

Collective Identity and Careers in a White Nationalist Forum

Amy Ellison Booth

Doctor of Philosophy

Aston Institute for Forensic Linguistics

Aston University

September 2023

©Amy Ellison Booth, 2023

Amy Ellison Booth asserts her moral right to be identified as the author of this thesis.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright belongs to its author and that no quotation from the thesis and no information derived from it may be published without appropriate permission or acknowledgement.

Aston University

Collective Identity and Careers in a White Nationalist Forum

Amy Booth

Doctor of Philosophy

2023

Thesis abstract

The *career* is a sociological concept which explores engagement with a community over time. It has largely been discussed in an abstract way, making it unclear how the concept can be operationalised. Using an inductive, multimethodological research design, this thesis applies sociological and linguistic methods to explore the career in the context of a white nationalist discussion forum, moving beyond teleological radicalisation arguments into a more nuanced understanding of how white nationalist collective identity and authority are developed and expressed.

The study begins with a multiple correspondence analysis as a novel method of systematically reviewing existing literature on careers. Two defining characteristics of careers are identified. The *temporal* element, or amount of time dedicated to the community, is investigated with sequence analysis of over 120,000 user careers. The *behavioural* element, or the nature of individuals' engagement with the community, is explored using quantitative and qualitative corpus linguistic techniques, characterising language within and across career types.

Four career types are identified based on duration. These career types of different lengths are shown to correlate with the use of several linguistic features, including epistemic stance features and pronominal features. However, analysis of individual careers shows that usage of most of these features decreases across the career, suggesting that no features are specifically associated with the late stages of long careers. Instead, careers are marked by the accrual of symbolic capital in the form of new linguistic resources, allowing them to perform authority and collective identity in unconventional ways without facing negative social consequences.

The thesis shows that, despite a level of individual variation, typical careers are identifiable on both the temporal and behavioural levels, thus offering a new perspective on this key debate in the literature. The value of the *career* and *collective identity* concepts for studies of online communities is demonstrated.

Keywords: white nationalism, far-right, corpus-assisted discourse analysis, multiple correspondence analysis, sequence analysis, collective identity, career, linguistic capital

For my Nannan

I did it!

Acknowledgements

First and foremost, I am grateful to my supervisors, Tim Grant, Graeme Hayes, and Helen Newsome, for the many hours of support and advice they have given me throughout this process. Thank you for answering my questions, even the daft ones, and for reassuring me whenever I needed it, which was often.

Thank you, now and always, to Mum and Dad for the emotional (and financial!) support through the years, and for always encouraging me to pursue my goals. Sorry for being chronically unemployed.

Thank you to my wonderful partner, Danny, for cheering me on at every stage, and for looking after me so well in the last few frantic weeks of me pulling this thesis together!

A huge thanks to the PhD and postdoc community at AIFL, past and present, for all the friendly advice, reassurance, and laughs you have provided, especially when I was just starting out, mid-pandemic. Thanks especially to my various officemates for supplying hot drinks and gentle peer pressure when needed.

Finally, thank you to all my pals in Sheffield and beyond for keeping me sane and tolerating my various nonsenses. Thanks especially to Katie for actually reading the thesis – you're a star! ♥

Table of contents

Chapter 1	Introduction	9
1.1	The far-right and white nationalism	9
1.1.1	Terminology	10
1.2	Radicalisation	11
1.3	The career	14
1.3.1	Language and the career	15
1.4	Aims	15
1.5	Structure	16
Chapter 2	Language and identity	18
2.1	Introduction	18
2.2	Goffmanian origins	18
2.2.1	Identity as performance	18
2.2.2	Footing	20
2.3	The social constructionist view of identity	20
2.4	The resource-constraint model	22
2.4.1	Linguistic repertoires	24
2.5	Language and identity in the online extremist context	25
2.5.1	Language and identity online	25
2.5.2	Distinguishing in-group and out-group identities in extremist contexts	27
2.5.3	Divisions within the in-group identity	29
Chapter 3	The career	31
3.1	Introduction	31
3.2	Career contingencies	31
3.2.1	The career and identity	32
3.3	Describing the career	33
3.3.1	The idea of the typical career	33
3.3.2	The career over time	35

3.4	Language and the career	37
3.4.1	Newcomers	38
3.4.2	Experts	40
3.5	The career in the context of social movements	42
3.6	Collective identity	45
3.7	Symbolic capital	49
3.7.1	Beyond social class	52
3.8	Bridging the gap: The career in the white nationalist forum	55
3.8.1	Collective identities online	55
3.8.2	Language and collective identity	59
3.8.3	Research questions	60
Chapter 4	Methodology	62
4.1	Introduction	62
4.2	An interdisciplinary approach	62
4.3	Data	64
4.3.1	Data access	65
4.4	Research design	66
4.4.1	Multiple correspondence analysis	68
4.4.2	Sequence analysis	69
4.4.3	Corpus-assisted discourse analysis	71
4.5	Ethics	74
Chapter 5	How to study the career: A multiple correspondence analysis	77
5.1	Introduction	77
5.2	Identifying variables: A qualitative literature review	78
5.2.1	Method	78
5.2.2	Findings	79
5.3	Synthesising key dimensions of variation: A multiple correspondence analysis	86
5.3.1	Method	86
5.3.2	Findings	90

5.4	Discussion	96
Chapter 6	The temporal career: A sequence analysis	99
6.1	Introduction.....	99
6.2	Method.....	100
6.2.1	Data and data restructuring	100
6.3	Findings.....	104
6.4	Discussion	111
Chapter 7	The behavioural career I: A quantitative linguistic analysis	115
7.1	Introduction.....	115
7.2	Exploratory corpus analysis	116
7.2.1	Data and method	116
7.2.2	Findings.....	117
7.2.3	Discussion.....	123
7.3	Correlation of linguistic features with career type.....	124
7.3.1	Selection of features	124
7.3.2	Calculation of correlation coefficients.....	131
7.4	Case study: Five individuals	136
7.4.1	Method	136
7.4.2	Findings.....	137
7.4.3	Discussion.....	173
Chapter 8	The behavioural career II: A qualitative linguistic analysis.....	176
8.1	Introduction.....	176
8.2	Method.....	177
8.3	Observations by user	178
8.3.1	User 1.....	178
8.3.2	User 2.....	187
8.3.3	User 3.....	190
8.3.4	User 4.....	196
8.3.5	User 5.....	203

8.4	Discussion	211
Chapter 9	Discussion	215
9.1	Thesis summary	215
9.2	Reflections.....	217
9.2.1	Collective identity and symbolic capital	217
9.2.2	Applying the career concept	221
9.2.3	The interdisciplinary approach.....	224
9.3	Limitations and future directions	226
9.4	Summary of contributions	228
9.4.1	Chapter 5 – How to study the career: A multiple correspondence analysis...228	
9.4.2	Chapter 6 – The temporal career: A sequence analysis.....228	
9.4.3	Chapter 7 – The behavioural career I: A quantitative linguistic analysis	229
9.4.4	Chapter 8 – The behavioural career II: A qualitative linguistic analysis	229
9.4.5	Practical implications	229
9.5	Concluding remarks	231
References		232
List of appendices		259

Chapter 1 Introduction

In this thesis, I take a longitudinal approach to the study of members' engagement with an online far-right community, and in particular of how these members assimilate to a shared identity over time. Sociological research into social movements has produced useful frameworks for understanding engagement in ideological groups. This includes the *career*, which accounts for sustained engagement over time, and *collective identity*, which conceptualises the developing relationship between the individual and the group. However, these studies have typically focused on offline mobilization contexts, with consideration of how these processes might unfold in online social movement groups only beginning to attract sustained attention. This is despite the increasing role played by social media and other online platforms in social movements of all kinds, and perhaps particularly in extremist movements such as the far-right (Europol, 2023). Meanwhile, the field of linguistics has much to say about the negotiation and renegotiation of identity in online spaces. The disembodied nature of interactions in such environments places language firmly at the centre of engagement with online communities. In this study, I aim to demonstrate the value of a novel interdisciplinary approach, combining elements of sociology and linguistics to study the development of collective identity over time in the context of an online far-right forum.

1.1 The far-right and white nationalism

'Far-right' is an umbrella term referring to a wide range of related ideologies. As the various papers presented in Rydgren (2018) demonstrate, the far-right is a global phenomenon and takes different forms depending on the local sociopolitical context and grievances. In this thesis, I am concerned with the western anglophone context. In this context, the far-right is primarily concerned with biological essentialist racist ideologies, including antisemitism, Islamophobia, and anti-immigration sentiment. This involves a belief in the supremacy of white people and a commitment to the furthering of their interests, including the protection of white people from a perceived 'white genocide' which seeks to destroy the 'white race' through physical violence and encouragement of interracial families. The far-right typically displays regressive and binary views towards sex and gender, including various kinds of queerphobia and a belief in traditional roles for men and women. There is an anti-government strand to far-right ideology, including anti-tax sentiment, the defence of personal liberties and corresponding belief that governments attempt to curtail these, and a pro-gun, militarist, and survivalist stance (particularly in the United States). These ideologies are often couched in antisemitic conspiracy theories which position white people (particularly straight

cisgender men) as the victim of alleged Jewish organisations such as the ‘New World Order’, ‘Zionist Occupied Government’, or ‘globalists’ who have infiltrated world governments.

Far-right ideology is found in various different contexts, including political parties (even governments), social movement groups, and subcultures of various kinds (Mudde, 2019). Increased attention has been brought to the far-right in recent years through the electoral success of far-right politicians and parties across Europe, the U.S., and beyond. This has been accompanied by an increase in deadly far-right violent extremism across the world (e.g. Europol, 2023; Institute for Economics and Peace, 2023; Pitcavage, 2023). In the U.S., the large majority of extremism-related murders across the last few years have been carried out by far-right actors (Pitcavage, 2023). Aside from violence, the prevalence of far-right extremist ideology has “contribute[d] to a climate of fear and animosity against minority groups” (Europol, 2020, p. 4). This problem is compounded by the increasing normalisation of racist and queerphobic talking points in mainstream political discourse (Mondon & Winter, 2020).

Authorities are increasingly concerned about the presence of far-right actors in online spaces, as these are key sites for radicalisation (Europol, 2023). The far-right has maintained a significant online presence from the very earliest days of the internet (Levin, 2002), using it as a tool for recruitment and networking. However, the ubiquity of the internet in modern life, and the rise of ‘spreadable media’ such as memes (Mihailidis & Viotty, 2017), have allowed the far-right to further normalise and spread their message of hate to new audiences, especially young people (Finlayson, 2021). It is one such online space that is the focus of this thesis.

The data for this research is taken from one of the largest and longest-running far-right discussion fora on the internet, in operation since the mid-1990s. The present dataset contains linguistic and metadata from over 120,000 user accounts from the years 2001-2018. As I discuss in greater detail in §4.5, I have chosen not to name the forum in this thesis, in order to avoid publicly associating it with my own name and research. The site describes itself as a white nationalist community, and members are invited to discuss a range of political and cultural issues, typically through this racist lens. Although based in the U.S., the forum also attracts a large international membership. Members of the forum have perpetrated a number of high-profile instances of violent extremism and are responsible for numerous murders and other crimes. The forum is described in more detail in §4.3.

1.1.1 Terminology

So far in this thesis, I have used the term ‘far-right’ to refer to the community at the centre of my research. Terminology is, however, a problematic issue in the study of such a

community. As I have already noted, the ‘far-right’ is an extremely broad movement, comprising political parties and other less formal groups with a wide range of related, but distinct, views. As a result, a number of different terms – both umbrella terms and more specific labels – have been employed by researchers. Common terms include, but are not limited to: the *far*, *alt*, *ultra*, *radical*, and *extreme right*; *white nationalism*, *supremacy*, and *separatism*; *(neo) fascism* and *(neo) Nazism*; *right-wing extremism* or *populism*; and the *white power movement*. Often, different terms are given apparently the same meaning within the same work, and if distinct meanings are intended, these are not explained; Eatwell (2004) comments on the apparently synonymous use of various words, with multiple different terms sometimes appearing “in thesaurus-like ways” (p. 5) within the same text. Different authors often employ the same term as one another, but explicitly define it in different ways (Mudde, 1996).

The field’s preferred terminology has changed over time, following shifts and trends in thinking about the community (Mudde, 2019). Currently, ‘far-right’ is the most popular umbrella term used to refer to the movement. However, I have reservations about the use of this term. Firstly, users of the forum I am concerned with here self-identify as *white nationalists*. The representation of ‘unlikeable’ research subjects such as extremists in their own terms is a contentious issue (e.g. Pasiaka, 2019; Rüdiger & Dayter, 2017; van Wyk, 2013) and I recognise that other researchers may take a different stance to my own. However, as the present work aims to understand and reflect the identities of members of this forum, I believe it is most accurate to represent these as *white nationalist* identities. More generally, I have concerns that the term ‘far-right’ is too euphemistic, as it makes no explicit reference to the racial or nationalistic elements that are so central to the ideology of the community. In the remainder of this work, I therefore use the term *white nationalists* to refer to the forum community I explore. When referring to other researchers’ work, however, I continue to use the terms that they have chosen, under the assumption that their choice was purposeful and accurate to their focal community.

1.2 Radicalisation

In the context of white nationalism and elsewhere, extremism has often been understood through the lens of *radicalisation*. Radicalisation can be described as “an increase in and/or reinforcing of extremism in the thinking, sentiments, and/or behaviour of individuals and/or groups of individuals” (Mandel, 2009, p. 111). As the frequent use of *and/or* in this definition suggests, there is little scholarly consensus on the exact meaning of the term. Indeed, Nasser-Eddine et al. (2011, p. 13) go as far as to claim that “[a]bout the only thing that

radicalisation experts agree on is that radicalisation is a process” rather than a single key moment in time.

Definitions of radicalisation typically also refer to the development of increasingly extreme ideologies, behaviour, or, frequently, both. Traditionally, radicalisation is seen as the series of ideological developments which precede engagement in violence and even terrorism, positioning ideological and behavioural extremism as two stages of the same process. In more recent years, however, researchers have begun to argue for a clearer distinction between what McCauley and Moskalenko (2017) call the ‘two pyramids’ of radicalisation: *radicalisation of opinion* and *radicalisation of action* (see also Abay Gaspar et al., 2020; Borum, 2011). This view seeks to avoid assuming that terrorism is the natural endpoint of a process of radicalisation. It is well-known that not everybody who holds extremist beliefs carries out acts of violence. However, it is also true that not all those who carry out extremist violence hold extremist beliefs (Horgan, 2012). This is also true for recruitment to extremist organisations, which may or may not precede the development of an extremist ideology (Ahmad, 2016; Blee, 2002; Winter et al., 2020).

In this two-pronged understanding of radicalisation, then, the ‘endpoint’ of extremism is not necessarily violent action, and we must shift our understanding of what it means to be more extreme. McCauley and Moskalenko’s (2017) ‘two pyramids’ approach identifies two endpoints, which may or may not coincide in the same individual. The *radicalisation of action* pyramid progresses through legal and illegal actions and culminates in terrorism (violent action targeting civilians). Low down in the *radicalisation of opinion* pyramid we find those who are sympathetic to the extremist cause, while at its peak are those who not only believe that violent extremist action is justified, but that it is their “moral obligation” (p. 212). Other researchers have also found that more extreme ideology features complete dehumanisation for outgroups – in the white nationalist case, typically racialised groups – who are perceived solely as ‘the enemy’ (Borum, 2003). The outgroup is entirely blamed for the injustices the extremist perceives themselves as facing (Dalgaard-Nielsen, 2010). This dehumanisation can serve as a justification for violent action, creating psychological distance between perpetrator and victim (Moghaddam, 2005).

Traditional models of radicalisation have understood it as a linear phenomenon, whereby all extremists pass through the same set of stages in the same order (e.g. Borum, 2003; Moghaddam, 2005; Wiktorowicz, 2005). These models compare radicalisation to a ‘pyramid’ or ‘narrowing staircase’, with fewer and fewer people progressing to each subsequent level. Of course, these models also imply that the radicalisation process funnels extremists towards the same ultimate outcome of violence. More recent approaches have emphasised

the individual radicalisation process, recognising that “one size does not fit all” (Borum, 2011, p. 8). For example, McCauley and Moskalenko (2017) and Pilkington (2023b) both recognise that radicalisation is not necessarily unidirectional, and that individuals may move back and forth between ‘stages’ of radicalisation. A wide range of push and pull factors affecting individuals’ level of radicalisation have been identified (Junk et al., 2020; Schmid, 2013). These include individual, group, and society level factors, such as personal grievances, pressure from social networks, and the sociopolitical climate (McCauley & Moskalenko, 2008).

Online spaces represent a key security concern, particularly in the context of right-wing extremism (Europol, 2023). Extremist ideas can spread quickly online across a range of platforms and therefore reach new audiences easily, introducing extremist views to some and intensifying them for others. In the U.K., the majority of convicted extremists were at least partially radicalised online (Gill et al., 2017; Kenyon et al., 2021). Rates of (fully or partially) online radicalisation in the U.K. have increased every year since at least 2007 (Kenyon et al., 2021), particularly since the Covid-19 pandemic (Kenyon et al., 2022).

However, despite the value of the radicalisation concept to the study of extremism, I will not be using this as a framework for my own study of engagement over time with a white nationalist forum. This is for two primary reasons.

First, despite some recent focus on “partial, stalled or non-radicalization trajectories” (Pilkington, 2023b, p. 3), there is an implicit assumption in much of the literature that radicalisation to violent outcomes is the natural endpoint of the process (see also Abay Gaspar et al., 2020). This is the case even for models which claim to separate radicalisation of belief from radicalisation of actions. For example, the peak of McCauley and Moskalenko’s (2017) *radicalisation of opinion* pyramid is the feeling of a “personal moral obligation to take up violence in defense of the cause” (p. 212). A model which ends in terrorist violence may be seen as ‘complete,’ accounting for every possible outcome, in a way a model of a ‘stalled radicalisation’ would not. For this, among other reasons, researchers of radicalisation have typically studied cases which have ended in violence (Pilkington, 2023b). As a result, most research on radicalisation is teleological. Early stages of the radicalisation process are implicitly understood as precursors to the eventual violent or terrorist act. It is, therefore, unclear how far our understanding of radicalisation can be applied to cases which have not, and may never, culminate in a terrorist act. In this study, I do not specifically focus on forum members who are known to have carried out violent extremist action. In Pilkington’s (2023b) terms, users in my dataset may follow pathways of *partial, stalled, or even non-radicalisation*.

Secondly, although the internet has been strongly linked to radicalisation, “[r]adicalisation processes rarely take place in either the online domain or the offline sphere exclusively” (Mutton et al., 2023, p. 4), leading Gill et al. (2017) to describe the idea of online and offline radicalisation as a “false dichotomy” (p. 114). It would therefore be inappropriate to treat my study of forum data as a study of radicalisation. Even assuming that some users can be said to have been radicalised, I would be unable to offer a full account of their radicalisation based only on their online activity on a single site.

For these reasons, I argue that the literature on radicalisation does not capture typical engagement with the white nationalist forum context, and consequently I have chosen not to apply the lens of radicalisation per se to this data. Another approach to understanding engagement with the forum over time is therefore required. For this, I use the concept of the *career*.

1.3 The career

As I describe in more detail in Chapter 3, the career is a way of understanding how individuals progress through a set of roles or stages in a given social environment. This environment can be, but is not necessarily, an extremist or other activist group (e.g. Fillieule, 2010). While radicalisation can be considered a kind of career, the career framework is much more wide-ranging and flexible. It allows for the study of processes beyond those which end in extremist violence, and of non-linear careers, such as those Pilkington (2023b) has recently begun to describe in the radicalisation literature.

The career may imply a set of objective changes in role, such as job titles in an occupational career. However, it also implies a set of subjective changes, corresponding shifts in how the individual perceives themselves in relation to others around them. The career is therefore a useful framework for investigating shifts in identity over time. In this study, I use the career framework to understand the development of an activist collective identity (Melucci, 1996; Taylor & Whittier, 1992), which has rarely been explored in an online environment (see §3.8.1).

The objective of this thesis is therefore to explore the career in an online white nationalist forum. In particular, I aim to investigate if there is a *typical* career in the forum. This represents a key fault line in the literature on careers (§3.3.1). Goffman (1961) and Becker (1973) take a structural approach, identifying career stages that are “importantly similar” (Goffman, 1961, p. 129) across individuals in a given context. On the other hand, Fillieule (e.g. 2001) argues that careers are highly individual, and that the structural approach therefore treats the career in a meaninglessly abstract way.

In Chapter 5, I develop a framework to operationalise the career, which has remained a largely theoretical concept in the literature. I identify two main elements which are important for the study of careers in any environment: the *temporal* element, and the *behavioural* element. The temporal element is primarily defined in terms of the duration of members' forum careers. As I outline in Chapter 5, the behavioural element can be interpreted in a context-specific way, according to the behaviour types relevant to the particular social environment. In an online context like the present one, however, these behaviours are largely linguistic in nature. In subsequent chapters, I apply this two-part framework to explore careers in the white nationalist forum in terms of their temporal and behavioural aspects, and consider whether my findings support an individualistic or a structural view of the career.

1.3.1 Language and the career

In order to explore the career in this white nationalist forum, I will combine theories and methods from both sociology and linguistics in a way that breaks new ground, both because of its interdisciplinary nature and its novel use of a large, longitudinal dataset.

As I have highlighted above, the career can be understood as a question of identity change over time. Linguistic analysis is therefore well-placed to offer a new perspective on, and tool for, the study of the career. As I detail in Chapter 2, a significant strand of linguistic theory and research has been dedicated to the linguistic construction and renegotiation of identity, including in online contexts such as discussion fora. Online, language is seen as particularly important for the construction of identity (Sergeant & Tagg, 2014; Tagg, 2015). Participants in online environments can choose to conceal their offline identity and may be functionally anonymous in their interactions. This is typically the case in the white nationalist forum at the centre of this thesis, as the extreme and criminal-adjacent nature of the community means that members may face social or even legal sanctions if their offline identities were to be linked to their forum participation. Therefore, identity is largely conveyed and understood through linguistic performance. In the framework of the career I develop here (Chapter 5), language is therefore also central to the study of the *behavioural* element of the career. As I describe in §5.2.2.9, behaviour in the context of the career may be construed broadly and include various different kinds of behaviour. However, in the text-based environment of this forum, behaviours of all types are enacted linguistically, as this is the only resource members have with which to engage with the community.

1.4 Aims

On the broadest level, the aim of this thesis is to shed light on the identity construction processes that take place on a large online white nationalist discussion forum, and how these processes change across users' careers. This is with a view to improving our

understanding of these environments and the white nationalist extremists who frequent them, in order to better tackle the rise in right-wing extremism and violence (Europol, 2023).

I aim to make a significant new contribution to the literature on careers by providing a framework for the operationalisation of this heretofore largely abstract concept. I will use this framework to empirically explore the nature of the career. Do the careers in this forum support the structural (Becker, 1973; Goffman, 1961) or the individual (Fillieule, 2010) view of the career? In other words, is there evidence of a typical career in this forum? Using the two-part framework I develop in Chapter 5, I aim to explore these questions from two perspectives, the temporal and the behavioural.

In exploring both temporal and behavioural elements of the career, I hope to demonstrate the value of an interdisciplinary approach to the concept. More specifically, I bring a linguistically-informed understanding and analysis of identity construction to sociological concepts of identity, namely: the development of a *collective identity* in social movement contexts; and the *career* as a process of identity change through involvement in a community. In doing so, I bring the two disciplines of sociology and linguistics into productive collaboration in a novel context.

