer and people find her draining.
Liam looks like a model in his new outfit.
She felt Brad was a joker and found him annoying.
She thought he was a sissy and teased him for it.
People think Ben is lovely since he is so polite.
She thought he was a friend and enjoyed his company.
He thought Chad was a muppet but he liked him anyway.
People will think you're a whore if you go out like that.
His teacher was bent and Dave didn't like him.
He shouldn't be such an ass if he wants to be liked.
She thought the actor was crap and wouldn't watch his films.
They said James was a moron because of how he acted.
He always avoided that bastard whenever he saw them.
Some girls called Jess a hooker because of her hobbies.
I wouldn't have thought he was a fag before you told me.
The dress looked like a tent because it didn't fit her.
They thought Martin was fun and liked when he visited.
They thought Will was a gossip and didn't trust him.
Everyone thought they were a laugh at the party yesterday.
Ella thought his fashion was camp and did not like it.
Robert was being a bully so they avoided him.
Rachel looks like a rockstar in her leather jacket.
Don't be such a prick around other people.
Trish is a collector of many board games.
She didn't seem like a dyke when I spoke to her.
You'd have to be a twat to have that haircut.
Josh is a liar and you shouldn't believe him.
Karen was a nuisance to the retail worker.
Lewis is really talented at writing new songs.
Matt thought his waiter was shit and didn't give him a tip.
They thought he was a liability so they fired him.
He thought she looked like a lezzer in the outfit she had chosen.
He called Joe a batty and started avoiding him.
Jake was acting like a poofter at the gym yesterday.
They are an annoyance because they are so needy.
His boss is a pain and he doesn't like him.
Amy would not invite a lesbo to her birthday party.
He knew Hannah was an expert in her field.