These aims will be aided by the use of a unique dataset (see §4.3) which is well-suited to the study of careers from both a temporal and a behavioural perspective. The dataset offers a complete record of the careers of some 120,000 white nationalist forum users, providing a complete set of linguistic and metadata from posts made to the forum between the years of 2001 and 2018. This allows me to make both broad and fine-grained observations about engagement with the forum, considering both the individual and the community level. This flexibility is valuable on a theoretical level, allowing me to consider the implications of the data for both the structural and individual perspectives on the career. On a practical, impactful level, the dataset also offers a significant insight into the online activity of a large white nationalist community. As I have established above (§1.2), the online environment remains a key security concern due to its prominent role in right-wing extremist radicalisation and violent attack planning (Europol, 2023; Mutton et al., 2023). The work presented here may therefore have implications for both theory and policy.

1.5 Structure

The thesis proceeds as follows.

In Chapters 2 and 3, I provide an overview of the literature on several relevant concepts from the disciplines of sociology and linguistics, including the role of language in identity (re)negotiation, the career, and collective identity. I highlight the conceptual overlap between

these literatures, particularly concerning the nature of identity as a socially constructed and negotiated concept, and show how these can be brought together into productive tension in order to provide new perspectives on both disciplines. In Chapter 3, I also present my research questions.

Chapter 4 outlines the research design for this thesis. As a novel interdisciplinary project, the research design follows a multimethodological approach which was developed inductively, with the findings of each study informing the design of the next. Chapter 4 narrativizes and justifies each step in this process. In this chapter, I also describe my dataset in greater detail, and address ethical considerations relating to the project.

Chapter 5 is the first empirical chapter. Here, I develop a framework for the operationalisation of the abstract career concept. I employ the novel method of multiple correspondence analysis (Le Roux & Rouanet, 2010) to systematise a highly disparate literature which explores careers across a range of contexts. I identify the *temporal* and *behavioural* elements of the career, which I explore in turn across the subsequent empirical chapters.

In Chapter 6, I employ a sequence analysis (Gabadinho et al., 2011; Gabadinho et al., 2010) to explore the *temporal* element of forum careers, particularly focusing on the duration and frequency of users' engagement with the community. I identify four key types of users based on the duration of their careers.

In Chapter 7, I take a large-scale quantitative corpus linguistic approach to explore the linguistic *behaviour* of users belonging to the four career types identified in Chapter 6, identifying linguistic features which correlate with career length. I then take a more detailed approach, quantitatively investigating the use of these features across the careers of a selection of individual users.

Chapter 8 continues the detailed linguistic study of *behaviour* across the career, this time taking a qualitative approach to investigate the discourse of a selection of users at the beginning and end of their careers.

Finally, Chapter 9 discusses the implications of the four empirical studies and their relation to existing theory and research, addressing the research questions identified in Chapter 3. I summarise the contribution of the thesis and consider directions for future research.

Chapter 2 Language and identity

2.1 Introduction

The strong link between language and the performance of identity is by now widely recognised across many disciplines (De Fina et al., 2006). In this chapter, I trace this idea from its sociological origins in the work of Goffman, showing how the idea has evolved over time into social constructionist and resource-constraint models of identity, before exploring how this concept has been applied in extremist contexts.

2.2 Goffmanian origins

2.2.1 Identity as performance

Goffman (1956) laid the groundwork for our understanding of how linguistic expression indexes identity through his metaphor of social interactions as dramaturgical performances. Goffman proposed that in social situations, the individual consciously or subconsciously presents themselves in particular ways through behaviour and discourse, in order that others – their *audience* of interlocutors – will gain a particular impression of them. The nature of the performed identity that the individual employs in a given interaction is dependent on contextual factors including the audience. They will act in the way they believe others expect them to, attempting to “glean clues” (Goffman, 1956, p. 1) from their interlocutor’s own behaviour and from their pre-existing knowledge of their interlocutor, as well as their own previous interactions which have taken place in similar contexts with interlocutors in similar social roles. In doing so, individuals attempt to ensure that the interaction runs smoothly, to prevent discomfort, embarrassment, or annoyance on the part of themselves or their interlocutors. This protects what Goffman calls the *face needs*, or reputational integrity, of both parties; in other words, both are able to retain the “positive social value a person [...] claims for himself” (Goffman, 1967, p. 5) which they attempt to maintain interactionally.

In summary, then, individuals “convey an impression to others which it is in his interests to convey” (Goffman, 1956, p. 3). This is true for all interlocutors, and so creates a state of *mutual monitoring*, which refers to each interlocutor’s awareness that they are “accessible to the naked senses of all others who are ‘present,’ and similarly find them accessible to him” (Goffman, 1964, p. 135). Each interlocutor therefore adapts their own behaviour to present themselves in particular ways to their audience. For Goffman, mutual monitoring is, therefore, key to all interactions, creating the situational norms that provide a model for the unfolding interaction (Goffman, 1963). Interlocutors co-construct the dramaturgical ‘scene’ and their respective expectations for it, and present themselves accordingly. Context is of crucial

importance to such performances, with no two situations demanding exactly the same identity presentation. Similarities which hold across all contexts are minimal, as Goffman (1963) summarises:

The rule of behavior that seems to be common to all situations and exclusive to them is the rule obliging participants to 'fit in.' [... The] individual must be 'good' and not cause a scene or a disturbance [...] he must keep within the spirit or the ethos of the situation (p. 11)

What counts as 'fitting in', however, varies across contexts; the styles of interaction expected on a date, for example, are very different from those expected in a business meeting. In each case, the individual must carefully design their talk and behaviour to ensure it is appropriate to their audience and the wider context.

This should not be understood to mean that all performances are consciously calculated. Certain performances – those which we understand to be 'neutral' or most situationally appropriate, especially in common contexts – become conventionalised in particular societies and are carried out almost unthinkingly by most members of that society. This links with individuals' awareness of context, such as their interlocutors' role or social status; individuals draw on their previous experiences with other, similar interlocutors as a model for how to act in new situations. While individuals must control their behaviour to be within social norms, this often feels natural and non-performative. Rather than being truly natural, however, such performances are simply the result of learned social norms. Repeated performances in the same contexts are the basis on which social relationships are developed – on which individuals come to know each other's identities.

Some of the most natural-feeling, unselfconscious performances are those which occur in what Goffman (1956) describes as the *backstage* or *back region*. Where multiple participants co-operate in a performance to an audience, this can be understood as the *frontstage*, so-called because such performances are more similar to the layperson's understanding of a stage performance, being sustained through careful planning and co-ordination. Frontstage performances have a corresponding backstage region where the frontstage performance is prepared, and "where the performer can reliably expect that no member of the audience will intrude" (Goffman, 1956, p. 70). Despite the terminology, backstage interactions should still be understood to be performances, as they involve interaction between individuals with access to the backstage region. Backstage interlocutors still engage in mutual monitoring to ensure that their own and others' behaviour is situationally appropriate. The difference is that in the back region, "a hide-out where certain standards need not be maintained" (Goffman,

1956, p. 76), a wider range of behaviours are likely to be acceptable. Therefore, backstage performances are likely to feel more natural and unconstrained.

2.2.2 Footing

In his later work, Goffman (1981) began to take a more explicitly linguistic approach in his understanding of how people establish their social relations with others. He did so through the concept of *footing*. Footing is the way an individual aligns their “projected self” with their audience, “as expressed in the way we manage the production or reception of an utterance” (Goffman, 1981, p. 128). While his earlier work focused on the ways individuals change their performances to match the particular conversational context they found themselves in, in his discussion of footing Goffman is clear that changes of footing are a “persistent feature of natural talk” (Goffman, 1981, p. 128) and may take place several times during a single short exchange. This results in the performance of highly context-dependent *microidentities*.

Goffman, then, realised that individuals are able to make their own position in relation to the ideas they express known through the language they use to express those ideas. In turn, this allows their interlocutors to understand not only an element of the speaker’s (or writer’s, although Goffman’s work concerns the face-to-face context) performed identity, but also how that identity position relates to their own. Goffman mentions several specific linguistic devices through which this self-positioning may occur. As Goffman highlights, when an individual produces a so-called ‘bare’ or unhedged assertion, it will normally be understood as “representing in some direct way the *current* desire, belief, perception, or intention of whoever animates [speaks] the utterance” (Goffman, 1981, p. 147) – that is, as representing the speaker’s own position. Speakers can manipulate this using various kinds of hedging to create distance between the proposition they are expressing and their own professed viewpoint. Goffman cites modal verbs as a key hedging resource. He also notes that the use of the third person to attribute propositions to other individuals can serve to distance the speaker from the proposition. In this way, Goffman demonstrates how individuals are able to use language to construct their own viewpoints and performances of identity.

2.3 The social constructionist view of identity

Following Goffman, a philosophy of identity known as *social constructionism* has developed which links the idea of the individual’s differing (linguistic) performances across contexts more clearly with identity (Graham, 2016). Self-evidently, the central idea of social constructionism is that identity is socially constructed – that is, negotiated and renegotiated through interaction with others. Under this view, identity can, therefore, be understood as wholly relational, requiring both a performer and an audience, and is defined as “the social

positioning of self and other” (Bucholtz & Hall, 2005, p. 586) or, in plainer terms, “who we are to each other” (Benwell & Stokoe, 2006, p. 6).

Because identity is interpersonal, it is enacted through discourse (Halliday & Matthiessen, 2004). Individuals do not simply express ideas in a vacuum; they are always speaking with or writing to some interlocutor. As a result, the individual’s discourse does not simply position themselves with relation to the proposition(s) they express. By expressing their attitudes towards that proposition *to an audience*, the discourse also serves to align or disalign the individual with that audience, depending on the audience’s own stance(s) on the proposition. Even language use that is not directly interactional serves to express the producer’s identity using the same mechanism, as their discourse positions them in relation to some imagined interlocutor (Bakhtin, 1981). For Bakhtin, all discourse is dialogic; the individual is familiar with existing discourses around a topic, and – whether explicitly or implicitly – positions themselves in relation to these discourses. Meanwhile, the individual uses their knowledge of these existing discourses to anticipate responses to their own contribution, and to construct their discourse accordingly, for example by opening up to, or closing off from, alternative viewpoints (Martin & White, 2005). The notion that the individual positions themselves in relation to both the proposition expressed, and consequently to their audience, is familiar from Goffman’s (1981) discussion of footing. This ‘triangle’ – from producer, to proposition, to receiver – is further elaborated and formalised by Du Bois (2007). The triangle is connected on all three sides in directly interactional contexts, where the original receiver responds to the original producer’s proposition, thus signalling their own (dis)alignment with the original producer with respect to the proposition. This reciprocity highlights the Goffmanian and social constructionist view that identity construction is a mutual endeavour, and that, as a result, identity is inherently relational.

The inherently interpersonal nature of identity implies another key element of the social constructionist view: that identity is to be understood as a social action, rather than a psychological construct. Because individuals negotiate, consciously or subconsciously, their behaviour in particular interactional contexts, identity cannot be understood as ‘passive’ or ‘received’. De Fina et al. (2006) conceptualise these self-positionings as ‘identity claims’, social acts “through which people create new definitions of who they are” (p. 3). These identity claims are constantly adapted to suit the interactional context and the identity the individual seeks to perform.

This view of identity as a carefully selected social action stands in contrast to more traditional, essentialist views. For social constructionists, there is no ‘essential’ or underlying identity which exists behind the discursive performance. Identity is not reflected in discourse,

but rather constructed solely through it (Benwell & Stokoe, 2006). As Antaki and Widdicombe (1998, p. 3) summarise, “the force of ‘having an identity’ is in its *consequentiality* in the interaction”: that is, identity is made relevant only by the interaction and the effect it has on the interaction, and has no meaning beyond it. Therefore, identity is to be understood properly not as a property of an individual *per se*, but as a property of social situations, and the way individuals act within them (Bucholtz & Hall, 2004). Contrary to popular understandings, the same is true for categories of social identity such as race and gender. While these *macroidentities* are commonly seen as natural groupings which pre-exist interaction, even as being based in biological reality, various researchers have demonstrated that these categories are in reality largely constructed discursively through language and other semiotic resources. This has been demonstrated for, among other social categories, gender (Bucholtz, 2007; Butler, 1990; Cameron, 1999; Eckert, 1996; Eckert & McConnell-Ginet, 1992; Pichler, 2009; Preece, 2009); race, ethnicity, and nationality (Doran, 2004; Lin & Kutoba, 2011; Lytra, 2016; Pichler, 2009; Shin, 2014; Walton & Jaffe, 2011; Wodak et al., 2009); social class (Block, 2016; Doran, 2004; Pichler, 2009; Shin, 2014; Walton & Jaffe, 2011); and age (Andrew, 2016; Bucholtz, 2007; Eckert, 1996; Rickford & Price, 2013). As the overlapping lists of references above show, individuals can perform multiple intersecting identities at once. Likewise, because identity is not fixed or essential, individuals can perform different identities in different contexts. These performances may be conscious – an individual can intentionally exploit resources to give their audience a certain impression of them, for example to deceive or to impress – or unconscious, as is typically the case for the performance of the social categories mentioned above.

2.4 The resource-constraint model

More recently, Grant and MacLeod (2020) have developed a new model of the link between language and identity, the *resource-constraint model*, building on the earlier *resource model* proposed by Johnstone (1996, 2009). This model aims to address a set of apparently contradictory observations. First, as established above, individuals are able to perform a wide range of macro- and micro-level identities both within and across contexts. However, theory and evidence from the forensic linguistic study of idiolect and authorship analysis shows that some linguistic elements of the individual’s performance may be constant across different situations (e.g. Coulthard, 2004; Grant, 2013, 2022; Nini, 2023; Wright, 2013, 2017), suggesting that some elements of identity, if not actually intrinsic to the individual, are nonetheless difficult to ‘leave behind’, even when attempting to disguise one’s identity. Additionally, despite the flexibility of possible identity performance, individuals may find themselves unable to perform appropriate identities in some, especially unfamiliar, contexts.

In other words, while individuals can perform many identities, this performance is not unbounded. Grant and MacLeod's (2020) model of identity accounts for these tensions.

Grant and MacLeod (2020) propose that individuals have a large repertoire of possible elements of identity performance available to them (see §2.4). These elements can be understood as *resources* for identity performance. These resources are derived from multiple sources. The first of these, which is primarily relevant in face-to-face contexts, is the individual's physicality. For example, an individual is more likely to be able to successfully perform a particular gender identity if their appearance and voice correspond closely with societal norms for that identity. As a result, individuals may choose to make changes to their physicality in order to perform gender identities in a more normative way.

A second source of identity resource is the individual's sociolinguistic history. This includes the dialectal and other sociolectal varieties of language that we are familiar with, and so relates to the social categories (e.g. gendered and racialised categories) that individuals see themselves, and are seen by others, as belonging to. Working within a social constructionist viewpoint, Bucholtz and Hall (2004, p. 376) suggest that "identities are not entirely given in advance but are interactionally challenged". As Grant and MacLeod (2020) highlight, 'entirely' here implies that some elements of identity pre-exist the specific interaction, an apparent contradiction to the social constructionist view of social identity categories. Grant and MacLeod (2020) address this problem by positioning elements of social identities as previously acquired *resources* on which individuals can draw to perform that identity in a given situation. Resources related to physicality and sociolinguistic history, which (as discussed for the case of physical gender presentation) are changeable over time but generally resistant to overnight change, may contribute to the element of individual, idiolectal stability mentioned above. From a Goffmanian perspective, the persistence of some identity features is interactionally beneficial. As Grant and MacLeod (2020) highlight, such features "allow others to predict our interactions with them, which in turn creates the possibility of smooth social interaction" (p. 24) and, therefore, protects the face needs of all participants.

The third and final source of identity resources is the interactional context itself. This returns to the Goffmanian view that the individual has expectations of the appropriate way(s) to perform in a particular context based on their previous experiences of interactions in similar contexts. The individual can also respond to the developing interaction, changing their self-positioning – or *footing* – throughout to ensure the interaction progresses smoothly.

However, individuals do not have infinite identity resources. The set of resources they *do* have also functions as a constraint, as the individual is not able to perform identities for which they do not have the pertinent resources. Most obviously, an individual who has never

learned French will be unable to draw on this resource to conduct an interaction and create an identity performance in French. The same principle also applies to more subtle resources; individuals would be unable to perform convincingly as a member of any social group without having the relevant resources, whether that group is a high-level social category, such as a racialised or gendered group, or a more local group, such as a family, a workplace, or a club of train enthusiasts. Individuals are able, as I have already intimated, to *acquire* resources, which allows them to perform new identities. Grant and MacLeod (2020) cite the community of practice framework (Lave & Wenger, 1991; Wenger, 1998; Wenger et al., 2002) as a model for how the individual can acquire identity resources. Crucially, this acquisition is not immediate, but developed through sustained effort and interaction with others engaging in similar activities and performances.

2.4.1 Linguistic repertoires

Resource-constraint theory shares strong similarities with the concept of linguistic *repertoires* (e.g. Gumperz, 1964). As the name suggests, the repertoire refers to the “range of possible expressions [... which are] accepted ways of formulating messages” (p. 137-138) across the individual’s languages, and, therefore, the “arsenal” (p. 138) of linguistic resources from which individuals choose. The repertoire, therefore, represents “a space both of restrictions and of potentialities” (Busch, 2012, p. 509) for individuals’ linguistic performances, much like the resources and constraints Grant and MacLeod (2020) describe. The resource-constraint model differs from Gumperzian repertoires in its explicit focus on how individuals use resources to create identity performances. Gumperz takes a less individual-centred view of his repertoire concept, using it to account for individuals’ knowledge of how to behave in a linguistically appropriate way *according to the norms of the communities they belong to*. The Gumperzian repertoire is, therefore, those linguistic resources “that people know how to use *and why* while they communicate” (Blommaert & Backus, 2013, p. 11; emphasis my own), placing a greater focus on the influence of the community and its norms than do Grant and MacLeod. Indeed, Gumperz later offers an explicitly communal definition of the *local linguistic repertoire*, or “the totality of distinct language varieties, dialects and styles employed in a community” (Gumperz, 1982, p. 155). Such repertoires, of course, feed back to the individual level, as the individual draws on those communally available variants for their own communication. In this form, the concept of the repertoire has echoes in the sociological literature on social movements, as we will see in more detail later (§3.7.1.2).

2.5 Language and identity in the online extremist context

2.5.1 Language and identity online

The nature of social media interaction clearly lends itself to a social constructionist view of identity (Seargeant & Tagg, 2014; Tagg, 2015). I here define *social media* in the broad sense, following Leppänen et al. (2014), as any and all “online environments which enable social interactions” (p. 113). This includes, but is not limited to, the social networking sites which are often considered synonymous with the term *social media*.

Social media users’ reliance on language and other semiotic resources for the construction and negotiation of their identities is clearer than ever in online contexts, where interactants may be otherwise entirely anonymous. Computer-mediated discourse (CMD) is typically *disembodied*, divorced from the extra- and paralinguistic information which usually provide additional clues about the speaker or writer’s identity, such as physical appearance, accent, handwriting, age, gender, facial expression, and tone of voice (Seargeant & Tagg, 2014; Tagg, 2015). This (at least theoretically) allows individuals a greater degree of freedom in their identity performance, as their performance is no longer affected by these constraints. boyd (2008) captures this phenomenon neatly with her claim that individuals *write themselves into being* online. Unsurprisingly then, studies of identity construction have been central to recent work into CMD (Barton & Lee, 2013).

In the early days of the internet and social media, there was an assumption that this high degree of anonymity would lead to an egalitarian online environment, with sources of offline social inequality being invisible and irrelevant online (Herring, 2001; Tagg, 2015). However, researchers soon showed that this was not to be the case. For example, Herring has extensively demonstrated that gendered differences in discursive practice are reproduced online, as participants “give off” gender (Herring, 2001, p. 621) due to the pressures of social conditioning learned offline (e.g. Herring, 1996; Herring, 2003; see also Herring & Stoerger, 2014 for a review). Even when linguistic identity disguise is attempted, this may not be entirely successful. While some resources can be consciously learned and deployed with relative ease, others are drawn on more subconsciously and may not be so easily disguised (Ford, 2021; Grant & MacLeod, 2020).

Nowadays, much of the internet is far less anonymous than it once was. Social media is increasingly integrated with individuals’ offline realities (Barton & Lee, 2013; Blommaert, 2018; Herring & Androutsopoulos, 2015; Tagg, 2015, 2020; Thurlow & Mroczek, 2011). In particular, social networking platforms like Facebook “work on the basis of existing relationships” (Tagg, 2015, p. 61), and users share aspects of – including information about and photos from – their offline lives (Tagg, 2020). CMD also plays a major role in the

workplace and education (Blommaert, 2018). The use of mobile devices to access social media platforms means that individuals are 'always on' (Baron, 2008), which further increases the integration of CMD and computer-mediated identity performance with everyday offline life. Online identity performance can therefore be understood simply as a "prosthetic extension" (Thurlow & Mroczek, 2011, p. xxv) of individuals' identity performance repertoire, rather than a wholly separate entity.

Nonetheless, individuals can be more selective about the elements of their identity performance which they choose to (de)emphasise in particular online contexts. The disembodiment of online communication means that individuals can shed unwanted or irrelevant aspects of their identities more easily and completely. These identity performances can be extremely context specific. Zappavigna (2011) highlights how Twitter users form temporary "communities of shared value" (p. 789) around unfolding news stories by engaging with trending hashtags. This is an example of what Blommaert (2018) would call a *light* community, which he defines as "smaller social sub-groups [...] of a] relatively superficial and ephemeral character" or "aggregations of people sharing just the rules of the encounters, but little beyond it" (p. 65). These light identities contrast with the *thick* identities which are often focused upon in sociological and sociolinguistic studies and which are traditionally (and still largely) formed offline, such as "race, gender, class or ethnicity [...] the state and the family, the guild and the church" (Blommaert, 2018, pp. 63-64). For Blommaert, social media facilitates the proliferation of light identity performances, allowing individuals to flit between identities in previously unprecedented ways.

Even in today's world of integration between the online and offline, there remain environments wherein a level of identity concealment is desirable to discourse participants. Primary among these are criminal contexts. For example, Chiang (2018) demonstrates how one child sex offender's quest for anonymity led him to perform at least seventeen different identity personae when interacting with different victims. Anonymity may also be desired by political dissidents, whistleblowers, and activists of all stripes who wish to avoid repercussions (Tagg, 2015). Members of the forum explored in this thesis are likely to seek anonymity both due to the hateful and potentially criminal nature of their discourse, and because of a fear of being targeted by a government they mistrust.

Paradoxically, even in anonymous environments, 'authentic' identity performances are highly valued (Tagg, 2015). Performances perceived as authentic are those in which "online behaviour [is] seen to consistently relate to an online persona [...] which in turn [is] seen to have parallels with who [the individual is] offline" (Tagg, 2015, p. 226). In other words, the individual is perceived as being who they say they are. This is especially crucial in criminal

or dissident contexts where identity disguise is practiced, as these are “low trust environments” (Grant & MacLeod, 2020, p. 167) where members feel at risk of infiltration by hostile law enforcement or government actors. Authenticity is also valuable as a resource for individuals to be perceived as powerful, authoritative, or influential. Newsome and Grant (in press) define power in such online contexts as the ability to “persuade [others ...] to listen to their point of view and potentially to act in a particular fashion” (p. 3). As Tagg (2015) points out, this power derives from an individual’s ability to “convince others that [their] experience and authority are genuine and based on real-life experience” (p. 151). This is reflected in Newsome and Grant’s (in press) findings on the resources employed by members of three online forum communities (both criminal and mainstream) to perform powerful identities. These include accounts of offline personal experience, reference to formal qualifications and accreditations, and reference to long-term membership of the community as evidence of a long history of community-specific experience and knowledge. Due to the value of authenticity, a perceived lack of it can, therefore, be face-threatening in both criminal and mainstream environments (Blommaert & Varis, 2013; Page, 2014) and is therefore undesirable.

As I will return to below (§3.8.1), some scholars have emphasised the importance of “[p]hysical copresence” (Collins, 2004, p. 23) for the development of collective or group identities, suggesting that computer-mediated contexts will always be “weak substitutes” (p. 62). However, there is evidence that at least some users of online extremist fora feel a strong sense of group identity and connection (De Koster & Houtman, 2008). Indeed, Blommaert (2018) argues that the internet provides “an ocean of ‘how to’ resources” (p. 70) for the performance of all kinds of light identities. Frequently, these identities have been developed entirely online and may have made their way offline, rather than the other way around.

2.5.2 Distinguishing in-group and out-group identities in extremist contexts

I turn now to the construction of identity in the context of extremist groups, with a particular focus on white nationalist and other far-right groups (see discussion of terminology, §1.1.1). The most important and commonly cited element of this identity construction is the distinction between the in-group and the out-group. This is often explicitly indexed through use of the pronouns *us* and *them* (e.g. Adams & Roscigno, 2005; Anahita, 2006; Baumgarten, 2017; Bowden, 2008; Brookes & McEnery, 2020; Douglas et al., 2005; Duffy, 2003; Perry & Olsson, 2009; Zickmund, 1997). From a social psychological perspective, Tajfel and Turner (2004) have theorised the importance of this key boundary to group identity formation in all social contexts, as it helps individuals to conceptualise the social

world and their place within it. However, the distinction becomes even more salient in the case of extremist groups (Hogg, 2012; Hogg & Adelman, 2013). Here, strong identification with a well-defined group identity can provide a model for individual members' identities, leading to *depersonalization* of the individual as their own identity is 'replaced' by that of the group (Hogg, 2012).

The in-group is often demarcated through frequent reference to the out-group. As Anahita (2006) highlights, this focus on the out-group indirectly serves to define the in-group identity by making it clear what the in-group is *not*. For example, in her study of a far-right forum, Baumgarten (2017) identified 188 different outgroup labels in a dataset of just 113 posts. Lorenzo-Dus and Nouri (2021) share similar findings of a "clear boundary demarcation between homogeneous, victimised in-groups and multiple, blameworthy out-groups" (p. 417). Brindle (2016) also notes how, by describing the alleged negative characteristics of gay men, the (male) white supremacist in-group is able to position itself as a totally opposite group. Across all these studies, the representation of out-groups is almost wholly negative. This is also the case in the discourse of violent jihadists (Baker et al., 2021) and extreme misogynists (Krendel, 2020), where out-groups are represented as unanimous collectives in order to dehumanise them and reduce them to their alleged negative characteristics.

Notably, these out-group collectives are often presented as a threat to the wellbeing or even existence of the in-group (Adams & Roscigno, 2005; Baker et al., 2021; Bostdorff, 2004; Brindle, 2016; Douglas et al., 2005; Duffy, 2003; Krendel, 2020; Krendel et al., 2022; Lorenzo-Dus & Nouri, 2021; Zickmund, 1997). In line with van Dijk's (1998) 'ideological square', while out-groups are represented negatively and their positive qualities ignored, in-groups on the other hand are represented positively and their own negative qualities ignored (Baker et al., 2021; Brindle, 2016; Douglas et al., 2005; Lorenzo-Dus & Nouri, 2021), providing the group with *positive social identity* and self-image (Tajfel & Turner, 2004). While out-groups are typically collectivised and represented as 'all the same', more individuality may be afforded to in-group members (Krendel, 2020; Krendel et al., 2022), serving to humanise this group. The out-group often receives more attention than the in-group, leaving the out-group identity more clearly defined than that of the in-group (Bowden, 2008; Krendel, 2020).

The boundary between the in- and out-group is strengthened by presenting the difference between the two as a biological fact, resulting in an essentialised and impermeable boundary. This serves in turn to increase the entativity or 'groupness' of the group (Lorenzo-Dus & Nouri, 2021) and strengthen feelings of identity within it. Across extremist groups, biological essentialist positions are taken towards race (Duffy, 2003; Lorenzo-Dus & Nouri,

2021; Perry & Olsson, 2009), gender (Heritage & Koller, 2020; Krendel, 2020; Krendel et al., 2022) and sexuality (Brindle, 2016). This essentialism is enacted linguistically by representing out-groups in nominal rather than adjectival forms (compare *gays* and *Jews* with *gay men* and *Jewish people*; Brindle, 2016). Relatedly, identities may be solidified by associating them with other identities that are tied to physical locations. Lorenzo-Dus and Nouri (2021) show how writers of an alt-right blog characterise U.S. citizens as being bound together by 'blood and soil' (borrowed from the Nazi idea of 'Blut und Boden'), anchoring essentialised white 'blood'-based identities to a physical location. Likewise, in the most extreme violent jihadist texts, "religious identity [is] intertwined with national identity" (Baker et al., 2021, p. 72) by referring to the 'land of Islam' and a range of named countries and regions.

Although, as mentioned above, in-group members are more likely than out-group members to be represented as individuals, in some contexts the in-group are nevertheless represented collectively (Baker et al., 2021; Booth, 2023). This has a dual function. It is used to emphasise and personalise the perceived threat to the in-group and their way of life from the out-group, representing each individual in-group member as being under threat. In some cases, it can also imply that the text producer is speaking for all members of the in-group, legitimising their own view in the process. Relatedly, some monologic texts collectivise the in-group by presupposing that the in-group audience share the same qualities and views (Baker et al., 2021; Bostdorff, 2004; Duffy, 2003; Lorenzo-Dus & Nouri, 2021; Prentice et al., 2011). This serves as a persuasive device by legitimising the views expressed by text producers, so may have a homologising effect on in-group readers and their group identity. A similar effect is achieved by the use of a shared specialist lexicon (Holt et al., 2020), especially where that lexicon is derived from sacred texts and therefore conveys a prestigious shared identity (Brookes & McEnery, 2020). As Darics and Gatti (2019, p. 246) highlight in a digital business context, this "shared linguistic repertoire" helps to place members of groups and teams 'on the same page'.

2.5.3 Divisions within the in-group identity

Although the most fundamental idea in the literature on extremist identity formation is the distinction between the in- and out-group, a number of studies show that this simple boundary does not tell the whole story of in-group identity. To capture this complexity, Berger (2018) proposes a three-tiered model of the in-group: 1) the *extremist in-group* of "active supporters" (p. 62) of an extremist movement; 2) the *eligible in-group* representing the "broad identity collective" (p. 62) on which the extremist group bases their identity and from which they recruit (white people, in the case of white nationalism); and 3) the *ineligible*

in-group, who are eligible on the basis of their social identity categories but actively “reject the extremist movement” (p. 63). Within the extremist in-group, attitudes towards the other tiers of in-group may vary. Berger (2018, p. 63) describes how the ineligible in-group may be treated as the extremist group’s “functional out-group” and draw more negative attention than ‘genuine’ out-groups due to their perceived betrayal of their own people. Likewise, for the most extreme groups even the eligible in-group may be treated with disdain, as extremist in-group members begin to “treat eligibility as obligation” (Berger, 2018, p. 63) to join the group. Elsewhere (Booth, 2023), I have empirically demonstrated some of Berger’s claims in a white nationalist forum, finding that the ineligible in-group of ideological opponents are represented as negatively as, if not more negatively than, racialised out-groups. I also show that forum members express dismay towards or criticism of eligible white people for their perceived betrayal of their race and other alleged wrongdoings. This serves to legitimise extremist in-group identities and position them as superior to other in-group members, as well as implicitly provide a model for how extremist in-group members should behave in order not to be ‘cast out’ with other, more peripheral in-groups. Baker et al. (2021) similarly show that some violent jihadist texts distinguish between ‘good’ and ‘bad’ Muslims, and lament that more Muslims have not ‘woken up’ to a violent jihadist ideology. Heritage and Koller (2020) find that extreme misogynist ‘incels’ (‘involuntarily celibate’ men) express resentment towards sexually desirable ‘Chad’ men, who represent the top level of a hierarchy of types of men. This may not always be the case, however; Lorenzo-Dus and Nouri (2021, p. 421) note that “[n]o negative in-group representations are provided” in their alt-right blog corpus.

Although Berger’s (2018) framework is valuable, it stops short of representing the whole range of variability within in-group identity. While he maintains that the extremist in-group has “strict boundaries” (p. 62), implying a high degree of homogeneity among its members, several studies have shown differences in identity expression within the extremist group. This takes the form of different claims about what the extremist in-group should be, do, and believe, hence drawing the boundaries of in-group membership in different ways. Anahita’s (2006) study of a racist skinhead forum finds explicit discussion of membership boundaries, describing ‘rules’ for skinhead identity (e.g. ‘only males are skinheads’; ‘we can’t have tattoos’). Brindle’s (2016) account of a white supremacist forum shows that members outline appropriate behaviour for white heterosexual men, including details of suitable clothing and film preferences. Discussion of the beliefs of ‘true’ white nationalists is also common (Booth, 2023; Brindle, 2016), which Brindle (2016) describes as a statement of “prerequisite[s] to being an in-group member” (p. 108). Extremist in-group members may also directly criticise other members of that group (Josey, 2010; Krendel et al., 2022).

Chapter 3 The career

3.1 Introduction

In this chapter, I introduce the sociological idea of the *career*. After outlining how the concept has been understood by other scholars, I link the career to the discussion of language and identity in Chapter 2, and discuss how these ideas have been applied to social movement contexts, particularly online.

The career refers to an individual's progression through a series of roles or stages within an institution or other, less formalised, organisation or group. The sociological meaning of the term is in many ways analogous to the layperson's understanding of the occupational or professional career. In fact, the sociological concern with careers began with Hughes' (1937) discussion of such traditional careers. As a result of this lay understanding, the application of the career concept is perhaps most intuitively understood in contexts like the occupational "where there is a formal hierarchy of rank, status and prestige" (Scott & Hardie-Bick, 2022, pp. 74-75). However, the term has broadened its meaning, and can refer to "any social strand of any person's course through life" (Goffman, 1961, p. 127). The concept has been applied to, among others: the process of patients' admittance to, and life in, psychiatric hospitals (Goffman, 1961); the process by which members of 'deviant' (socially non-normative) communities move further away from general social norms, and closer to local norms (Becker, 1973); participation in sports and leisure activities (Scott & Hardie-Bick, 2022); and, as we will see later, activism (Fillieule, 2001; 2010; see §3.5).

3.2 Career contingencies

The way in which the individual moves through their career is determined by *career contingencies*. These are "factors on which mobility from one position to another depends" (Becker, 1973, p. 24); thus, the continued career is *contingent on* these factors. The aim of studies of careers is therefore typically "to uncover the recurrent or typical contingencies" (Scott, 2014, p. 59) of careers in a given context. There are both *objective* and *subjective* career contingencies (e.g. Becker, 1973; Scott & Hardie-Bick, 2022; Strauss, 1959).

Objective contingencies are structural and clearly defined facts which relate to the individual's social position. This can include their current and previous roles, which may lead to subsequent roles, especially in an occupational context (i.e. needing to occupy a junior position before progressing to a senior position). Other such contingencies may include qualifications, skills, socio-economic status, and the reputation of both the individual and any other actors or associations they are involved in. In other words, objective contingencies are

any facts of the social structure and the individual's position within it which open the door for that individual to progress through their career. Subjective contingencies, on the other hand, are much more abstract and difficult to observe directly, being internal to the individual's own mind. These contingencies concern the individual's own perception of themselves and how they relate to others, both within and beyond the pertinent community; their changing knowledge of and capacity for community-specific behaviour; and their changing motivations which may inspire them to take certain career steps. Subjective contingencies, then, are a question of how the individual's experience of themselves – that is, their identity – changes over time (§3.2.1).

The development of these subjective career contingencies is described by Becker (1973) as a learning process; the individual learns to see the world from new points of view, allowing them to progress to new states of mind. Various but interrelated kinds of knowledge and inclination must be learned in order for the individual to progress through their career. Becker (1973) himself describes careers in deviant communities, whose practices are looked down upon by society at large. The average person is assumed to be motivated to avoid these deviant practices and communities, so as to avoid the reputational damage that results from being associated with them. Therefore, individuals who become members of deviant communities must have learned new motivations and desires which are sufficiently strong to overcome the socially normative ones they have presumably been brought up with and socialised into. Initial forays into the deviant community and its practices may simply be the result of "casual experimentation [... or] a random impulse" (Becker, 1973, p. 30). However, sustained participation in the community – in other words, a career – requires the 're-programming' of normative motivations, in order for the individual to perceive the benefits of community membership as more important than the reputational costs. This involves learning the meanings of the community's practices, and coming to understand, appreciate, and feel the value of these for the community. Of course, this is linked to the learning of new motivations; a person is unlikely to be motivated to continue their career without gaining this understanding. Particularly in the case of deviant practices, however, individuals may be able to appreciate or 'get something out of' these practices while still being affected by the pressures of social normativity, so may not progress in their career.

3.2.1 The career and identity

As noted above (§3.2), the career can also be seen as the individual's process of learning to perceive their own identity in new ways. Goffman (1961) describes the careers of patients who willingly admit themselves to psychiatric hospitals, thus moving from what he calls the *prepatient* to the *inpatient* phase of their career. Their decision to do so, he says, must be

the result of a re-evaluation of their understanding of themselves, deriving “[p]resumably [...] hav[ing] found themselves acting in a way which is evidence to them that they are losing their minds or losing control of themselves” (p. 131). In other contexts, individuals begin to experience feelings of “belonging and solidarity” (Scott & Hardie-Bick, 2022, p. 74) within the community, and, thus, to consider themselves part of some shared or collective identity (a concept I return to in §3.6). The learning of new motivations or understandings of the symbolic meaning of community practices feed into this, allowing the individual to further assimilate to this collective self (Collins, 2004).

This learning is achieved interactionally, either through direct teaching from other members of the community, or through observation and imitation of others (Becker, 1973). Such interaction is fundamental: an individual cannot learn how to be a member of the community, and so to progress in their career, without participating in that community. Notably for my purposes here, Becker (1973) highlights that “[t]he vocabularies in which deviant motivations are phrased reveal that their users acquire them in interaction with other deviants” (pp. 30-31). This indicates not only that individuals develop their new, community-oriented identities through interaction with other members, but also that these new identities are performed linguistically, in their newly acquired group sociolect (§3.4). Individuals are, therefore, able to acquire new identity resources (Grant & MacLeod, 2020; §2.4) as their career progresses.

As in society at large, these community-specific performances are policed by a system of mutual monitoring (Goffman, 1964; see §2.2.1). In a given social context, norms are norms precisely because members of that community implicitly agree on the ‘proper’ ways to act or think in that context. This is true even of deviant communities, which have their own norms that are very distinct from those of wider society. When members interact with the community, they “take into account the way their fellows will evaluate what they do, and how that evaluation will affect their prestige and rank” (Becker, 1973, p. 183). Therefore, the rest of the community, especially its more established members, act as gatekeepers for the individual’s career, facilitating or blocking access to the community (Lindesmith et al., 1999; Scott & Hardie-Bick, 2022). Only if the individual behaves in a way that is sanctioned by the community will they be able to progress to later stages, and themselves become an established member of the community.

3.3 Describing the career

3.3.1 The idea of the typical career

The foundational literature on careers is divided on the question of whether it is possible or desirable to identify a typical career in a given context. On the one hand, Becker’s method of analytic induction in the context of deviant careers sought to identify a “general statement of

the sequence of changes [...] which always occurred" (Becker, 1973, p. 46) among each of his 50 informants. Similarly, Goffman's approach is to discount "unique outcomes", focusing instead on processes of change which are "basic and common" (Goffman, 1961, p. 125) to members of the relevant group. Although concerned with these abstract *general* or *common* career paths, both researchers accept that deviations from these normative paths occur. For example, Becker (1973) details how individuals may step away from participation in the community, creating a 'gap' in their career before their eventual return. The psychiatric patients Goffman (1961) observed are described as experiencing both progress and setbacks in their recovery careers, often "in different directions and at different rates" (p. 163) from one another, while still remaining in the *inpatient* career stage. Both also mention the possibility of individuals halting their career progression at a particular stage, remaining part of the community without moving on to subsequent stages. However, these deviations are not considered to represent alternative career paths. The career stages described by Goffman and Becker are therefore necessarily more abstract, capturing "importantly similar" (Goffman, 1961, p. 129) structural properties of individuals' careers despite lower-level variation.

This contrasts sharply with the position taken by Fillieule (2001, 2010), who argues that in the study of activist careers "l'unité pertinente est l'individu" [the proper unit of analysis is the individual] (Fillieule, 2001, p. 203). While recognising that careers will logically follow a very broad pattern of initial engagement, followed by a period of more or less active participation, and finally disengagement, Fillieule emphasises the "variations d'intensité" [variations in intensity] (Fillieule, 2001, p. 206) and duration between individuals' patterns of engagement. Fillieule argues that what may nominally be considered as the same career stage varies, in reality, in many significant ways, depending on any number of career contingencies (§3.2) which may differentially impact individuals. He highlights this with the example of the *disengagement* stage (Corrigall-Brown, 2012; Klandermans, 1997), which may "vary as a function of what provokes it, the cost of defection, the manner in which it takes place, and therefore what becomes of those who leave" (Fillieule, 2010, p. 2). This position can be extrapolated to an assumption that there are as many career paths as there are individuals in a group.

This debate remains unresolved, in part because most of the researchers who have explicitly discussed the career concept have done so in a largely abstract and theoretical way. Each of Becker, Goffman, and Fillieule describe the career in a single context, offering no suggestion as to how the concept may be applied to new contexts. This dearth of evidence makes it unclear how careers in different contexts might actually unfold over time. However, as I show in Chapter 5, a number of papers exist which do not explicitly work within the

career framework, but may nonetheless be understood to describe careers. These papers introduce typologies of roles or stages which individuals might pass through across a range of different social contexts. In the next section, I explore what both the foundational careers literature, and this new set of papers, can tell us about how careers develop over time.

3.3.2 The career over time

Starting with the foundational literature on careers, the typical path with which Becker (1973) concerns himself is characterised by increasing engagement over time before a plateau of “occasional” (p. 44) or casual engagement. He recognises that some individuals may deviate from this path, displaying career gaps and periods of decreasing engagement. Goffman echoes this, recognising that within the framework of the typical career he describes, individuals may progress in “different directions and [...] at] different rates” (Goffman, 1961, p. 163). Both Becker and Goffman lead us to believe that the typical career may be relatively long. In deviant communities, including Becker’s community of marijuana smokers and the online white nationalist forum I investigate in this thesis, Becker argues that members develop a shared non-normative view of the world, viewing themselves as “in the same boat” and in turn “solidif[ying] a deviant identity” (Becker, 1973, p. 39), making members less likely to leave the group. The career Goffman describes takes place within a psychiatric hospital, where individuals cannot leave easily or at will. In this context, the career may last for the rest of individuals’ lives. This career is, of course, unusually constrained, and we may expect careers elsewhere to be shorter. This may particularly be the case in online contexts, such as the forum; literature concerning a range of online environments shows that careers are often vanishingly short (e.g. Arguello et al., 2006; Halfaker et al., 2012; Jensen et al., 2011; Lampe & Johnston, 2005; Preece & Shneiderman, 2009; Resnick et al., 2010; Wang et al., 2012).

Beyond the literature which works explicitly within the career framework, we find two main approaches to the question of how the career unfolds over time. One approach does not take a diachronic perspective at all, instead describing a set of distinct roles without considering how members progress towards those roles after joining the community (e.g. Chiang et al., 2021; Gill & Horgan, 2013; Marcoccia, 2004; Waters & Gasson, 2005). The existence of different roles naturally implies that members follow different career paths as they develop their ability to fulfil these distinct roles. This is implicitly recognised by Altier et al. (2022), who note the existence of “pathways” (p. 753) to reach these distinct roles; however, these pathways are not explored further. While not offering any information on career development over time, this contributes to the debate on the idea of the typical

career. It suggests that there may not be one typical career path, but rather a set of typical paths leading to different end goals – that is, to different roles within the community.

The second approach focuses on career progression through time. These papers implicitly advocate for the idea of a single typical career path which all members are implied to follow. In many cases this is limited to a simple set of labels amounting to *beginning*, *middle* and *end* of career, such as Brush et al.'s (2005) *new*, *active* and *past* users; Klandermans' (1997) *initial engagement*, *sustained participation* and *disengagement*; and McAuley and Leskovec's (2013) *novices*, *experts* and *leaders*. These imply a simple linear career path shared by all members, likely comprising increasing engagement, reaching a peak, then dwindling to nothing. Fillieule's (2001, 2010) concerns can be echoed here, as a single typical career may only be identifiable at an extremely high level of abstraction.

There are also a small number of papers which take into account the possibility of both multiple career paths, and of progression over time. However, this is usually only in a limited capacity. Smith Risser and Bottoms (2014), Wenger (1998), and Dennen (2008) focus primarily on one central linear career path, but recognise a secondary, 'peripheral' path. The peripheral path is used to describe members, sometimes long-term members, who remain on the fringes of and do not participate fully in the community. Others (Golder & Donath, 2004; Viégas & Smith, 2004; Welser et al., 2011) take a similar approach to the first strand. These papers identify a set of roles, including new users as a distinct type of member. However, no account is given of the paths newcomers take towards fulfilling other roles. Authors who more fully explore both approaches include Corrigan-Brown (2012), who mirrors Fillieule (2001, 2010) in the assumption that each individual will follow a unique career path based on biographical and ideological factors. Corrigan-Brown describes activist careers which include periods of continued engagement (*persistence*), career gaps (*individual abeyance*), and permanent *disengagement*, all taking place on varying timescales. Preece and Shneiderman's (2009) review provides a flexible view of the online career, proposing a simple linear career path alongside an extensive account of ways in which users deviate from this path. This includes: members remaining static at a particular career stage, either continuing at their preferred level of engagement indefinitely or leaving without progressing to a subsequent stage; 'skipping' stages entirely and proceeding to a 'later' stage; and returning to 'earlier' stages. This is echoed by Scott and Hardie-Bick (2022), who, working explicitly within the career framework, suggest that while some individuals follow "a simple, linear sequence of progressive steps towards the earning of new status," others may "move erratically" between stages (p. 75).

3.3.2.1 *In white nationalist fora*

Two papers have investigated the development of the career over time in the context of white nationalist fora. Kleinberg et al. (2021) identify two main types of user: typical or 'normal users', and 'super users'. They found that the majority of posts to the site were made by a small proportion of users, and defined 'super users' as those in the 99th percentile of posting frequency. These users had an average career length of just under six years. The label 'normal users', who had an average career duration of just nine months, is therefore applied to a very wide range of careers. This again suggests that a typical career may only be found at a high level of abstraction.

Similarly, Scrivens et al. (2022) use group-based trajectory modelling to "identify latent groups of cases that share similar patterns of behavior over time" (p. 1723) – in other words, different career paths – in two right-wing extremist fora. They distinguish careers based on initial posting rate – *low*, *moderate*, *high*, and *explosive* (that is, starting high and quickly rising) – and level of continued participation – *early* and *moderate* desisters, whose engagement decreased over time, and *chronics*, whose engagement continued at a similar level throughout the career. These careers were short, continuing for up to nine months, and career gaps were common. In both fora, by far the most common career is the *low-onset early desister*. Across the datasets, most users are desisters; that is, their posting intensity decreases over time, rapidly or otherwise.

3.4 Language and the career

As far as I am aware, very little literature explicitly links the career with changing linguistic expression. The exception is Becker (1973), who, as I discuss in §3.2.1, briefly notes that individuals acquire new "vocabularies" (p. 30) through their career. However, there is a significant literature which takes a longitudinal approach to studying language use in various contexts, examining how members of communities change their linguistic usage across their membership. In this section, I outline some of this literature. Due to the subject of this thesis, this review focuses on how these changes have been described in online contexts.

This literature typically considers only two distinct stages of the career. Newcomers, or *newbies*, receive significant attention in the literature, with 'post-newcomer' language simply contrasted with that of newcomers. As this typically highlights the relative expertise of the 'post-newcomer' member, the latter are often referred to as experts or *veterans* (Newsome & Grant, in press), terms I will use here to refer to this second career stage.

3.4.1 Newcomers

The interest in newcomers may stem from the marked nature of newcomers' behaviour relative to the rest of the community. Newcomer identities may be stigmatised due to their lack of knowledge of community norms (Boostrom, 2008; Golder & Donath, 2004; Graham, 2016), which may cause annoyance or embarrassment to more established members. In this way, newcomers threaten the *negative face needs* (Brown & Levinson, 2006) of both themselves and others. Negative face needs are the needs of individuals not to be imposed upon or made uncomfortable by others. Newcomers therefore employ negative politeness strategies (Brown & Levinson, 2006), such as indirectness, to avoid imposing on the community. Less frequently, they may also employ positive politeness strategies, such as praise and flattery, to appeal to interlocutors' positive face needs, or "desir[e] that [their] self-image be appreciated" (Brown & Levinson, 2006, p. 311).

Newcomers' language is frequently characterised by explicit and implicit references to their newcomer identity, protecting existing members' negative face needs by showing that they are aware of their own lack of knowledge and stigmatised status (Chiang, 2018).

Newcomers' first act is often to produce a dedicated introductory post, announcing their arrival to the community (Chiang, 2018; Graham, 2016; Nguyen & Rosé, 2011). These posts are often of a personal nature, exhibiting a higher rate of usage of first person singular pronouns than more experienced community members (Danescu-Niculescu-Mizil et al., 2013; Nguyen & Rosé, 2011). The posts often include a significant narrative element. This may include biographical information about themselves (Graham, 2016), depending on the nature of the community and the level of anonymity expected. Most often, these narratives focus on the connection that the newcomer shares with the community. The newcomer's motives for joining the community are often recounted – these may be a significant life event such as a medical diagnosis (Angouri & Sanderson, 2016; Arguello et al., 2006; Galegher et al., 1998), or simply the development of a shared interest, whether socially acceptable (Graham, 2016) or deviant (Chiang, 2018).

Newcomers also orient to their newness in more indirect ways. As Benwell and Stokoe (2006) summarise, newcomers may "do' being cautious" (p. 273) or even frightened (Honeycutt, 2005). This is a negative politeness strategy, emphasising their low status and lack of authority within the community. In this way, newcomers perform an unassertive role, showing their respect for the status quo and the community. Other negative politeness strategies that may be employed by newcomers include the use of modal verbs and other hedging devices to show low epistemic commitment to their propositions (Benwell & Stokoe, 2006; Golder & Donath, 2004), avoiding claims of authority through overly assertive claims.

Chiang (2018) also finds that requests to join a community may be made indirectly, making reference to “looking forward” (p. 193) to joining. She notes, however, that a smaller number of newcomers simply make direct requests.

Newcomers may also employ the negative politeness strategy of self-deprecation and criticism of their own skill levels in the activity central to the community (Benwell & Stokoe, 2006). Chiang (2018) reports that newcomers highlight their own lack of experience, and seek understanding and tolerance for this from other members. Likewise, Giles (2006) finds that newcomers often give a “disclaimer” (p. 467) of their status when offering advice or opinions to others, signalling that their newness may be seen to affect the quality of their contribution. Relatedly, they may offer “unnecessary apolog[ies]” (Golder & Donath, 2004, p. 12), not for the content of their posts but for the very fact of their posting. Newcomers also employ positive politeness strategies (Brown & Levinson, 2006), which protect other users’ face through praise and flattery. For example, newcomers commonly express thanks for other members’ contributions and responses (Chiang, 2018; Nguyen & Rosé, 2011), positioning them as experts who offer useful advice (Rudolf von Rohr et al., 2019). This behaviour also positions the newcomer as “an authentic advice seeker” (Rudolf von Rohr et al., 2019, p. 243) who respects the community and does not wish to waste its time.

Uncertainty and lack of authority is also indexed by newcomers through the frequent asking of questions (Chiang, 2018; Galegher et al., 1998; Newon, 2016; Nguyen & Rosé, 2011). Golder and Donath (2004, p. 13) describe this as the “single most prevalent behaviour” exhibited by newcomers. Ang and Zaphiris (2010) provide quantitative data showing that question-asking is not only a common interaction type among new users, but also that rates “dro[p] dramatically” (p. 610) as individuals gain experience over time. When asking questions, polite formulae are used (Golder & Donath, 2004), again to mitigate the imposition placed on other members. For instance, members may ask for any help or *any information* (Golder & Donath, 2004, p. 14), lowering the force of their demand and the sense of obligation for readers to respond.

More rarely, newcomers perform other kinds of identity. Some newcomers attempt to emphasise their legitimacy within the community right from their first post, performing an expert persona that Chiang (2018, p. 203) refers to as a “competent newbie” identity. In particular, Chiang draws attention to the role of narratives in newcomers’ introductory posts. These newcomers describe their interests and history in the relevant (criminal) context, highlighting the expertise they already have beyond the forum. As Galegher et al. (1998) argue, these narratives are a means for the newcomer to establish the legitimacy of their claim to membership of the group, allowing the newcomer to emphasise common ground

between themselves and the community. Newcomers may also refer to a period of *lurking*, silent observation of an online community without contributing posts. Lurking equips newcomers with community-specific knowledge which they can draw on when making their first posts in order to appear more competent (Golder & Donath, 2004; Preece et al., 2004; Radin, 2006). By referring to their lurking behaviour, newcomers can emphasise their competence, making it clear they have 'done their homework' on the community and its practices (Chiang, 2018). As Galegher et al. (1998) highlight, it also allows newcomers to position themselves as already being a (silent) community member rather than a complete outsider. Chiang (2018) finds that 'competent newbies' may go further than simply demonstrating their alignment with the community, by also outlining the value they can offer to the community in terms of content, support, and skills. Besides introductory posts, newcomers may assert themselves interactionally, contrary to typical patterns. Benwell and Stokoe's (2006) case study shows that newcomers may display their own expertise, expressing their own opinions and even openly disagreeing with other members. They may also challenge existing members' authority "by invoking [their] own" (Benwell & Stokoe, 2006, p. 276), responding in kind to hostility from other members and positioning themselves as having the authority to threaten.

3.4.2 Experts

Where newcomer language is primarily characterised by question-asking, veterans frequently prioritise question-answering, advice-giving, and other information provision. Naturally, the exact nature of this is community specific. Examples from the literature include: citing personal experience to offer advice on a medical condition (Angouri & Sanderson, 2016; Rudolf von Rohr et al., 2019); giving technical support on software (Singh, 2012); giving instructions in online games (Ang & Zaphiris, 2010; Newon, 2011); or explaining topic-specific vocabulary (Peuronen, 2011). Unusually in this literature, Singh (2012) offers a longitudinal view of how this behaviour changes as newcomers advance through their career to become more expert users of a site. Newcomers in the very beginning of their career often ask questions in an indirect way, opening with non-explanatory statements like *I'm stuck*. Upon receiving advice, they ask many clarification questions. A little later on, these users begin to ask "direct, to the point question[s...] giv[ing] suitable details of what they have tried before asking the question" (Singh, 2012, p. 71), showcasing greater knowledge of the topic and the community's norms. They later transition into giving advice.

Experts also index their knowledge in other ways. In particular, many researchers cite the use of specialist and topic-specific vocabulary (Danescu-Niculescu-Mizil et al., 2013; Newon,

2011; Nguyen & Rosé, 2011; Peuronen, 2011). Users may position themselves as experts in order to offer support and reassurance to newer members. This is particularly noted in the context of health or illness fora, where members share their experiences of diagnoses and treatments (Angouri & Sanderson, 2016; Nguyen & Rosé, 2011; Rudolf von Rohr et al., 2019). Here, advice- and information-giving are combined with expressions of empathy and shared experience. Members often reinforce the knowledge shared by others, replying with explicit agreement to others' posts. This not only serves as further reassurance for the advice-seeker, but also creates a performance of group expertise (Rudolf von Rohr et al., 2019). Relatedly, members may also increase confidence in the forum and community by explicitly praising and advocating for it, encouraging newcomers to return and continue seeking advice there (Angouri & Sanderson, 2016).

Veteran members display authority within the group by managing the community in various ways. This includes guiding newcomers and ensuring that they behave in appropriate ways for the community. While some members respond to newcomers' introductory posts with a simple welcome, others provide information and advice which serve to 'induct' these newcomers into the community (Angouri & Sanderson, 2016). Some take this even further by providing ongoing mentorship to newcomers (Lampe & Johnston, 2005). This involves explicitly assessing others' performance on the site (Newon, 2011) and making any consequences for ignoring instructions clear (Lampe & Johnston, 2005). In communities which focus on providing support and advice such as health support fora, more established members may praise others' efforts, positively reinforcing particular behaviours (Rudolf von Rohr et al., 2019). Elsewhere, members take a more explicitly managerial role; Ang and Zaphiris (2010) describe how experienced players of an online game restructure their group by assigning roles, as well as organising members by deciding in-game meeting points. Group management may also involve the maintenance of group boundaries (Golder & Donath, 2004; Honeycutt, 2005). I have already outlined in §2.5 how boundaries can be (re)negotiated and reinforced in extremist online environments. Honeycutt (2005) emphasises that this may involve overt hostility and threats until newcomers are deemed to have earned their place in the community.

Researchers commonly report the use of inclusive first person plural pronouns among veteran users (Angouri & Sanderson, 2016; Danescu-Niculescu-Mizil et al., 2013; Nguyen & Rosé, 2011; Pluempavarn et al., 2011; Rudolf von Rohr et al., 2019). These pronouns are widely used to position the writer as a representative of the community's voice, emphasising the user's authority (Rudolf von Rohr et al., 2019). Complementing this, Danescu-Niculescu-Mizil et al.'s (2013) longitudinal approach finds that while use of first person *plural* pronouns quantitatively increases over time, first person *singular* pronouns decrease. They suggest

that this finding “might be attributable to a user’s increasing identification with the community” (Danescu-Niculescu-Mizil et al., 2013, p. 310). Nguyen and Rosé (2011) link the use of inclusive pronouns to other community-specific and familiar terms of address used by veteran users to address their fellow members. Indeed, these users show a greater preoccupation with the development of rapport and sense of community than newcomers do. Nguyen and Rosé’s (2011) mention of the strongly personal nature of interaction between users is one example of this, signifying “emotional involvement with other forum members” (p. 83). Veteran members can increase solidarity with one another, and particularly with new users, through the use of hedging and low epistemic commitment. As Angouri and Sanderson (2016) remark in the case of a health support forum, this positions the writer as a comforting presence, reassuring fellow users by downstepping potentially worrying statements or advice. Newon (2011, 2016) notes the solidarity-building function of hedges, which “mitigat[e] social difference” (Newon, 2016, p. 295) between expert members giving instructions and information on the one hand, and their less experienced addressees on the other. In this way, veteran members may, like newcomers, attend to the negative face needs of other members.

Relatedly, users may engage in off-topic talk, showing a broader connection with the community than simply the central topic or issue which brought them to the site (Ang & Zaphiris, 2010; Angouri & Sanderson, 2016). Ang and Zaphiris (2010) find that this talk, which often takes a light and joking tone, is exclusive to veteran users, with newcomers beginning to take part in it as their career progresses. For Angouri and Sanderson (2016), off-topic discussion by veterans, which they describe as *rapport-oriented*, serves as “modelling of appropriate topics” (n. p.) for other users. This aligns with the claims of Graham (2016) and Golder and Donath (2004), who note that veterans have the “social authority” (Graham, 2016, p. 313) to determine community norms. This includes the power to flout these norms without facing the repercussions that newcomers would face (Graham, 2016). In this way, veterans users have the ability “to push and extend boundaries of acceptable conversation” (Golder & Donath, 2004, p. 22). This is mirrored in the discussion of collective identity below (§3.6). Once individuals have assimilated to a collective or group identity and the community norms associated with that identity, they can begin to contribute to the process of interactional renegotiation of the collective identity.

3.5 The career in the context of social movements

The notion of the career was introduced to social movement studies by Fillieule (2001, 2010, 2020). Through this lens, activism is recognised as a “long-lasting social activity articulated by phases of joining, commitment, and defection” (Fillieule, 2010, p. 4). Every stage of the

career is understood in relation to both the individual's personal history and the context of the particular movement or organisation (Fillieule, 2010). Past experiences inform individuals' present attitudes, perceptions, and behaviours; in turn, both past and present experiences will go on to influence the activist's future career path. This model allows the researcher to understand decision points in the activist's career not as isolated 'moments', but as part of a continuous process. This context makes the individual's decisions more readily explicable.

While particular movements and organisations provide 'push' and 'pull' factors that influence all their members' careers (e.g. Bunnage, 2014), Fillieule strongly emphasises the importance of the individual's life history. For Fillieule, the activist career is to be seen as one facet of the individual's life as a whole; it necessarily interacts with other elements of the individual's life, meaning that no two individuals' careers are exactly the same (Fillieule & Neveu, 2019; see also Horgan, 2009; Pilkington, 2023b). As a result, he recommends the individual, rather than the movement or organisation, as the proper unit of analysis for the study of careers (Fillieule & Neveu, 2019).

The notion of activism as a continuous career allows for a range of issues addressed by social movement scholars to be brought together as a single focus of study. These include the exploration of factors which give individuals a "predisposition to [...] activism" (Fillieule, 2010, p. 11), and encourage them to get involved with activism; the motivations to continue engaging in activism, and to move to different kinds of activism; and eventually, the causes of disengagement. By grouping these elements as stages of a single career, we can explain the similarities found between the motivations for these different phenomena. Rather than the same explanations being offered for multiple separate events, such factors can be understood as career contingencies, which affect the same career at different points.

By far the most frequently mentioned activist career contingency in the social movements literature is the social ties individuals have to other activists (Van Dyke & Dixon, 2013). The fact of being involved with existing networks of activists, even very informal ones, is very often a key factor in the individual's first forays into activism (Blee, 2002; McAdam, 1986; Somma, 2009). Snow et al. (1980) in particular emphasise the importance of social networks as a catalyst to engagement, stating that "[h]owever reasonable the underlying assumption that some people are more (psychologically) susceptible than others to movement participation, that view deflects attention from the fact that recruitment cannot occur without prior contact with a recruitment agent" (p. 789). Of course, this implies that other factors contribute to the individual accepting that invitation to engage; nonetheless, this engagement relies fundamentally on social ties. Social networks are also instrumental in the choice to

continue the career, as the stronger social ties developed through participation help activists to feel appreciated and more likely to remain (Klandermans, 1997), and facilitate mutual reinforcement of beliefs (Nepstad, 2004). On the other hand, social ties can also contribute to disengagement from an activist career, either when an individual's relationship with other activists breaks down (Horgan, 2009) or when an individual follows their close contacts in disengaging from a group, creating a 'negative bandwagon effect' (Sandell, 1999).

Other commonly cited factors influencing activists' career paths include *biographical availability*, defined as "the absence of personal constraints that may increase the costs and risks of movement participation, such as full-time employment, marriage, and family responsibilities" (McAdam, 1986, p. 70). Greater biographical availability is associated with greater rates of activist recruitment and retention – that is, with individuals deciding to embark on or continue an activist career. Meanwhile, certain factors, such as finding full-time employment or having children, make the individual less likely to commit time to a social movement, and therefore to disengage or simply never join (Corrigan-Brown, 2012; McCarthy & Zald, 1973; Oliver, 1984).

Perhaps contradictorily, the possession of material resources (such as money) and symbolic resources (such as high levels of education) is also linked by some researchers to engagement and sustained participation in activist communities (Schussman & Soule, 2005; Verba et al., 1995). While potentially limiting participation, full-time employment can also encourage engagement as it is associated with greater resources (McAdam, 1986). This combination of low biographical availability with high possession of resources can result in activists continuing to engage infrequently, rather than entirely withdrawing from their activist career (Corrigan-Brown, 2012). Just as Snow et al. (1980) say regarding social ties, McAdam (1986, p. 87) concludes that "regardless of their level of ideological commitment to the project, it is the extent and nature of the applicants' structural locations vis-à-vis the project that best accounts for" engagement in a social movement.

The particular movement or organisational context can also impact activists' careers and decisions to (dis)continue engagement. Organisations which provide their members with opportunities to develop new resources in the form of activism-related skills are likely to foster continued participation, as individuals feel empowered by these skills (Van Dyke & Dixon, 2013). When the structure of an organisation is such that activists are regularly brought together, social ties between activists are strengthened, keeping them engaged (Nepstad, 2004). Feelings of connection between the individual and the movement are also strengthened (Corrigan-Brown, 2012). If an organisation changes its structure, activists who preferred the previous structure may decide to disengage (Klandermans, 1994).

These contingencies often mentioned in relation to activist career trajectories are largely *objective* in nature, referring to the social structures surrounding the individual which influence their participation at different times. Fillieule and Neveu (2019) favour this highly biographical approach, emphasising the importance of external factors and how an individual's activist career relates to the rest of their life. While these factors show us how activism sits within the wider context of the individual's life course, less is known about the internal or subjective elements of activism, and the ways the individual's perceptions change over time through their activism – in other words, about the life course of activism itself.

A small number of researchers make reference to such subjective contingencies. These contingencies can be understood both as motivating factors in the development of the individual's career, but also as evidence for the transformation of identity implicated by the progression through a career. Bunnage (2014) summarises the link between subjective factors and the activist career, noting that “[c]ommitment is about what individuals believe, how they perceive themselves, and their [feelings of] responsibility and ability to engage in change making [...] the extent to which one feels ready to engage in long-term activism” (p. 437). Pre-existing political and ideological commitments can be a factor in initial engagement (Schussman & Soule, 2005), but some activists only develop relevant ideological commitments after joining a movement or group, becoming more sympathetic to the movement's ideology through interacting with other activists who believe in it (Ahmad, 2016; Blee, 2002; McAdam, 1986). Individuals may also be motivated by a desire to create change (Klandermans, 2004), and a belief that they can use their (perhaps newly developed) skills to do so (Van Dyke & Dixon, 2013). If the individual finds that they are unable to create change and therefore fulfil their expectations of activism, and that the costs of activism therefore outweigh the benefits, they may become disillusioned and disengage (Klandermans, 1997). Benefits of activism can, therefore, also be understood as subjective career contingencies, as their presence is a factor in the individual's progress through their career. The most significant subjective contingency or benefit mentioned in the literature on social movements is *collective identity*.

3.6 Collective identity

Collective identity is a “notoriously abstract concept” (Flesher Fominaya, 2010, p. 393), which has been defined in various ways by a number of social movement scholars. Put simply, collective identity can be understood as a sense of ‘we’-ness shared by members of a social movement organisation (Snow, 2001; Taylor & Whittier, 1992), which is derived from “real or imagined shared attributes and experiences” (Snow, 2001, p. 2213). Fundamentally, collective identity allows groups to recognise themselves, to define the boundaries and

criteria of group membership, and to understand the form and meaning of the action they undertake (Melucci, 1996). It defines and justifies, among other things, “[p]rocesses of mobilization, organizational forms, models of leadership, ideologies, and forms of communication [... as well as] relationships with the outside – with competitors, allies, and adversaries” (Melucci, 1996, p. 4).

Collective identity represents a social constructionist view of identity (§2.3) in social movement settings, with these identities understood to be negotiated through interactions both within and beyond the boundaries of the group (Melucci, 1996). Although studies of collective identity have typically been carried out in face-to-face contexts involving like-minded individuals in the same physical location, as well as ‘structural location’ with respect to society at large, these proximities alone are not sufficient to create a collective identity (Snow, 2001; Taylor & Whittier, 1992). Instead, such identities are “created in the course of social movement activity” (Taylor & Whittier, 1992, p. 110). As has been demonstrated for other kinds of identities (Busch, 2012; Grant & MacLeod, 2020), collective identity creates “a field of [both] opportunities and constraints” (Melucci, 1996, p. 4), which influence its members’ values and actions.

For Taylor and Whittier (1992), the interactional construction of collective identity has three key, intertwined strands: *boundaries*, *consciousness*, and *negotiation*.

The first of these, the drawing of *boundaries*, is already familiar from the discussion of online extremist identity creation (§2.5). It refers to the ways in which movement groups delineate their own identity from that of various out-groups, “promot[ing] a heightened awareness of a group’s commonalities” and “highlighting differences” between their own group and others (Taylor & Whittier, 1992, p. 111). Typically, social movement groups represent marginalised populations, whose boundaries are drawn for them by dominant groups with an interest in maintaining their own dominance. Collective identity therefore involves “a kind of reverse affirmation of the characteristics attributed to it by the larger society” (Taylor & Whittier, 1992, p. 111); in other words, the group’s reclamation of its own identity.

The second element of collective identity is the raising of *consciousness*. Consciousness is members’ understanding of how the group is socially positioned in relation to, and oppressed by, the dominant group, and how this oppression manifests in daily life. In other words, it is “a dominated group’s own explanation of its position” (Taylor & Whittier, 1992, p. 117). By creating this consciousness among its members, collective identity becomes a lens through which to see reality. Instead of attributing their position to personal failings, “collective actors must [learn to] attribute their discontent to structural, cultural, or systemic causes” (Taylor &

Whittier, 1992, p. 114), which creates both a positive image of the self and a negative image of the out-group, tying back into the idea of boundary creation.

The third and final strand Taylor and Whittier describe is the *negotiation* of the meaning of everyday items, symbols and practices, such as styles of dress, music, words and slogans, gestures and signs, relationships, and other lifestyle choices (see also Melucci, 1996; Saunders, 2008). Through repeated association with the movement and its ideologies (Collins, 2004), these practices are imbued with a political meaning, which allows members to redefine their everyday life as a form of resistance to their (perceived) oppression, and of commitment to the movement.

Collective identity is used to account for members' continued motivation to commit to a social movement group even when the group is not currently engaged in collective action, or when its actions are not achieving the desired results (Flesher Fominaya, 2010; Snow, 2001). As Melucci (1996, p. 63) explains, "the simple premise of common interest is not enough to explain mobilization without the introduction of a structure of incentives and individual advantages". This principle has long since been recognised: Olson (1965) describes the so-called 'free rider problem', which notes that individuals who stand to benefit from a movement's actions are nonetheless likely to 'sit back' and let the movement do the work on their behalf rather than get involved themselves. Collective identity offers a solution for this problem, as it provides its own rewards to those who choose to engage in action (Saunders, 2008). These rewards include a shared sense of meaning and belonging (Flesher Fominaya, 2010; Saunders, 2008) and positive affect towards other members and the group as a whole (Jasper, 1998). Because it increases motivation for individuals to continue their careers, collective identity is a subjective career contingency, and has been recognised as such in the literature (Bunnage, 2014; Corrigan-Brown, 2012; Klandermans & de Weerd, 2000; Taylor et al., 2009).

Collective identity has been variously understood as an ongoing *process*, and as a fixed *product*. Melucci (1996) is a strong proponent of the process view, arguing that collective identity is constantly "constructed and negotiated through the ongoing relationships linking individuals or groups" (p. 67). This is consistent with the social constructionist view already described (§2.3), as it rejects the idea that collective identity has a permanent 'essence' in favour of seeing it as created relationally through interaction. Melucci's formulation, while influential, has been criticised as too abstract, as the process is not empirically observable ("never a definite datum"; Melucci, 1996, p. 76) and, therefore, can be seen as lacking empirical evidence (see Flesher Fominaya, 2010). The product view, on the other hand, characterises collective identity as "something people [...] recognize and respond to"

(Flesher Fominaya, 2010, p. 397). Flesher Fominaya is here referring to movement outsiders, such as its opponents, allies, and potential recruits. However, the same principle can be applied to other actors such as the researcher, or even current members of the movement group, who, like all actors, intuitively try to understand their own identity in a clear, static way (Melucci, 1996). Under this definition, then, the collective identity is a 'thing' that can be directly 'pointed to', observed, and described (Saunders, 2008).

The value of the product view for the researcher is therefore clear. Even Melucci (1996) concedes that the collective identity process results in 'reifications', artefacts which reflect that identity. Among these, he lists "organizational structures, leadership patterns, [and] membership requisites" (Melucci, 1996, p. 75); we may also include the negotiated practices, symbols and items described above (Taylor & Whittier, 1992). Melucci recognises these as "empirical indicators of a possible collective identity" (Melucci, 1996, p. 75), but denies that these *are* the collective identity in themselves. Instead, he says, the collective identity "process itself [...] is always concealed behind these forms" (Melucci, 1996, p. 72). Collins (2004) also demonstrates how reifications – what he calls *rituals* – are part of a larger ongoing process of a collective identity that sustains itself and evolves over time. For Collins, the repetition of rituals across time is of fundamental importance, as each repetition 'charges' the reification with movement-related meaning, and "if the ritual is not carried out for a time, the sacredness [shared meaning of the ritual] fades away" (Collins, 2004, p. 17). As Wenger (1998) has described elsewhere, the repetition of a practice across different situations allows the meaning of that practice to shift over time, as the group "extend, redirect, dismiss, reinterpret, modify or confirm" (p. 52) previously-understood meanings. The collective identity is, therefore, never fixed, but is being constantly renegotiated through the very practices which reflect it in its present form. Reified collective identity 'products' can, therefore, only ever be seen as a single moment in time for the movement group. They are the product of that moment, but collective identity can never be understood as a *finished* product.

This has implications for our understanding of the career. The literature on collective identity is clear that "individual identities are brought to movement participation and changed in the process" (Johnston et al., 1994, p. 12) as individuals become increasingly bound to the group and its identity (Saunders, 2008). This identity transformation can be understood as the individual's acquisition of group-specific resources that allow them to perform the collective identity as part of their repertoire of resources (Grant & MacLeod, 2020). We may expect that this acquisition of resources will be reflected in the earlier stages of an individual's career within a group, with their linguistic behaviour shifting to become more similar to that of other group members. Implicit in the literature on collective identity,

however, is a view that assimilation to collective identity is a finite process, which is 'accomplished' or 'completed'. This occurs when, for example, the "set of individuals *become* a collective identity" (Flesher Fominaya, 2010, p. 393; emphasis my own); when the "networ[k] *transform[s]* its members into political actors" (Taylor & Whittier, 1992, p. 104; emphasis my own); or when the collective identity exercises its "*binding* effects on social movements" (Saunders, 2008, p. 228; emphasis my own). This view seems to predict that the individual's repertoire of linguistic behaviours will remain static once they reach this target.

However, if we understand collective identity as a process, this identity becomes a constantly moving target (Johnston et al., 1994), which continues to shift even after a given individual has become enmeshed in the group and its identity. Identity is (re)negotiated through interactions, and the meanings ascribed to different kinds of movement-related interaction are renegotiated with each interaction (Wenger, 1998). Once the individual is part of the group, they regularly take part in these interactions, and, in doing so, contribute to the renegotiation of the collective identity. The individual, therefore, not only continues to shift their own behaviour towards the moving target of collective identity, but is also themselves involved in moving the target by contributing to the negotiations that move it. The later career, we may therefore hypothesise, is characterised by the individual's contributions to these negotiations of what it means to be a member of the group.

Throughout this section, I have linked the concept of collective identity both to the idea that individuals move through *careers* in a social movement context, changing their behaviour and identity as they do so, and to the link between language and (changing) identity. These three threads lead us to the question: how might an individual's discourse indicate their progression through a career path of movement-based collective identity? Taylor and Whittier (1992) provide a possible framework for this. The idea of *boundaries* allows us to consider how the individual discursively positions themselves in relation to other actors, including the group itself as well as its opponents and allies. Relatedly, a consideration of *consciousness* sheds light on the individual's representations of the social relations between in- and out-groups. Finally, the *negotiation* of the meaning of practices and symbols offers a view into individuals' usage of community-specific linguistic resources, including how they use this repertoire of resources to signal their identification with and belonging to the community at different points in their career.

3.7 Symbolic capital

So far, I have laid out how, across their careers, individuals are able to acquire and leverage new resources in order to perform an identity which is situationally appropriate for the

movement group context. This process implicates the acquisition of another, related resource: that of *symbolic capital*. Developed by Bourdieu (e.g. 1984, 2004), the idea of symbolic capital builds on, but is differentiated from, a “narrowly materialist” (Crossley, 2014, p. 86) Marxist understanding of capital. Bourdieu’s formulation recognises that power and authority – defined in this thesis, following Bourdieu (1991) and Newsome and Grant (in press), as the status of being *listened to*, able to persuade and influence – are derived not only from financial wealth (economic capital) but also from other, more symbolic means (symbolic capital). Symbolic capital is defined as cultural and social resources which allow those who hold them to claim authority and prestige in a given context.

Unlike economic capital, symbolic capital is not necessarily related to the intrinsic value of objects or practices. Social recognition, rather than intrinsic value, is central to symbolic capital (Crossley, 2014). Certain objects and practices become, in a given society, imbued with socially determined value which is often mistaken for intrinsic value.¹ These objects and practices are, therefore, associated with power and prestige, and those who partake in them can be said to have symbolic capital; the objects and practices convey that capital (Moore, 2014). Individuals who share similar types of capital will recognise one another’s capital and may perceive themselves as part of the same group (Julien, 2015). However, capital is also often recognised as such even by those who do not have capital themselves. This ability to recognise different types of capital, who has it, and who does not, is the basis of social class and other groupings.

Bourdieu describes a number of different types of symbolic capital across his oeuvre, but those most commonly referred to in the literature are *social* and *cultural* capital. Social capital is “the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition – or in other words, to membership in a group” (Bourdieu, 2004, p. 21). Social capital allows individuals to claim and draw on the reputation or “collectively owned capital” (Bourdieu, 2004, p. 21) of the group. It therefore “exerts a multiplier effect” (Bourdieu, 2004, p. 21) on the individual’s own capital. The most valuable networks are large and consist of people who have capital in their own right. This includes the possession of *cultural capital*.

Cultural capital can be summarised as the individual’s proximity to and familiarity with prestigious forms of culture and cultural expression; it is “the symbols, ideas, tastes, and preferences that can be strategically used as resources in social action” (Scott, 2014, p.

¹ Of course, objects and practices that are seen as socially desirable are often priced accordingly, and expensive items may be seen as socially desirable as a result of their cost. This relationship is empirically common, but not necessary.

142). Cultural capital takes three forms. The first is *objectified* capital, which refers to the “cultural goods” (Bourdieu, 2004, p. 17) which are associated with cultural prestige or ‘legitimate culture’ (Bourdieu, 1984). Examples of objectified cultural capital include artworks, instruments, writings, galleries, libraries, opera houses, and other cultural artefacts which are considered ‘distinguished’ rather than ‘popular’ or ‘vulgar’ (Bourdieu, 2004; Moore, 2014). The second form of cultural capital is *institutionalised*. This is cultural capital that is formally recognised or reified by some institution. For example, the acquisition of a degree functions as a certification that the holder has cultural capital. The degree-holder can then ‘demonstrate’ their capital without needing to perform it interactionally. The third and final form of cultural capital is *embodied* capital. This is the “long-lasting dispositions of the mind and body” (Bourdieu, 2004, p. 17) that allow individuals to enact their cultural capital unconsciously, as second nature. Embodied capital is synonymous with what Bourdieu calls the *habitus* (Moore, 2014), the individual’s natural preferences for certain types of socially valued practices. While anyone can engage with objectified forms of cultural capital, for example by visiting an art gallery, only those with embodied cultural capital are able to engage with these unselfconsciously, as part of their natural cultural milieu (Bourdieu, 1984).

As the nature of embodied capital suggests, symbolic capital is not easily acquired. There is more to symbolic capital than simply ‘going through the motions’ of participating in certain practices. Particularly in the case of social class, symbolic capital largely reproduces itself through the class system; children raised by parents who have symbolic capital unavoidably acquire symbolic capital of their own, while children who are not raised by such parents are unlikely ever to acquire symbolic capital. In other words, “the fast, easy accumulation of every kind of useful cultural capital, starts at the outset, without delay, without wasted time [...] for the offspring of families endowed with strong cultural capital” (Bourdieu, 2004, p. 18). Under such conditions, it is acquired “quite unconsciously” (Bourdieu, 2004, p. 19).

Symbolic capital can also be acquired subsequently, but this is time-consuming and effortful, in much the same way as the acquisition of new identify performances (§2.4). Capital is only truly acquired when it is *embodied* – in other words, when it is reflected in the individual’s natural predispositions to act in particular ways and to engage in particular practices.

Acquiring these internal “logics of selection” (Moore, 2014, p. 104) takes time, as it requires consistently making the ‘right’ choices in the ‘right’ contexts until it is done unthinkingly. The importance of repetition is reminiscent of Collins’ (2004) discussion of rituals. For Collins, rituals are imbued with symbolic meaning only through repeated practice. This meaning “builds up over time, as one comes to feel the [...] resonances more deeply” (Collins, 2004, p. 153). In other words, initial engagement with ritualised or normalised group practices involves self-consciously ‘going through the motions’ of the practice. Through repetition,

however, the individual comes to understand the meaning of the practice, and to feel natural in their engagement with it. In this way, participation in or use of group-specific *rituals* (Collins, 2004), *reifications* (Melucci, 1996), or *resources* (Grant & MacLeod, 2020) across time and the career reflects the acquisition of a group-specific embodied symbolic capital.

The assimilation towards an interactionally (re)negotiated collective identity performance across the career can, therefore, be framed as the acquisition of symbolic capital. This allows us in turn to understand the implications of capital acquisition within the framework I have developed so far in this chapter. First of all, the acquisition of symbolic capital across the career endows individuals with greater power and authority within the community. For Bourdieu, capital and power “amoun[t] to the same thing” (Bourdieu, 2004, p. 16). The late-career individual, who has acquired capital within the group, therefore has authority within it. This may link back to the power of established members of the collective identity group to shift that collective identity through interaction (see §3.4.2 and §3.6). Relatedly, the acquisition of capital means that individuals have more resources at their disposal for their identity performance; new resources do not replace existing ones, but rather are added to the individual’s repertoire. As Bourdieu (1991) highlights, the possession of more capital, and thus more resources, places individuals in a privileged position compared to those with fewer resources. A larger repertoire means that individuals are able to choose from a wider range of potential performances in a given context. Meanwhile, their increased authority means they can ‘get away with’ a wider range of performances, including those which less powerful individuals may be criticised for. Individuals with less capital are, therefore, more constrained in their performances, having less access to resources and less freedom to perform.

3.7.1 Beyond social class

Although Bourdieu’s theory of symbolic capital was developed to account for the self-reproducing nature of social class, it can be taken beyond class analysis to investigate power, and the acquisition of that power, in any given group context. Forms of symbolic capital that are relevant to the present work include *linguistic* capital and *activist* capital.

3.7.1.1 *Linguistic capital*

Bourdieu (1991) recognises linguistic capital as a subtype of cultural capital. Further, it can be understood as an *embodied* form of capital, concerning the repertoire of linguistic resources individuals have which allow them to perform appropriately and with authority across interactions in a given cultural or subcultural context. As Thompson (1991, p. 13) summarises, the possession of linguistic capital gives individuals “a feel for the game” of an interactional context.

Bourdieu's discussion of linguistic capital stems from a criticism of linguists like de Saussure (1995) and Chomsky (1965) for their focus on *langue* or *competence* as the object of linguistic study, without attention to the social context in which a language and its different forms are used. In this way, Bourdieu (1991) argues, linguists may uphold restrictive and state-sponsored ideologies of language, by subscribing to the notion that languages only have one legitimate or 'true' form. As well as the grammatical competence described by Chomsky, Bourdieu argues that speakers also have a "practical competence [...] by virtue of which they are able to produce utterances that are appropriate in the circumstances" (Thompson, 1991, p. 7), where *appropriate* means effective and socially valued in the group or cultural context. The difference between grammatical and practical competence is the difference between producing grammatical sentences and "sentences that are likely to be *listened to*" (Bourdieu, 1991, p. 55). Individuals with linguistic capital have a high level of practical competence across contexts and interactions.

Much like other forms of cultural capital, then, some linguistic variants – those which Bourdieu (1991) calls *legitimate language* – convey power and prestige, while others do not. The socially dominant have the power to set the standard of legitimate language, and judge whose linguistic behaviour is acceptable and practically competent. In this way, language-based differences reflect social relations in the society or group at large, a "*re-translation* of a system of social differences" (Bourdieu, 1991, p. 54) in linguistic terms. As summarised above, individuals with linguistic capital not only have access to prestigious and powerful resources, they also have the ability to draw on more resources from their larger repertoire, performing comfortably in a wider range of interactions. In this way, language users with more linguistic capital have a high degree of agency in a given interactional context, and are able to determine and even set the standard for which options are appropriate in that context.

The concept of linguistic capital has typically been used at the level of whole societies, characterising the relative prestige of different languages or regional forms of the same language. Many researchers have described the linguistic capital afforded by prestige languages (particularly English) as an important access requirement in many international and transnational contexts (e.g. Morrison & Lui, 2000; Roth, 2019; Silver, 2005; Smits & Gündüz-Hoşgör, 2003). However, I will take a different approach to linguistic capital. Rather than comparing the linguistic capital afforded by resources associated with different communities, I consider how individuals are able to differentially use community-specific linguistic resources to index their own power – or lack thereof – in an online, text-only environment. In line with the literature discussed in §3.4.2, I hypothesise that those with the longest careers will possess the most linguistic capital, and, therefore, have greater linguistic

agency, having developed the capacity to choose between a wider range of available linguistic resources.

3.7.1.2 *Activist capital*

Several researchers have made reference to the concept of *activist capital*, although there is little consensus on what this kind of capital might be, or how it is acquired. Most of these conceptions are removed from Bourdieu's own understanding of capital. Van Dyke and Dixon (2013) introduce 'activist human capital', which they define as "activist-relevant skills and knowledge" (p. 198). This is to be understood as a career contingency, as the acquisition of such capital empowers individuals to continue their activism, feeling that they are able to use their skills to help the cause. Activist human capital is explicitly linked with social capital, as individuals acquire it in part from other activists in their networks.

Elsewhere, Chidgey (2021) describes how celebrity feminists can develop 'activist capital' – the appearance of authentic belief in the cause – through performances of care and emotional labour. In turn, social movements can benefit from the 'celebrity capital' – essentially, media exposure – that celebrities bring to the cause. Zihnioğlu (2023) adopts the alternative civic engagement definition of social capital associated with Coleman (e.g. 1988) and Putnam (e.g. 1995) – see Julien (2015) and Adkins (2008) for a comparison of these views with Bourdieu's – and adapt it to the activist context, and defining it as "the capacity to collectively defy, resist and contend" (Zihnioğlu, 2023, p. 124). She uses this concept of activist capital to explain how movement groups keep up momentum for the cause during times of enforced abeyance in the face of political and police pressures. Only Husu (2013) explicitly works within a Bourdieusian framework. She does not consider activist capital, but rather how other forms of symbolic capital, such as cultural capital, affect activist movements. Activists can draw on the symbolic capital associated with their identities in order to advance their cause. As Husu highlights, the possession of symbolic capital allows activists to "cognitively inven[t], interpre[t] and articulat[e]" (Husu, 2013, p. 272) social problems in a way that appears legitimate to socially dominant groups.

Related to the concept of activist capital is that of collective action repertoires (Doherty & Hayes, 2019; Tilly, 1993), defined as "a limited set of routines that are learned, shared, and acted" (Tilly, 1993, p. 265) for the purpose of collective action. These 'routines', which can be understood as resources for action, are protest tactics such as petitions, strikes, and public meetings. While a property of movements and groups rather than individuals (Tilly, 1993), collective action repertoires share similarities with individual (linguistic) repertoires in that they represent a set of options for "public performance" (Doherty & Hayes, 2019, p. 271) from which groups can choose in order to promote their cause, and, simultaneously, encode and present their collective identity. Repertoires are linked directly to identity by Jasper

(1997), who notes that tactics are “emotionally and morally salient in these people’s lives [...] and] express protestors’ political identities” (p. 237). If action repertoires are linked with collective identity, then, the ability to draw on these repertoires as part of a collective can be linked to the individual’s assimilation towards that collective identity, and their acquisition of group-specific symbolic capital. The activist human capital that Van Dyke and Dixon (2013) describe, which individual activists develop through their career, gives activists the skills to contribute towards the tactics in their movement’s repertoire. This mirrors the way the individual’s own linguistic repertoire is drawn from, and in turn contributes to, the community-level *local linguistic repertoire* (Gumperz, 1982; see §2.4.1).

3.8 Bridging the gap: The career in the white nationalist forum

3.8.1 Collective identities online

It has long been known that collective identity is concerned with the ongoing production of shared meaning through interactional (re)negotiation (Melucci, 1996). Despite social media being “one of the main meaning-making machines of our times” (Milan, 2015, p. 897) and “one of the main places where the symbolic work of contemporary movements occurs” (Milan, 2015, p. 889), research into the construction of collective identity online remains in its infancy. Flesher Fominaya (2019) describes it as a ‘new frontier’ in social movement research. Indeed, much of the research which has been completed to date appears in a single special issue dedicated to the topic (Gerbaudo & Treré, 2015).

The very idea that collective identity can emerge online remains somewhat controversial in the field. Critics have focused on the supposed *networked individualism* (Wellman et al., 2003) of the internet, which fosters loose networks of individuals rather than the strong solidarity associated with collective identity. For Bennett and Segerberg, key scholars in the case against online collective identity, online activism involves “self-motivated [...] sharing of already internalized or personalized ideas” (Bennett & Segerberg, 2012, p. 753), rather than group negotiation of shared meaning. Social media has also been criticised as facilitating only short-term and fragile commitments rather than collective identities which endure over time, leaving online movements constantly vulnerable to collapse and disappearance (Juris, 2012). As Poell and van Dijck (2018) highlight, the design of social media (especially social networking) platforms can be at cross purposes with the development of collective identity: topics and hashtags can only trend when there is a spike of engagement, so sustained engagement over time will reduce visibility of the movement. These qualities of social media and their effects on movements online are also recognised by scholars who accept and promote the idea of computer-mediated collective identities (e.g. Bakardjieva, 2015; Firer-Blaess, 2016; Gerbaudo, 2015; Kavada, 2015; Milan, 2013, 2015). As we will see below,

these researchers incorporate these qualities of social media movements into their conceptualisations of collective identity, shifting the concept away from Melucci's (1996) formulation in its adaptation to a very different new environment, while still retaining key elements of Melucci's (1996) definition.

Some scholars have also argued for the importance of "[p]hysical copresence" (Collins, 2004, p. 23) for the development of collective identity. For Collins, this is essential for the imbuing of practices with shared meaning. He argues that online rituals will always be "weak substitutes" (Collins, 2004, p. 62) of their offline equivalents, citing weddings and funerals, sports games and musical performances, and religious ceremonies as evidence for his claim. His core argument is that "[w]ithout bodily presence, it is hard to convey participation in the group and to confirm one's identity as member of the group" (Collins, 2004, p. 54). In the social movement context, Kavada (2015) contrasts online and face-to-face participation in the Occupy Movement, noting both that the practicalities of involving online participants in decision making processes meant that they were often "relegated [...] to spectators" (Kavada, 2015, p. 880) of the movement, and that most of the activists she interviewed felt that "it was regular involvement face-to-face that actually rendered someone an 'Occupier'" (Kavada, 2015, p. 879). Notice, however, that Collins (2004) cites only examples of long-established offline social rituals. Kavada's (2015) example is of a movement with a necessarily face-to-face component, namely the physical occupation of the Wall Street district and other similar locations worldwide. It is difficult to imagine, on the other hand, that Zappavigna's (2011) live tweeters, Newon's (2011) World of Warcraft guild members, or Firer-Blaess' (2016) hacktivists feel that there is 'something missing' from their practice or that they are any less a member of a group or community. These practices are developed online and are by their nature only performed online. The collective identities that develop around these shared practices are, therefore, also online in nature.

A small number of scholars have described collective identity in a selection of online-only and hybrid movements. For Futrell and Simi (2004) and Bakardjieva (2015), the accessibility of activism via social media plays a key role in solidarity-building, pulling in new and peripheral members and maintaining commitment of key members by creating a sense of "participating in a much larger movement culture" (Futrell & Simi, 2004, p. 37). The ubiquity of social media means that its use for activism purposes also serves to imbue this everyday practice with movement-related meaning, which Taylor and Whittier (1992) identify as a key element of collective identity construction (§3.6). Taylor and Whittier's (1992) consciousness raising can also be achieved online; the individual may realise that their "personal problems are actually political" (Bakardjieva, 2015, p. 985) through content shared by members of their loose online networks. Additionally, this everyday engagement with the movement

“continuously activate[s] the relationships that maintain collective identity” (Milan, 2015, p. 893) according to Melucci (1996). Social media also allows for the archiving of, and easy access to, movement-related artefacts and cultural products. These “archives can be useful tools for community consciousness” (Firer-Blaess, 2016, p. 39) for both new and existing members. As Kavada (2015) highlights, the digitally-mediated nature of these artefacts means that they can be “living documents” (p. 879) which reflect renegotiations of the collective identity in real time. Other collective identity artefacts, such as iconography and jargon, can also be created and sustained online. As Gerbaudo (2015) points out, these contribute to the self-recognition as, and the recognition by others of, a collective, which is key to collective identity (Melucci, 1996). Gerbaudo (2015) discusses this in the context of movement-related profile pictures on social networking sites, arguing that this is a significant act of identity involving “the temporary renunciation of one’s own individual identity” in order to highlight that “they are also particles in a collective actor” (pp. 925-6).

As noted above, the consideration of collective identity online involves “re-interrogatio[n]” (Flesher Fominaya & Gillan, 2017, p. 398) of the concept as described by Melucci (1996) in the context of face-to-face movements. The first of these concerns the increased individualism associated with social media. A degree of individualism is not inconsistent with Melucci’s (1996) views, which naturally leave room for individual opinion and disagreement (Bakardjieva, 2015; Firer-Blaess, 2016) – without this, the collective identity would never be renegotiated, and this renegotiation and “self-reflexive capacity” (Bakardjieva, 2015, p. 984) is central to Melucci’s understanding. Consensus need not be total but only sufficient to sustain action, and it is never stable. Firer-Blaess (2016) describes the collective identity of the entirely online and highly decentralised Anonymous movement. He finds that while not all members accept all elements of the overall collective identity, “all accept and express a number of them that is sufficient to legitimate their own belonging to the movement and to be recognised most of the time by other Anons as such” (Firer-Blaess, 2016, p. 190). More controversial elements of the collective identity are discursively renegotiated, as Melucci (1996) predicts. Milan (2015) shares a similar perspective, arguing that consensus emerges out of individual ‘votes’ in the form of “selecting, highlighting and sharing” content and ideas, “thus determining what fits into the collective narrative” (Milan, 2015, p. 895). Milan also argues that online individualism may be seen as a strength for collective identity. Individuals share their own experiences and perspectives that they feel a personal connection with. The sum of these individual contributions creates “personalized yet universal narratives, connecting individual stories into a broader context that gives them meaning” (Milan, 2015, p. 894). This “facilitates the alignment of personal and collective identities” (Milan, 2015, p. 895), thus strengthening individuals’ connection to the movement.

Responding to criticisms of 'slacktivism' (e.g. Morozov, 2009), Gerbaudo (2015) highlights that online collective identity networks "are not meant as substitutes for 'real' political action, but are crucial in creating the sense of 'we' which constitutes a fundamental pre-condition for protest movements to emerge" (p. 927). This contrasts with Melucci's (1996) strong focus on collective action in his discussion of collective identity (both scholars agree that collective identity is a prerequisite for collective action). While disagreeing on the nature of the relationship between online spaces and 'real political action', Firer-Blaess (2016) follows Gerbaudo (2015) in the belief that there is more to collective identity than simply collective action. Although contrasting with Melucci's approach, Firer-Blaess defends this position with reference to Melucci's own (1996) concept of the *submerged network*, "a social network larger than the interactions concerned with activist actions" (Firer-Blaess, 2016, pp. 72-73). For Melucci, collective identity is (re)negotiated within the submerged network: "brief and intense public mobilization campaigns [...] are fed by the submerged life of the networks and their self-reflective resources" (Melucci, 1996, pp. 392-393). Firer-Blaess (2016) argues that collective identity must, therefore, logically include other elements beyond collective action. He identifies several such illustrative examples in the context of the collective identity of Anonymous: a deviant counterculture based on offensive content; a horizontal organisation structure; anonymity; and, interestingly, *universality*, a shared "claim of non-identity" of the movement, which Firer-Blaess (2016) characterises as "a collective identity definition in itself" (p. 188).

Scholars of online collective identity also recognise the often temporary nature of these identities, and this has been used to question the very idea of online collective identity formation. Reflecting on these criticisms, Bakardjieva (2015) asks: "Does their 'we' last long enough to become self-reflexive as an agent?" (p. 987). As the discussion above shows, several scholars have argued that these communities do reflect on and renegotiate their collective identity. For Milan (2015), a collective identity that turns out to be short-lived may nonetheless be "strong in the present" (p. 896). This seems comparable to the distinction between *thick* and *light* identities introduced by Blommaert (2018; see §2.5.1). Traditional, face-to-face social movements often represent *thick* collective identities based in "race, gender, class or ethnicity [...] the state and the family, the guild and the church" (Blommaert, 2018, pp. 63-64). Meanwhile, social media spaces lend themselves to *light* identities, often ephemeral, in which "aggregations of people [may] shar[e] just the rules of the encounters, but little beyond it" (Blommaert, 2018, p. 65). The rules of these encounters may be understood as the (temporarily) shared meanings and norms that make up the *light* collective identity. Although ephemeral, these shared identity performances are no less strong or real "in the present" (Milan, 2015, p. 896).

As Goffman (1956) outlined, group identity (re)negotiation often takes place backstage. In order to give an impression of authenticity to the group identity, and thus to protect its members' face needs, only unified performances of that identity appear in front of audiences on the frontstage, leaving the (re)negotiation process invisible to those outside the group. This is mirrored in the context of collective identity, which is (re)negotiated within the *submerged network* (Firer-Blaess, 2016; Melucci, 1996) and only displayed for an audience in "public mobilization campaigns" (Melucci, 1996, p. 392). However, this is not the case in online forum interactions, where all posts can be read by any member or guest to the community. This environment allows for no backstage area where members can (re)negotiate their collective identity and the meaning of their practice; instead, these interactions are necessarily carried out in frontstage forum posts, in full view of any reader of the site. Online forum interactions therefore provide privileged access to the collective identity (re)negotiation process as it is expressed discursively on the site. This transparency allows the researcher the same access to the collective identity as the aspiring community member, therefore allowing for the process by which individuals assimilate to that collective identity to be empirically observed.

3.8.2 Language and collective identity

Despite the Goffmanian sociological origins of the concept of identity performance (Goffman, 1956), and the more explicitly linguistic perspective Goffman took to this in his later work on footing (Goffman, 1981; see §2.2.2), since that time the overlap between sociological and linguistic work on (collective) identity creation has been minimal.

This is not due to a lack of recognition in the social movements literature that language plays a key role in the construction, maintenance, and renegotiation of collective identity. Indeed, this fact is mentioned frequently, if not always overtly. Melucci's (1996) central belief is that collective identity is realised through "exchanges, negotiations, decisions, and conflicts" (p. 4), which are naturally carried out discursively. He also explicitly mentions that groups may codify elements of their collective identity "within a language [...] that is specific to a group" (p. 70). Both Melucci (1996) and Polletta and Jasper (2001) recognise that collective identity is expressed through "cultural artifacts" (Melucci, 1996, p. 70) or "cultural materials" (Polletta & Jasper, 2001, p. 285), including "names, narratives, symbols, [and] verbal styles" (Polletta & Jasper, 2001, p. 285). Collins also recognises that jargon, front- and backstage discussions and negotiations, and topics of discourse are crucial "membership symbols" (Collins, 2004, p. 86). Polletta (2006) expands on the importance of narratives, finding that they are used to encourage participation, recount motivations, and outline moral meanings in accessible terms. Hund and Benford (2004) similarly show that talking through past

collective actions can imbue these with new meaning, which serves both to align personal and collective identities and thus sustain participation, and also to provide new opportunities and ideas for future action. They summarise: “[f]undamentally, collective identities are *talked* into existence” (Hund & Benford, 2004, p. 445; emphasis in original), mirroring boyd’s (2008) observation that online identities are *written into being*. In the online context, Gerbaudo and Treré (2015) recognise that new symbols, lexicons, and naming practices have played a central role in the development of recent social movement identities. In the domain of careers, Becker (1973) notes that “[t]he vocabularies in which deviant motivations are phrased reveal that their users acquire them in interaction with other deviants” (pp. 30-31). In other words, group members’ changing linguistic performance is evidence that the individual has assimilated towards the identity and norms shared by the group. However, although mentioned in passing, the link between collective identity and language has rarely, if ever, been explored further in a linguistically-informed way.

As I have made clear, however, language cannot be ignored or hand-waved when discussing the construction of identity, especially in largely text-based computer-mediated contexts. The individual’s progression through their career, their assimilation to the collective identity, their acquisition of capital, are all facilitated and evidenced only through language. In the online forum context of this thesis, we can explore language in a social movement in its ‘pure form’, unmediated by other elements which arise in face-to-face contexts.

3.8.3 Research questions

In light of the above discussion, I can now refine the broad research aims outlined in Chapter 1 into a set of research questions. Ultimately, this project aims to provide a better understanding of the nature of white nationalist identity as expressed in a particular online forum, and in particular to consider whether and how individual identities change over time as they become further enmeshed in this extremist community across their career. While being shaped by the literature, these research questions have also, as I describe in more detail in Chapter 4, been developed through an inductive research process in which subsequent research questions are informed by the findings of earlier studies. This process has resulted in the following major research question:

What does the career look like in the context of this online white nationalist forum?

I approach this research question using a set of more focused sub-questions:

RQ1: *How, if at all, can the structure of the career be conceptualised in a stable and reproducible way?*

RQ2: *How far are the individualised vs. structural approaches to defining the career supported in this forum context?*

a: *From a temporal perspective, to what extent, if any, can a typical career path be said to exist?*

b: *To what extent, if any, are longer careers associated with the accrual of linguistic capital in the form of a wider repertoire of linguistic resources? What resources are acquired, if any?*

c: *To what extent, if any, do members appear to radicalise across the career? For example, through more extreme boundary drawing and dehumanisation of out-groups, and/or greater acceptance of violent means to achieve ideological ends.*

RQ3: *How does the career observed in the forum intersect with the development of collective identity, in terms of its three elements (boundaries, consciousness, and negotiation)?*

Chapter 4 Methodology

4.1 Introduction

This chapter describes the research design for this thesis. I begin with a high-level discussion of my interdisciplinary and multimethodological research design, justifying this approach with reference to my social constructionist research philosophy. The dataset, introduced in §1.1, is described in more detail, followed by the various sociological and linguistic methods employed in the thesis. The chapter concludes with a consideration of ethical issues.

4.2 An interdisciplinary approach

As I have established across the previous chapters, in this thesis I am taking an interdisciplinary approach, which “combines components of two or more disciplines” – here, sociology and linguistics – “in the search or creation of new knowledge” (Nissani, 1997, p. 203). Interdisciplinary research approaches remove the conceptual limits imposed by individual disciplines and, in doing so, introduce novel and creative insights to each respective discipline (Lyll & Meagher, 2012; Nissani, 1997). As Gioia et al. (2013) summarise, “[a]dvances in knowledge that are too strongly rooted in what we already know delimit what we can know” (p. 16). By considering both sets of perspectives, gaps and shortcomings from each discipline can be identified and addressed in ways which are not always apparent to researchers working within a single discipline (Nissani, 1997). In Chapters 2 and 3, I have already demonstrated how the sociological and linguistic literatures on (collective) identity and careers overlap with and supplement one another. This has allowed me to develop a set of research questions (§3.8.3) which would not be meaningful through the lens of either discipline alone. In this chapter, I show how the interdisciplinary approach can be used to move beyond disciplinary constraints by selecting the most suitable methods from both disciplines in order to answer these research questions.

Naturally, then, this necessitates the adoption of a multimethodological approach. Following Clark et al. (2021), I define this as the “combination of more than one method within a particular [research] design” (p. 516). As I detail in §4.4, in this thesis I draw on a range of methods, both qualitative and quantitative in nature, and taken from both the sociological and linguistic traditions. A multimethodological approach allows me to address my research questions more fully, offering a “more comprehensive account” (Bryman, 2006, p. 106) of the career and collective identity in this forum. As I demonstrate in Chapter 5, studies of the career should include consideration of two main elements. The first is the *temporal* element,

which I investigate in Chapter 6 using *sequence analysis* (Gabadinho et al., 2011; Gabadinho et al., 2010), a quantitative method developed for sociological research. The second is the *behavioural* element, which in online contexts is best explored using linguistic methods (Seargeant & Tagg, 2014; Tagg, 2015); I therefore undertake linguistic analyses in Chapters 7 and 8. The findings of Chapter 5 show that I “cannot access everything [I] need to know” (Clark et al., 2021, p. 558) using just one method. These plural methods allow me to “offset” (Bryman, 2006, p. 106) the weaknesses of each method by drawing on the strengths of other, complementary, methods. In particular, as I note later in this chapter (§4.4.3), corpus linguists frequently point to the value of combining quantitative and qualitative corpus-based approaches, as I do in Chapters 7 and 8 respectively, to find a balance between breadth and depth of analysis (e.g. Baker, 2006).

Lyall and Meagher (2012) highlight that interdisciplinary projects, more so than others, often require a greater level of initial testing and exploration of methods in order to find the best approach to answering the research questions. This has been the case for the present research, and I have therefore taken an inductive approach to developing my research design. Contrary to a theory-testing deductive approach, the inductive approach builds theory through the research process by “drawing generalizable inferences out of observations” (Clark et al., 2021, p. 20). An inductive approach is recommended for novel research contexts such as the interdisciplinary one presented here, which offers little pre-existing knowledge or testable theory (Jebb et al., 2017; Miles & Huberman, 1994). As I detail in §4.4, the design of each subsequent study in this thesis follows directly from the findings of earlier studies.

The approach I have described above is consistent with the social constructionist philosophy I have already outlined in relation to identity in §2.3. In the context of identity, social constructionism is the view that identity is constructed entirely through linguistic performances and (dis)alignment from other social actors. More broadly, the philosophy holds that all aspects of (social) reality are socially constructed, emerging through repeated interactions which (re)negotiate and reinforce participants’ view of the world. Under this view, both identity and reality are seen as dynamic and unfixed.

Social constructionism is a broadly relativist philosophy (Andrews, 2012; Burr & Dick, 2017; Irwin, 2011), seeing knowledge as dependent on human perceptions of the world. This contrasts with a realist approach to understanding the nature of knowledge, which assumes that there is a single, objective truth which is “external to actors” (Clark et al., 2021, p. 28) – that is, independent of any given individual or society’s perception of the world. Andrews (2012), following Hammersley (1992), adds nuance to this characterisation of social

constructionism as a wholly relativist ontology, arguing that while the world does have a reality which is independent of human perceptions, there can be no “direct access to that reality” (Andrews, 2012, p. 42). Burr and Dick (2017) illustrate this position using the example of trees. Trees have “various ‘natural’ characteristics” (p. 59), such as their height, bark, and leaf shape. However, what we perceive as a tree, rather than, say, a shrub or a flower, is “largely a consequence of how classifications [...] are produced through language rather than being a natural consequence of our perceptual capabilities” (p. 59). Phenomena we perceive as objective reality are therefore simply the result of ideas becoming “institutionalised by society to the extent that future generations experience this type of knowledge as objective” (Andrews, 2012, p. 41).

A social constructionist philosophy, therefore, “looks beyond [...] taken-for-granted notions” about the world (Irwin, 2011, p. 100). This has two main implications for this work. The first is that, because knowledge is not assumed to be essential or objective, the social constructionist researcher should avoid imposing *a priori* assumptions or hypotheses on the data. Instead, they should allow theory – or classifications of the world – to emerge from the data through an inductive research process, as I do for the present study. The second implication stems from the idea that these ‘taken-for-granted’ claims are normalised and propagated by the socially powerful (Burr & Dick, 2017). Discourse reflects and reproduces the social positioning of actors in relation to one another and to the world. Social constructionism therefore serves as an analytical strategy, allowing me to uncover the identity claims – claims about the collective, and the position of the individual in relation to it – which are made through linguistic performance in the forum.

4.3 Data

The data for this project is taken from a long-standing online forum which describes itself as a ‘white nationalist community’, where members primarily discuss social and political issues, with a particular focus on race. However, users may also engage in non-ideological talk (Holt et al., 2020; Wong et al., 2015). The site is organised into a large number of topic-specific subfora such as *History & Revisionism*, *Culture and Customs*, and *Musicians*. The primary language of the forum is English, although local subfora exist which use other languages.

Forum members can begin new threads in the forum, or respond to existing threads. Typical of a discussion forum, communication is asynchronous, meaning that participants need not be (virtually) co-present for the discourse to occur. Once created, posts remain on the site forever. The site’s guidelines for posting warn members of this, and advise them to exercise caution in their posts for this reason. In rare cases, posts may be deleted by the site’s administrators. For example, posts made by the convicted terrorist and mass murderer

Anders Breivik are no longer archived on the site. Members are not, however, able to delete their own posts. The length of posts is “effectively limitless, or at least, it is larger than practical limits on how long most people are willing to type and others are willing to read” (Herring, 2007, n.p.). There is apparently no limit on thread length, although there may be practical constraints. Some popular threads have numbered titles, with initial posts in these threads commenting that a previous thread was ‘too big to function properly’.

The forum is primarily text-based, although members can also include images and emoji in their posts. They can also include hyperlinks to external resources, including audio and video, but these are not directly embedded into forum posts. Users are also able to directly quote others’ posts in order to respond to these posts using a dedicated quoting function which embeds (part of) another user’s post in a new response. Posts appear beneath one another in the order they were posted in the thread, with no option to reply directly to a particular post, but this quoting function makes adjacency pairs clear (Herring, 2001). Alternatively, posters may use each other’s usernames to signal who their own post is in response to. While private asynchronous messaging between members is available, this study focuses on publicly available posts.

Typically, posting threads and comments to the site requires registration with an email address for an account on the site, which is free of charge. However, there are a small number of ‘open’ subfora where individuals without member accounts (‘guests’) are able to post to the site. Non-members are, however, able to read most other threads on the site without registering for an account. This is with the exception of one restricted subforum, which is available for viewing and posting only by *sustaining members*, who support the site financially.

The poster’s status – as a guest, regular member, sustaining member, or other role such as administrator – is displayed alongside each of that individual’s posts to the site. Also displayed is the individual’s selected username (except for guests) and optional avatar. Posts to the site are pseudonymous, with individual members choosing the extent to which they wish to disguise their offline identity with their username. Most appear to adopt some level of disguise, although a limited number of users appear to have chosen their own name as a username. Users can also choose to include information about their location, which also appears alongside their username and member status. Finally, the member’s join date (month and year), and number of posts to the site, is non-optionally displayed.

4.3.1 Data access

The dataset used in this study had been scraped from the white nationalist site previously by other researchers (Kredens & Pezik, 2021) and was accessed via the Forensic Linguistics

Databank held at Aston University. The full dataset contains every message posted to the site between the years of 2001 and 2018, a total of around eight million posts representing over four billion tokens. This includes messages from over 120,000 user accounts. The dataset includes both linguistic data and metadata. Along with the text of each post, the dataset also includes the username of its poster, the date and time the post was made, thread title, and the post's position in the thread (where 1 = initial post, 2 = first reply, etc). Although as noted above posters can include images and emoji in their posts to the forum, these elements are excluded from the present dataset. Where the poster has used the quote function to copy part of another user's post into their own contribution, these quoted sections are excluded from the dataset in order to avoid repetition – only the responder's commentary on or response to the quote is preserved. Posts in languages other than English were excluded from my dataset by excluding local subfora which use a different primary language, and by running a language detection script on the remaining posts.

Each of the studies recounted in Chapters 6-8 make use of a different subset of the dataset described here. The specific characteristics of each sub-dataset are described in more detail in these chapters.

4.4 Research design

As I have described above (§4.2), I have developed my multimethodological research design following an inductive and sequential approach, allowing the design for each part of the study to be shaped by the findings generated by the preceding part(s). In this section, I provide more detail on the methodology of each study in turn, showing how each successive phase of analysis influences the next stage, combining within one holistic and iterative research design. First, multiple correspondence analysis (§4.4.1) enabled me to systematise a diverse literature on careers across contexts, which led me to investigate temporal and behavioural elements of forum careers. This provided the rationale for a sequence analysis (§4.4.2) exploring the temporal properties of careers and identifying four broad career types based on duration. The language used by these four user types was compared using a quantitative corpus-assisted analysis (§4.4.3), which identified features such as epistemic stance resources, pronouns, and aspectual verbs which appear to correlate with career length. Finally, the quantitative and qualitative usage patterns of these features were investigated in more detail across the careers of five individual user careers (§4.4.3). Figure 1 summarises this process in the form of a flowchart.

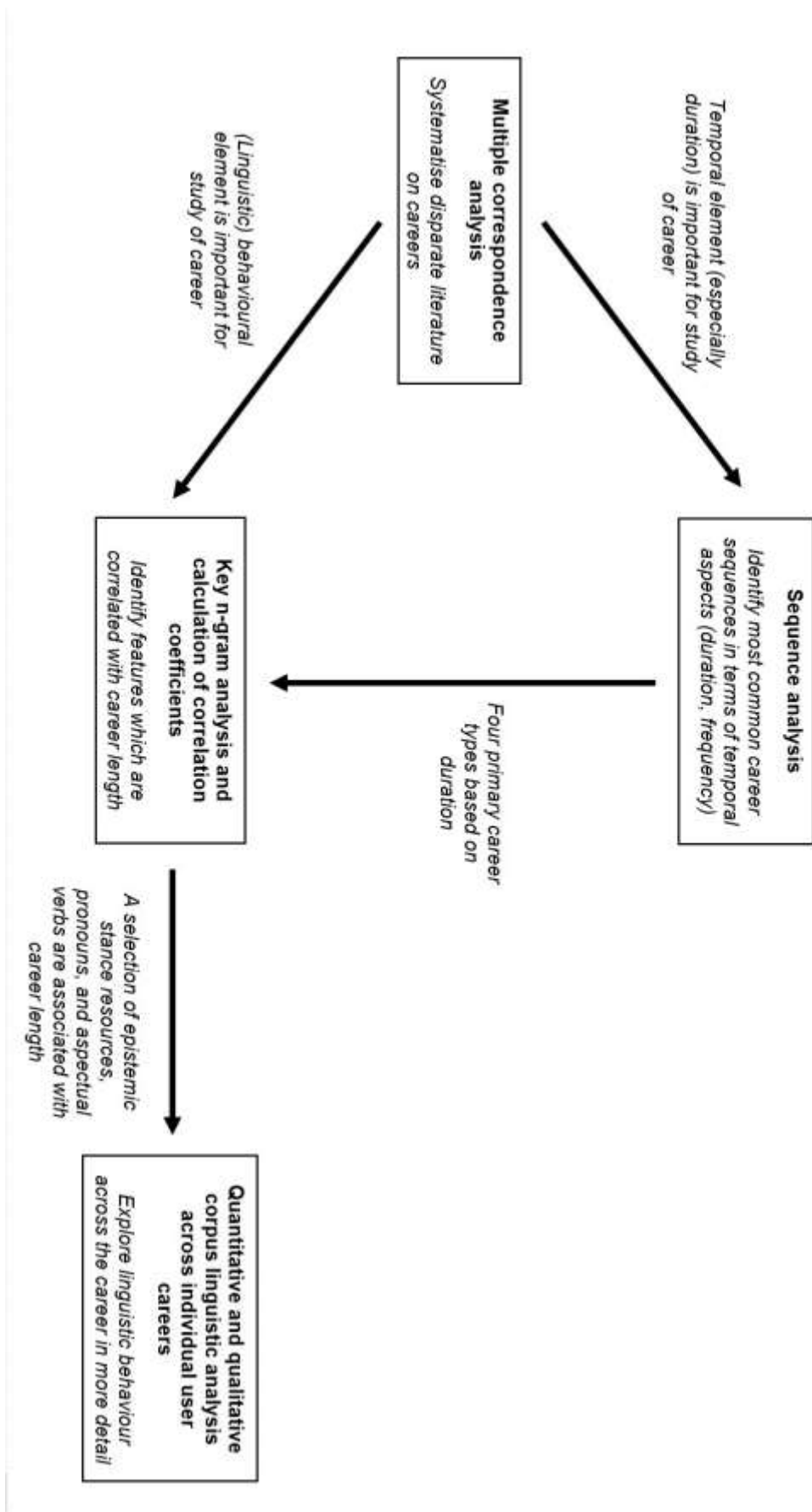


Figure 1: Simplified flowchart of emergent research methodology

4.4.1 Multiple correspondence analysis

As I discussed in Chapter 3, the sociological literature which has explicitly worked within the framework of the career has typically been abstract and theoretical in nature. It therefore leaves us with little understanding of how the concept of the career can be operationalised for empirical research of the career in a particular social context. However, beyond those papers drawing explicitly on the notion of the career, a number of studies exist which do empirically investigate what we might understand as the career – that is, they explore the roles and/or stages which members of some social group might pass through during their time as part of the group. These studies can therefore offer some guidance for the researcher of the career. However, these studies span a number of disciplines; they employ a range of different methods; and they demonstrate a variety of possible ways of differentiating careers, and their constituent roles or stages, from one another. This literature therefore requires systematisation before it can lead us to an understanding of the best way to operationalise the career concept.

Given the disparate nature of these papers and the career typologies described within them, I decided that a qualitative approach to grouping the papers and typologies and examining the underlying principles of organisation would not, by itself, be sufficient. The differences between them made it difficult to categorise the typologies into discrete, mutually exclusive groupings. This was particularly the case as some typologies were similar to one another in some ways, while differing in others. In Chapter 5, I therefore combine an initial qualitative analysis with a multiple correspondence analysis in order to explore patterns of co-occurrence in the features used by researchers to create career typologies.

Multiple correspondence analysis (MCA; Le Roux & Rouanet, 2010) is a dimension reduction technique, similar to factor analysis or principal components analysis, but designed to work with categorical datasets. For this study, I developed a dataset through manual qualitative analysis of the career typology literature, collecting information regarding the data, methods, and findings of each study. Studies of careers in both social movement and online community contexts were included, to ensure the results of the analysis were of value for the present research context. The manual analysis was systematised as far as possible by asking the same questions of each typology, including: the methodology; the nature of the community; the nature of the final typology; and the basis on which the researcher distinguished between different career roles or stages.

MCA allowed me to identify how the features I identified covary with one another, and thus to determine the main ways in which these career typologies have been developed.

Systematising literature in this way is a novel application of MCA. The dataset of career typologies I use for the analysis includes a smaller number of datapoints than is typical for MCA, which is traditionally used to explore survey data with hundreds of respondents (see Bourdieu, 1984, for a classic example). Nonetheless, it is appropriate here due to the qualitative nature of the dataset, which contains datapoints which covary in complex and sometimes unexpected ways.

The most significant finding of the MCA for the development of this thesis was that, across the literature, two key elements of the career have typically been investigated. The first is the *temporal* element, which I define as the amount of time the individual has dedicated to the group across the career. This is to be understood primarily in terms of the *duration* of the individual's commitment, but also the *frequency* with which they engage with the community. The second is the *behavioural* element. As I describe in §5.2.2.9, *behaviour* is construed broadly, and has a different meaning depending on the nature of the community being investigated and the activities members engage in as part of the community. However, further findings from the MCA, as well as other literature on online communication (Seargeant & Tagg, 2014; Tagg, 2015), suggest that linguistic behaviour is central in the case of online communities such as the white nationalist forum context. The nature of MCA means that these two elements are often researched separately, rarely appearing in the same paper. However, a "more comprehensive account" (Bryman, 2006, p. 106) of the career can be developed through a multimethodological approach. I therefore used these two elements as the backbone of the remainder of the thesis, detailed in the subsequent sections and in Chapter 6, and Chapters 7 and 8, respectively.

4.4.2 Sequence analysis

In Chapter 5, I identify that temporal elements, and particularly duration, are central to the investigation of careers. As I have shown in §3.3.2, there is little consensus in the literature on what careers may look like from the temporal perspective, and this must be explored in the specific context of the white nationalist forum. The findings of this investigation also have implications for a more fundamental disagreement in the careers literature, namely whether a typical career path might be identified (Becker, 1973; Fillieule, 2010; Goffman, 1961). In order to answer these questions, in Chapter 6 I conduct a sequence analysis to investigate patterns of engagement with the community over time.

In the social sciences, sequence analysis (Gabadinho et al., 2011; Gabadinho et al., 2010) is a method of visualising different periods (called *states*) or events in the life course or, here, career. For this analysis, I conceptualise careers as a sequence of engagements with the

community over time. Sequence analysis helps to visualise these sequences of engagement; compare different sequences; and identify more common sequences.

Following from the results of the MCA (Chapter 5), each state in the career sequence is identified as the number of posts made by each user in each calendar year of their career. Sequence analysis is designed for categorical data. Therefore, *posts per year* was reanalysed as a categorical variable, creating three categories corresponding to *low*, *moderate*, and *high* rates of engagement in a given year. This three-state model followed from patterns observed in the data and from the example of other literature examining engagement over time (Scrivens et al., 2022). Using categorical data also allows for easier identification of typical career paths, by grouping similar, if not identical, career sequences together. Sequences could then be compared in terms of frequency and of duration (number of calendar years of posting).

My characterisation of careers in this white nationalist forum is naturally limited to an analysis of those forms of engagement which leave a trace on the forum and therefore in the dataset. Some elements of the career are inherently unavailable to a web scraper, namely those periods of time where a user is *not* posting to the forum. In the absence of forum posts, it is impossible to determine an individual's true level of engagement with the site. In other words, we can only observe certain kinds of engagement.

The act of reading without posting, commonly known as lurking, is extremely common in online communities, and may be understood to form part of the career. Nonnecke et al. (2006) describe lurking as *non-public participation*, reflecting the view that lurkers experience many of the same benefits of community participation as posters do (Mo & Coulson, 2010; van Uden-Kraan et al., 2008), including a sense of group identity (Nonnecke & Preece, 2000, 2001). Lurking, then, can be understood as a legitimate career stage. However, the present dataset cannot tell us how long, if at all, users lurked prior to making their first post. Where gaps appear in a user's career, it is impossible to know whether that user was truly absent from the site, or whether they lurked in the meantime. For this reason, lurking cannot be taken into consideration in this study. The word 'career' in the context of this study should therefore be understood to refer to the visible elements of the career, perhaps more accurately described as the 'posting career'.

With this caveat, the findings reported in Chapter 6 show that most forum users follow a single typical career path, comprised of a single year of low-frequency posting. However, I judged that this most typical career was not by itself a suitable one to investigate in more detail. Given the minimal amount of time dedicated by these typical users to the forum community, we may assume that these users do not exhibit a career in the subjective sense

– that is, experiencing a change in identity through shifting perceptions of self and other over time (Becker, 1973; Scott & Hardie-Bick, 2022; Strauss, 1959), which is indexed through changing linguistic performance (e.g. Grant & MacLeod, 2020). Identifying only a single typical career also obscures the large variation in career types which I demonstrate (Chapter 6).

To facilitate the linguistic analysis in Chapters 7 and 8, I therefore decided to distinguish four broad career types, each of a different duration – one year, two years, three to six years, and seven or more years. This was determined to be a manageable number while still capturing some of the significant variation in the sequence analysis results. The four types were emergent from these results, representing 76%, 10%, 10% and 4% of users respectively.

4.4.3 Corpus-assisted discourse analysis

In Chapters 7 and 8, I return to the (linguistic) *behavioural* element of the career identified in Chapter 5. Following from the *temporal* analysis presented in Chapter 6, I take the four career types identified there as the starting point for the linguistic analysis, comparing the linguistic behaviour across the four types to investigate the relationship between behaviour and career length. In this way, the findings of Chapter 5 serve as a means of *purposive sampling*, allowing for suitable data to be selected for linguistic analysis. This is another benefit of the multimethodological approach (Bryman, 2006).

I have chosen to use a corpus-assisted approach to linguistic analysis. Corpus linguistics is suitable for the analysis of large datasets which are representative of particular varieties of language (Baker, 2006; McEnery et al., 2006), in this case the varieties produced by the four user types. In the interest of representativeness I created four very large subcorpora, comprising around 30 million words each (see §7.2.1). These were uploaded to the online corpus analysis software Sketch Engine, which is able to process extremely large datasets.

Due to the size of the corpus, I first carried out a key n-gram analysis (§7.2) as a “way in” to the data (Baker, 2012, p. 248). Keyness analysis allows for the comparison of each subcorpus with the remainder of the corpus, identifying the most distinctive or salient n-grams (word strings) for each career type. I manually analysed the key n-gram lists as a preliminary step to identify apparent differences between the subcorpora representing the shortest and longest career types, assuming that these types would present maximally different linguistic behaviour. These features could then be explored in more depth with subsequent analyses. These features were emergent from the key n-gram lists, but in places matched the existing literature on newcomer and expert language usage (§3.4), for example the expression of epistemic commitment and pronoun usage.

I then conducted corpus searches for the linguistic features identified through the key n-gram analysis, which I hypothesised to be associated with career length. I recorded the relative frequency of each feature for each of the four career types, then calculated the correlation between relative usage of each feature and length of career. The correlation calculation was conducted not as an end in itself, but was used as “first piece of the [...] puzzle” (Hunter, 2022, p. 77) by identifying features of interest for further investigation. Following Evans (1996), features with a weak or very weak correlation (60% of features) with career length were discarded from further analysis, while those showing at least a moderate correlation (40%) were used for further analysis. The threshold of a moderate correlation was used to ensure that a range of features were investigated, as the strongest correlations were primarily shown only for pronominal features.

At this stage, I have identified features which correlate with career length, in the sense that they are more commonly associated with either the shortest or longest career type. In Chapters 7 and 8, I investigate whether the same pattern is observed with individual careers – that is, are features associated with short careers *also* associated with the early stages of long careers, and are features associated with long careers primarily associated with the later stages of those careers?

In order to investigate the use of linguistic features across the career, I selected the five most prolific users from the longest career type. The selection of users with long careers for this sample allowed for the exploration of these users’ identity and linguistic behaviour across the entire career. Although the features assumed to be associated with early career stages were based on the linguistic behaviour of users with short careers, no users with short careers were selected for the sample, as it was assumed that these users were unlikely to show identity change across their careers. The decision to include five users for more detailed analysis, out of over 4,000 users who follow the longest career path, was made partially due to time constraints. It would not have been possible to carry out this analysis on a representative proportion of those 4,000+ users. However, I judged that the inclusion of five users would allow me to determine whether or not any general patterns of linguistic behaviour emerge across different users, and if so, to describe those patterns with a degree of accuracy.

In §7.4, the quantitative corpus-assisted approach is continued on the individual level. Five corpora were created, each containing every post made by one of the five selected users. The posts were ordered chronologically. This allowed me to use Sketch Engine’s *distribution of hits* tool to explore the relative frequency of each feature of interest across ten

chronological deciles of each user's career. In this way, the individuals' linguistic behaviour across the career can be quantified and visualised.

In Chapter 8, the linguistic behaviour of the five users across the career is explored from a qualitative perspective, analysing concordance lines, or examples of the features of interest in their discourse context. Bryman (2006) has argued for the value of combining quantitative and qualitative approaches in order to offset the shortcomings of each method, and offer a more complete understanding of the phenomenon under investigation. These advantages have been widely recognised among corpus linguists (e.g. Baker, 2006; Brindle, 2016; Egbert et al., 2020). Egbert et al. (2020) recommend concordance analysis as a way of moving past the abstractions of quantitative data and “get[ting] back to the actual language contained in the corpus” (p. 52), and demonstrate how “rich information” and patterns may be “missed entirely” (p. 57) by using only quantitative analysis. In the present study, qualitative analysis is therefore necessary to understand how the five users employ the features of interest discursively to index their (changing) identity through (dis)alignment with others.

The features identified and explored quantitatively in Chapter 7 were used to guide the selection of concordance lines for analysis, to maximise the productivity of the qualitative analysis of shifting linguistic identity performance. Due to time constraints, and in order to – at least theoretically – capture maximal difference across the career, only tokens of the features of interest taken from the first and tenth deciles of each user's career were analysed.

I annotated each concordance line, commenting on the content and function of each token. This analysis was primarily inductive, allowing themes to emerge from the data, but was supported by the existing literature on newcomers and expert members (§3.4). In order to maximise rigour in this inductive approach, I followed guidance from the literature on inductive content analysis, with which this analysis shares methodological similarities.

One important principle of inductive research is *stability*, “keep[ing] track of coding decisions” (Bengtsson, 2016, p. 13). This was achieved by referring back to annotations already made for other users and features to ensure that the data is coded consistently throughout; these annotations formed a codebook (Clark et al., 2021, p. 516) to refer back to.

Several researchers recommend that coding of data should be carried out on two separate occasions to ensure *reliability* (Bengtsson, 2016; Elo et al., 2014; Schreier, 2014; White & Marsh, 2006). Time constraints did not allow for a full second round of annotation. However, once the first round was completed, I replicated this effect by using the codebook of annotations from the first round to double-check the consistency of annotation.

Finally, I took steps to ensure the *confirmability* of my analysis by presenting my findings transparently “in a way that allows the reader to look for alternative interpretations” (Elo et al., 2014, p. 6; see also Gioia et al., 2013; White & Marsh, 2006) rather than taking my interpretations for granted. This involves the reproduction of language data throughout Chapter 8.

Following annotation and compilation of results, I used the most frequently occurring themes to characterise each user’s linguistic behaviour at each of the two deciles. The two deciles for each user were then compared, identifying both common themes across the career, and those which are unique to the career beginning or end. These were used to create ‘narratives’ characterising each user’s career. Finally, I compared early-career and late-career themes across all five users, identifying any similarities which may suggest that there is a typical path of behaviour change across user careers in this forum as a whole. These findings are discussed in Chapter 8.

4.5 Ethics

As in any sensitive context, this research raises a number of ethical concerns which must be considered. In this section I outline some key issues and the steps I have taken to address them. These measures were approved by the Aston Institute for Forensic Linguistics Research Ethics Committee. A copy of my ethics application form and other relevant documentation can be found in Appendix A.

Questions have been raised concerning the nature of social media data, and the consequences of this for its ethical use in research. Some researchers have highlighted the “significant textuality” (Bassett & O’Riordan, 2002, p. 234) of such data, arguing for an approach to ethics which treats social media data only as text, rather than as human behaviour per se. This is an ethically uncomplicated position, avoiding the issues of informed consent that arise from research with human subjects. However, this approach is clearly unsuitable for the present research context, which views language as a social phenomenon through which we can understand the changing identities of its producers.

Researchers who apply the human subjects model of research ethics to social media data, as I do here, have identified the classification of data as *public* or *private* as a central ethical concern. As a general consensus, *public* data are seen as clearly suitable for use by researchers, while *private* data raise additional issues (Sveningsson Elm, 2009). However, *public* and *private* are not always clear-cut concepts. For example, social media users’ expectations of privacy may not align with the actual accessibility of their data (boyd, 2011;

Lüders, 2015), especially in the case of more sensitive discussion topics (Sveningsson, 2003).

The data used for the present research can reasonably be considered public. Although the content is sensitive, including “ideologically controversial topics such as religious and political opinion” (Page et al., 2014, p. 72), the forum is accessible to anybody with an internet connection, with no registration required to read the site. The site also signposts its public status in several ways, including the existence of the ‘open fora’ where guests are invited to post. The site’s guidelines for posting also remind users that their posts are permanent, viewable and useable by individuals with varying motivations. It is therefore unlikely that users of the forum have a high expectation of privacy for their posts.

It is generally agreed that seeking informed consent is not required in the case of public data (Page et al., 2014). However, Mackenzie (2016) argues against this blanket approach, instead recommending a context-sensitive approach to the question of consent. For this research, I have chosen not to seek consent from the forum users I investigate. This is due to concerns for researcher safety. Several researchers have highlighted the potential dangers that face researchers who become known to extremist communities (Aiston, 2022; Conway, 2021; Massanari, 2018). This typically takes the form of social media abuse, but doxxing (posting individuals’ personal information online without consent) creates the possibility of offline harm. Conway (2021) therefore advises researchers against contacting extremist communities for any reason, including to seek informed consent. This sentiment is echoed in the ethical guidelines of the British Association for Applied Linguistics, which recognise that it may not always be “appropriate or safe” to seek consent when investigating “criminal interactions online” (BAAL, 2021, p. 5).

The decision not to seek informed consent makes it particularly important to protect users’ privacy. While user accounts in the forum are pseudonymous, they are not necessarily anonymous, as users may choose to incorporate a link to their ‘real-world’ identity in their forum username. Usernames are therefore not reported in this thesis; the five users studied in detail in Chapters 6 and 7 have each been assigned a numerical identifier. I have taken care, wherever possible, not to select quotes from the data which include potentially identifying information. Where this has not been possible, such information has been replaced with generic labels such as *[country]*.

When selecting quotes, I have also been conscious of the unpleasant and potentially distressing nature of this data. By reproducing the hateful language found in the dataset, I arguably give these views a new platform (Hawley, 2017). On the other hand, by selecting only the most “reasonable and erudite” quotations, (Hawley, 2017, p. 8), I risk whitewashing

the community's dangerous views. Jane (2014, 2015) takes the latter position, arguing that the selection of only mild quotes serves as "a kind of modesty-protecting 'fig leaf'" (Jane, 2015, p. 73) which allows the community's abhorrent views to fly under the radar. While Jane raises an important point about the risks of sanitising a movement's vitriol, I am wary of using genuine messages of hate directed at members of society for shock value. I am also aware of my position as a white researcher; while unpleasant to read, the typical sentiments expressed on the forum are not personally damaging to me. I must therefore be prudent about circulating the most extreme vitriol to audiences for whom this may not be the case. Therefore, while the language quoted in this thesis is frequently racist and distressing, I have avoided including the most violently hateful quotes.

I have also taken other steps to mitigate the effects of this potentially distressing data. I have had access to both group and individual psychological support, provided through the Aston Institute for Forensic Linguistics for researchers working on such data. This has allowed me to discuss strategies to cope with such data mentally and to develop resilience, as well as to practice good "work hygiene" as recommended by BAAL (2021, p. 15).

Chapter 5 How to study the career: A multiple correspondence analysis

5.1 Introduction

As I showed in §3.3.2, research into the career in social movement contexts has so far followed two main strands. The first, foundational strand has been largely theoretical, laying the groundwork for the concept of the career with broad reference to a case study context (e.g. Becker, 1973; Goffman, 1961). The second strand, which does not always explicitly work within the career framework, has detailed what can be seen as career contingencies in activist careers, factors which condition initial engagement, continued participation, and disengagement from a movement (e.g. Bunnage, 2014; McAdam, 1986; Somma, 2009; Van Dyke & Dixon, 2013). However, these papers often leave us with little understanding of what a typical career path or paths might look like in any given context, and how the researcher can practically go about identifying different career paths or stages from one another.

In order to consider how a typical career might look and how this has been established, I have, therefore, looked beyond the literature directly concerning careers and contingencies. A number of papers have been identified which investigate the different roles or stages in the careers of members of particular groups. These papers are taken from across a wide range of contexts and disciplines, using a number of different methods and reporting diverse findings. Given this variety, and the fact that each paper focuses on only one community each, it is unclear how generalisable these results might be regarding the nature of the career. This is complicated by the fact that the community that I am investigating in this thesis is of a hybrid nature. While its members can be considered to be activists for the white nationalist movement, the online-based nature of the community sets it apart from those typically examined in the social movement literature, particularly with regard to careers.

In this study, I will therefore synthesise this disparate body of research concerning career roles and stages. This will include research taken from the two disciplines of social movement studies and internet studies, as well as the few studies focusing on communities which fit both descriptions (Kleinberg et al., 2021; Lindekilde et al., 2019; McCauley & Moskalenko, 2008; Scrivens et al., 2020; Scrivens et al., 2022). I do this with a view to understanding the principles of organisation applied to communities by researchers in order to identify and categorise different career types and/or different stages of the same career. This synthesis is complicated by the wide variety of literature included in the review. The chapter therefore proceeds in two main parts, as detailed in §4.4. First, I describe the

qualitative analysis which was used to manually identify important features of the included career typologies. The identified features are described in §5.2.2. This preliminary qualitative analysis is then combined with the statistical technique of multiple correspondence analysis (Le Roux & Rouanet, 2010; see §5.3.1), for which the features identified in §5.2.2 are used as variables. The results of the multiple correspondence analysis are then interpreted.

5.2 Identifying variables: A qualitative literature review

5.2.1 Method

I began by gathering together papers in which researchers break down the membership of groups or communities into a set of roles or stages. I collected papers using a snowballing approach, following citations in each paper to find more relevant literature. I also received recommendations from my colleagues and supervisors. This approach is recommended in the case of research topics which are not easily summarised into a small set of search strings (Greenhalgh & Peacock, 2005; Sayers, 2007; Wohlin, 2014). This is the case for the present study, where the papers of interest differ on a number of important levels, including: their field; the communities or groups researched; the type of data used; the methods used to categorise the members; and the labels used in the final typologies of roles or stages. In some cases, the primary goal of the paper is not to categorise the membership of the focal community. Instead, the researchers do so as a preliminary step, and use the reported categories to investigate their stated research question. It would, therefore, be extremely challenging to identify an appropriate list of keywords which would capture a body of relevant literature. This is a direct consequence of the nature of this study, which seeks to bring together a very disparate body of literature for the first time. If there were a single clear way of identifying and describing stages or roles, or a small number of keywords which would be used to find all of the papers included, there would be very little point carrying out this work.

I included two main types of literature in the dataset: papers investigating careers in social movements or activist organisations, and those exploring online communities where interaction is only (or predominantly) virtual and linguistic in nature, to examine the online forum career. These capture the main features of my own community of interest. I also included a small number of more theoretical papers (Benne & Sheats, 1948; Lave & Wenger, 1991; Wenger, 1998) which deal with stages or roles in groups generally, so are theoretically applicable to both kinds of career. Another small set of papers (Lindekilde et al., 2019; McCauley & Moskalenko, 2008) deal explicitly with extremist activist communities, and allow for both online and offline careers. A pilot version of this study comprised 33 typologies from 31 papers. Following recommendations and a second stage of snowball searching, I added an additional 13, giving a full dataset of 46 typologies from 43 papers. Notice that

some papers contain multiple types of categorisation, which do not constitute a single system. For example, one typology I will refer to as Maloney-Krichmar and Preece (2005 [1]) distinguishes members from one another in terms of their duration and frequency of their participation in the community. In the same article, the typology I call Maloney-Krichmar and Preece (2005 [2]) captures a series of 'Clubs' of which community members may consider themselves to be part.

I read each paper in detail and made notes on various aspects, including: the methodologies used; the characteristics of members that the researchers had selected to distinguish them from other members, thus creating the typologies or categorisations; properties of the data, such as whether it concerned an online or offline community; and properties of the final typology, such as whether or not it presented a series of logically sequential stages. As noted in §4.4.1, this method was systematised as far as possible by asking the same questions, relating to the factors above, of each typology. Based on these criteria, I have sorted the typologies contained in the papers into a large number of overlapping groupings, which are described qualitatively below.

5.2.2 Findings

5.2.2.1 Literature review (4 typologies)

In some cases, researchers have conducted a literature review in order to determine their categories, giving a synthesis of roles that have appeared in other studies (Altier et al., 2022; Corrigan-Brown, 2012; Perliger et al., 2016; Preece & Shneiderman, 2009). Corrigan-Brown (2012) relies on the literature for some of her career stages (persistence, disengagement) but introduces others based on her own case studies. This example highlights the overlapping nature of the categories of typology I present here.

5.2.2.2 Case study (13 typologies)

Some studies (Ahmad, 2016; Benne & Sheats, 1948; Corrigan-Brown, 2012; Gill & Horgan, 2013; Kim, 2000; Klandermans, 1994, 2004; Lave & Wenger, 1991; Lindekilde et al., 2019; McCauley & Moskalenko, 2008; Milbrath, 1965; Perliger et al., 2016; Wenger, 1998) are based on the researchers' own generalised observations, and often contain specific case studies to demonstrate the different roles. These are often more theoretical book-length texts (e.g. Lave & Wenger, 1991; Wenger, 1998) or older studies (e.g. Benne & Sheats, 1948; Milbrath, 1965).

5.2.2.3 Social network (14 typologies)

Several studies used elements of a social network approach to consider how individuals were interactionally positioned in relation to others. Some of these studies (Ahmad, 2016;

Gleave et al., 2009; Lindekilde et al., 2019; Pedahzur & Perliger, 2006; Perliger et al., 2016; Smith Risser & Bottoms, 2014; Turner et al., 2005; Welser et al., 2011; Welser et al., 2007) explicitly employ social network analysis. Others (Klandermans & Linden, 2004; Maloney-Krichmar & Preece, 2005 [1]; Preece & Shneiderman, 2009; Singh, 2012; Waters & Gasson, 2005) do not, but do have elements of a social network approach, taking into account factors, such as the level of response members receive to an online post, or how frequently members interact with others, in order to distinguish certain membership roles from others.

5.2.2.4 Content analysis (7 typologies)

Some researchers have conducted content analyses of posts in online environments to identify different kinds of contributors (Dennen, 2008; Marcoccia, 2004; Radin, 2006; Scrivens et al., 2020 [1]; Singh, 2012; Waters & Gasson, 2005; Welser et al., 2007). In most cases content analysis was the stated methodology. This was not the case for Singh (2012), but the research questions and analysis make it clear that a kind of content analysis was performed.

5.2.2.5 Sentiment analysis (2 typologies)

Two typologies (Kleinberg et al., 2021; Scrivens et al., 2020 [2]) use sentiment analysis, an automated and more limited form of content analysis.

5.2.2.6 Conversation analysis (1 typology)

Two other linguistic methods are used to distinguish between types of members, both by only one typology each. The first of these appears in Marcoccia (2004), who uses elements of digital conversation analysis to distinguish between members who initiate sequences and topics and those who do not.

5.2.2.7 Move analysis (1 typology)

Another linguistic method is employed by Chiang et al. (2021), who use move analysis to highlight distinct patterns of posting.

5.2.2.8 Longitudinal studies (4 typologies)

A few researchers have taken a longitudinal approach, following the course of individual group members as they progress in their careers (Kleinberg et al., 2021; Scrivens et al., 2020 [2]; Scrivens et al., 2022; Singh, 2012).

5.2.2.9 Behaviour (28 typologies)

One of the largest categories of typology is those which consider the nature of participation which members engage in (Ahmad, 2016; Altier et al., 2022; Benne & Sheats, 1948; Brush et al., 2005 [2]; Chiang et al., 2021; Dennen, 2008; Gill & Horgan, 2013; Gleave et al., 2009;

Introvigne, 1999; Kim, 2000; Klandermans, 2004; Lave & Wenger, 1991; Leshed, 2005; Maloney-Krichmar & Preece, 2005 [1]; Marcoccia, 2004; McCauley & Moskaleiko, 2008; Milbrath, 1965; Perliger et al., 2016; Preece & Shneiderman, 2009; Radin, 2006; Singh, 2012; Smith Risser & Bottoms, 2014; Turner et al., 2005; Viégas & Smith, 2004; Waters & Gasson, 2005; Welser et al., 2011; Welser et al., 2007; Wenger, 1998). Naturally, as these typologies describe a range of online and offline groups, this category covers many different types of behaviour, including but not limited to: function of online posting; initiation vs. response behaviours; lurking vs. posting behaviours; and level of effort or personal risk involved in activities undertaken.

5.2.2.10 Group loyalty (14 typologies)

Other studies considered members' loyalty to the group (Brush et al., 2005 [1]; Corrigan-Brown, 2012; Kim, 2000; Klandermans, 1994, 1997, 2004; Klandermans & Linden, 2004; Kleinberg et al., 2021; Maloney-Krichmar & Preece, 2005 [1]; Perliger et al., 2016; Saunders et al., 2012; Scrivens et al., 2020 [2]; Scrivens et al., 2022; Smith Risser & Bottoms, 2014). The choice of the word 'loyalty' is inspired by Hirschman (1970), who contrasts *loyalty* and *exit* to refer to the individual's choice to remain in or leave the group. Loyalty here has two main dimensions. The first can be described as *duration*, whereby members are categorised according to when they began participating in the group, relative to others. Duration is to some extent quantifiable – members are 'newer' or 'older' than one another. Example typologies include Smith Risser and Bottoms (2014), Brush et al. (2005 [1]), and Kim (2000). The second can be described as *persistence*, whereby members are distinguished based on their decision to, on the one hand, persist or remain as a member of the group, or on the other, to take some other path and leave the group. Here, members are distinguished according to whether they are still part of their (original) group or not; example typologies include Corrigan-Brown (2012) and Klandermans (1997). The two labels pick out two shades of the same meaning, both referring to the individual's continued commitment to a group. Indeed, Saunders et al. (2012) is an example of a typology which seems to span both labels. Here, individuals are distinguished based on their choice to continue engaging or to disengage, rather than on the duration of their commitment in months or years. Nonetheless, the threshold between roles is in part determined by durational milestones.

5.2.2.11 Frequency (11 typologies)

Another category captures those typologies which take into account the frequency or intensity with which members participate in the community or its associated activities (Brush et al., 2005 [1]; Kleinberg et al., 2021; Leshed, 2005; Maloney-Krichmar & Preece, 2005 [1]; Marcoccia, 2004; Saunders et al., 2012; Scrivens et al., 2020 [2]; Scrivens et al., 2022; Turner et al., 2005; Viégas & Smith, 2004; Welser et al., 2007). This is related to the *group*

loyalty category, as it concerns how much time an individual dedicates or has dedicated to a group. The two are recognised as distinct categories, however, as several typologies explicitly take each as separate considerations. To again use Saunders et al. (2012) as an example, the roles of *returners* and *repeaters* are distinguished only by the level of persistence over time (protests attended over the lifetime); meanwhile, *repeaters* and *stalwarts* are distinguished by the frequency of protest in the last year alone.

5.2.2.12 Expertise (6 typologies)

'Expertise' is often used as a measure with which to distinguish members from one another. Expertise is necessarily vaguely defined here. The nature of expertise varies across communities, as Lave and Wenger (1991) and Wenger (1998) make clear. McAuley and Leskovec (2013) describe expertise only as "some quality that users gain over time" (p. 897) and, for the purposes of their study, as "an *interpretation* of our model's latent parameters" (p. 902; emphasis in original). In other studies, expertise is the result of having attended particular training workshops, courses, or camps (Ahmad, 2016; Introvigne, 1999; Perlinger et al., 2016).

5.2.2.13 Trust (1 typology)

Radin (2006) distinguishes between stages in a member's career based on the level of trust in the community that the member displays. She identifies levels of trust based on relevant literature; this is beyond the scope of this work.

5.2.2.14 Self-reported data (3 typologies)

From a methodological standpoint, certain studies base their role typologies not on observations of participants' behaviour, but on participants' own self-reporting of their involvement in the community through surveys (Brush et al., 2005 [2]; Leshed, 2005; Saunders et al., 2012).

5.2.2.15 Community labels (4 typologies)

For the labels given to particular roles, some researchers rely, at least in part, on labels used by the community to refer to its own members (Gleave et al., 2009; Golder & Donath, 2004; Maloney-Krichmar & Preece, 2005 [2]; Welser et al., 2011). These labels may be used by members to self-identify as members of social cliques (Maloney-Krichmar & Preece, 2005 [2]) or functional roles (Gleave et al., 2009; Welser et al., 2011), or alternatively to describe other members, usually in negative terms regarding their behaviour (e.g. *flamer* or *troll* in Golder and Donath (2004)).

5.2.2.16 User profiles (1 typology)

In one study (Vaast, 2007), the researcher uses the user profiles produced by members of a social media site to distinguish between different types of member.

5.2.2.17 Sequentiality (14 sequential typologies; 22 non-sequential; 8 with elements of both; 2 with elements of neither)

The typologies examined here differ from one another with respect to their sequentiality. Some typologies (Ahmad, 2016; Brush et al., 2005 [1]; Corrigall-Brown, 2012; Kim, 2000; Klandermans, 1994, 1997; Lave & Wenger, 1991; Maloney-Krichmar & Preece, 2005 [1]; McAuley & Leskovec, 2013; McCauley & Moskalkenko, 2008; Preece & Shneiderman, 2009; Radin, 2006; Saunders et al., 2012; Singh, 2012) are *sequential*. The roles they describe can be understood as *stages* in a process; members must pass through 'earlier' roles or stages in order to attain subsequent ones, although there is no expectation that every member will continue to progress throughout the entire set of stages to the 'later' ones (e.g. Preece & Shneiderman, 2009; Singh, 2012). Despite the temporal implications of the words *sequential*, *earlier*, *later*, and *subsequent* I have used here, not all of the typologies included in this list are based on the length of time members have spent in the community. Some sequential typologies (e.g. Lave & Wenger, 1991; McCauley & Moskalkenko, 2008; Preece & Shneiderman, 2009; Radin, 2006; Singh, 2012) are, in fact, based on behavioural, rather than temporal, stages, where new behaviours have to be learned before the member can progress to the next role. These recall the social constructionist nature of identity performance described above (§2.3), whereby people acquire new community-specific performance resources over time.

Additionally, not all of the typologies which are based on temporal aspects are sequential. For instance, Scrivens et al. (2022) distinguish between *early desister* trajectories (those whose forum posting career is very short, disappearing shortly after their first post) and *chronic poster* trajectories (who join and continue to post persistently over a long time). While the distinction between the two is a matter of persistence or loyalty to the forum, there is no element of sequentiality, as members cannot move from one role to the next; if a given member account is an *early desister*, they cannot then become a *chronic poster*, as the former precludes the latter (and vice versa). This typology is, therefore, among those categorised here as *non-sequential* (alongside Benne & Sheats, 1948; Brush et al., 2005 [2]; Chiang et al., 2021; Cortese, 2015; Gill & Horgan, 2013; Klandermans, 2004; Klandermans & Linden, 2004; Kleinberg et al., 2021; Leshed, 2005; Lindekilde et al., 2019; Maloney-Krichmar & Preece, 2005 [2]; Marcoccia, 2004; Milbrath, 1965; Pedahzur & Perliger, 2006; Scrivens et al., 2020 [1] and [2]; Scrivens et al., 2022; Turner et al., 2005; Vaast, 2007; Valocchi, 2012; Waters & Gasson, 2005; Welser et al., 2007). These typologies do not

suggest that a member must 'develop' (over time or through progressively learned behaviours) to reach the 'next level' of membership, or that any particular role must precede any other. However, this does not imply that members may not have different roles at different points in their career (Welser et al., 2007) or even hold multiple roles simultaneously or in short succession, varying from forum post to forum post as they engage in behaviours associated with more than one role (Maloney-Krichmar & Preece, 2005 [2]; Waters & Gasson, 2005).

A few papers (Altier et al., 2022; Dennen, 2008; Golder & Donath, 2004; Introvigne, 1999; Smith Risser & Bottoms, 2014; Viégas & Smith, 2004; Welser et al., 2011; Wenger, 1998) propose typologies with both sequential and non-sequential elements. An illustrative example is Golder and Donath (2004), who show that a single poster can move from a *lurker* to a *newbie* to a *celebrity*, but outside of this typical career path posters may also drop in as *flamers*, *trolls*, or *ranters*. Two papers (Gleave et al., 2009; Perliger et al., 2016) do not propose typologies per se, but instead describe a range of factors which may contribute to typologies for particular groups. These are, therefore, neither sequential nor non-sequential.

5.2.2.18 *Continua (8 typologies)*

Some typologies (Klandermans, 2004; Leshed, 2005; Marcoccia, 2004; McAuley & Leskovec, 2013; McCauley & Moskalenko, 2008; Pedahzur & Perliger, 2006; Scrivens et al., 2020 [2]; Scrivens et al., 2022) are defined (in whole or in part) in terms of continua. They do not draw the (generally somewhat arbitrary) dividing lines between discrete categories of membership as in other typologies, but instead individual members can be assigned a position on each dimension and thus have greater or lesser similarity with other members.

5.2.2.19 *Mode (26 online typologies; 16 offline typologies; 4 typologies which may be either)*

A key distinction to be made between different types of typologies included in this study is whether they describe *online* or *offline* communities. As highlighted above (§5.2.1), two main bodies of literature have been brought together in this review. One concerns social movements and activist communities, including those of an extremist nature, which have typically (although not always) focused on offline contexts. The other focuses on online communities of various kinds, including, in some cases, activist or extremist communities. The community of interest here, users of an online white nationalist forum, straddles both types. The categories of *online* communities (Brush et al., 2005 [1] and [2]; Chiang et al., 2021; Dennen, 2008; Gleave et al., 2009; Golder & Donath, 2004; Kim, 2000; Kleinberg et al., 2021; Leshed, 2005; Maloney-Krichmar & Preece, 2005 [1] and [2]; Marcoccia, 2004; McAuley & Leskovec, 2013; Preece & Shneiderman, 2009; Radin, 2006; Scrivens et al., 2020 [1] and [2]; Scrivens et al., 2022; Turner et al., 2005; Vaast, 2007; Viégas & Smith,

2004; Waters & Gasson, 2005; Welser et al., 2011; Welser et al., 2007) and *offline* ones (Ahmad, 2016; Altier et al., 2022; Benne & Sheats, 1948; Corrigall-Brown, 2012; Cortese, 2015; Gill & Horgan, 2013; Introvigne, 1999; Klandermans, 1994, 1997, 2004; Klandermans & Linden, 2004; Milbrath, 1965; Pedahzur & Perliger, 2006; Perliger et al., 2016; Saunders et al., 2012; Valocchi, 2012) are for the most part mutually exclusive. However, a handful of typologies can be placed in both categories. Some papers (Lave & Wenger, 1991; Wenger, 1998) are more theoretical in nature, and do not focus their descriptions on any individual communities. Although specific case studies are offered, it is clear that these papers aim to develop a framework which can be applied to all manner of communities and contexts, including online (Wenger, 1998, pp. 6-7). Others focus more narrowly on (extremist) activism, but either do not focus on specific communities in the same vein as the other theoretical papers (McCauley & Moskalenko, 2008), or explicitly refer to careers realised both online and offline (Lindekilde et al., 2019).

5.2.2.20 Content analysis of interviews (4 typologies)

There is a second category of typologies which, like those outlined in §5.2.2.4, also employ content analysis to identify differences between membership types. However, rather than analysing posts made by group members when participating in the community, these typologies (Altier et al., 2022; Cortese, 2015; Klandermans & Linden, 2004; Valocchi, 2012) involve content analysis of interviews conducted with individuals speaking about their own activist identities. This is a more meta-level analysis, considering individuals' own self-perception rather than their actual behaviour or language use. The category is distinguished from the self-reported data and community labels categories (§5.2.2.14 and §5.2.2.15 respectively), as the categorisations are based on the researchers' interpretations of participants' responses rather than the participants' own words.

5.2.2.21 Leadership-assigned roles (3 typologies)

This is another category which is in some senses similar to that of *community labels*, in that the categories are defined by members of the group and merely reported by the researchers (Altier et al., 2022; Gill & Horgan, 2013; Introvigne, 1999). However, in this second category, roles, if not necessarily the labels used to describe them, are assigned to members by group leaders, sometimes against the wishes of the individual member (e.g. Altier et al., 2022). The nature of the typologies included in this category implies that, as we might expect, this phenomenon is more likely to be associated with offline, activism-focused communities.

5.2.2.22 *Level of financial contribution (1 typology)*

In one case (Introvigne, 1999), members are officially distinguished from one another into membership levels depending on the level of financial contribution that the individual has made to the organisation.

5.2.2.23 *Personality traits (1 typology)*

Two typologies (Klandermans & Linden, 2004; Lindekilde et al., 2019) suggest that an individual's political socialisation and other elements of their life prior to entry to the group can also affect that individual's career trajectory within the group.

5.2.2.24 *Level of radicalisation (1 typology)*

McCauley and Moskalenko's (2008) 'pyramid model' proposes that adherents of social movements are distinguished from one another along a continuous axis of level of radicalisation. Those who broadly agree with a certain political position occupy the bottom of the pyramid, while those few who have been radicalised into extremist support for the movement are at the top.

5.2.2.25 *Group-based trajectory modelling (3 typologies)*

Finally, a selection of typologies (Scrivens et al., 2020 [1] and [2]; Scrivens et al., 2022) employ a method called Group-Based Trajectory Modelling (GBTM). GBTM was developed by Nagin and Land (1993) to analyse criminal careers but has been expanded to careers in other domains, such as online forum careers, as demonstrated here. The method identifies individuals with similar trajectories and groups them together, revealing distinct trajectory types which distinguish between member types. The typologies included here shows that forum members' careers differ in terms of their initial volume of posting, whether this volume rises or falls over time, and how long an individual continues to post for before leaving the community.

5.3 Synthesising key dimensions of variation: A multiple correspondence analysis

5.3.1 Method

Multiple correspondence analysis (MCA) is a dimension reduction technique, similar to principal component analysis but suitable for categorical or qualitative data (Le Roux & Rouanet, 2010). MCA identifies patterns of co-occurring features in a dataset and groups them together, reducing the number of *dimensions* or *axes* of variation and thus rendering that variation more readily interpretable. I used the *soc.ca* package for R (Larsen et al., 2016) to conduct the analysis.

Like all methods of dimension reduction, MCA is predicated on the idea that where two or more variables have a high level of covariance (i.e. are similarly distributed across the dataset, or have a reasonably consistent relationship to one another across the dataset), then those variables can be grouped together as a single complex variable with no considerable data loss. Figures 2 and 3 below illustrate this, showing how two-dimensional data can be reduced to a single dimension while retaining much of the information provided by two dimensions. By combining those variables which pattern similarly across individuals, and which, therefore, do not help to distinguish one individual from another, the main points of difference between individuals can be understood more easily, without the noise associated with considering a large number of variables at once.

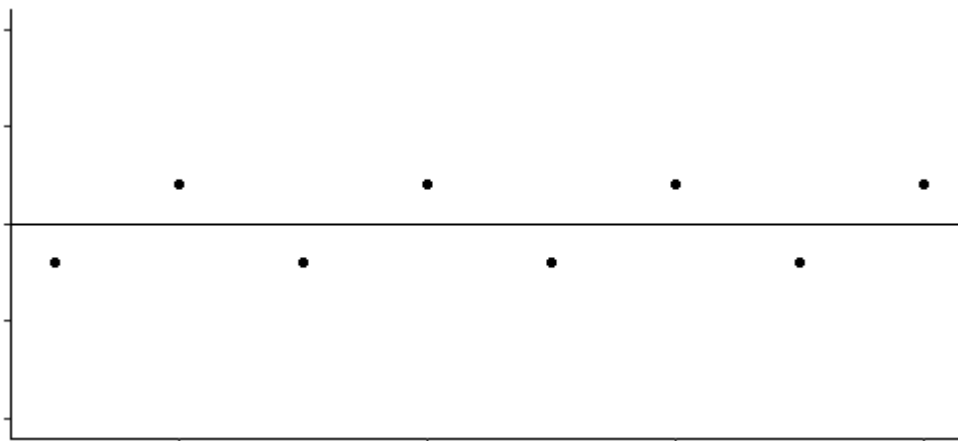


Figure 2: Example of two-dimensional data

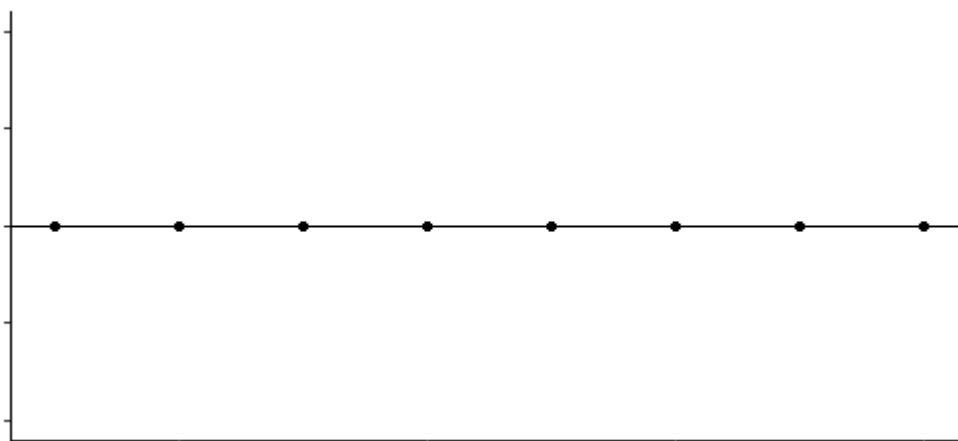


Figure 3: Example of one-dimensional data, illustrating how a single dimension can be derived from two dimensions without notable data loss

MCA identifies sets of features which tend towards mutual inclusivity or exclusivity of one another. These features are values of variables: for example, a binary *variable* might have the *values* 1/0 or Y/N. For the purposes of MCA, these features or values are referred to as

categories or *modalities*. MCA identifies a set of modalities which tend to co-occur together, and contrasts these with another set of co-occurring modalities. Crucially, it identifies modalities which frequently co-occur *within*, but not *across*, these sets. The distinction between two sets – say, Set A and Set B – creates a *dimension*, with Set A at the positive end of the dimension and Set B at the negative end.² Individuals may fall anywhere along the dimension; if they have all of the modalities from Set A, and none from Set B, the individual will be located at the positive end of the dimension. Individuals with a mixture of features will be closer to the middle of the dimension. MCA, therefore, allows us to identify which individuals are more similar to and more different from one another, and importantly to *explain* the dimension(s) along which they differ in the simpler terms of ‘A’ and ‘B’ rather than in more complex, multivariate terms. The dimensions identified through MCA can therefore be understood as the key ways in which individuals vary from one another in the dataset, or the key principles that can be used to organise these individuals. To illustrate this more intuitively, Table 1 demonstrates how bundles of individual, frequently co-occurring features (*modalities*) can be interpreted together and understood as elements of a single complex variable.

	Modalities associated with dimension		Interpretation of modalities		Interpretation of dimension
Positive dimension	Shoe size over 10	→	‘Tall’	→	HEIGHT
	Arm span over 1.8m				
	Inseam over 34”				
Negative dimension	Shoe size under 3	→	‘Short’	→	
	Arm span under 1.5m				
	Inseam under 27”				

Table 1: Illustration of multiple correspondence analysis using the intuitive dimension of height

The qualities of having a larger shoe size, a wider arm span, and a longer inseam measurement can be understood to contribute to a single complex quality, that of being tall. In other words, multiple modalities can be interpreted together as having a single, complex meaning. This complex variable can then be contrasted with another, opposing variable, such as being short. The dimension can be interpreted as the continuum of difference between these two extremes: here, as *height*, with tall individuals found on one end of the dimension, short on the other, and average-height individuals (with qualities of both ends of the dimension) in the middle.

For this study, the dataset comprises 46 individuals, the 46 typologies identified in the membership categorisation literature. I manually marked each typology for the presence or

² The terms *positive* and *negative* are used in the geometric sense and do not imply a value judgement.

absence of each of the binary features identified in the qualitative analysis above. Two variables were not marked in this binary fashion; the *sequentiality* variable also includes a modality to account for those typologies with both sequential and non-sequential elements, while the *mode* variable includes not only *online* and *offline* but also *both*, which identifies the more theoretical typologies which may apply to either kind of community.

Not every variable identified in the qualitative analysis was included in the dataset for the MCA, however. The purpose of MCA is to identify key points of difference in the dataset. If every individual displays the same value for a single variable, this variable is of no use in distinguishing (groups of) individuals from one another, and, thus, in identifying dimensions of variation. For this reason, MCA emphasises rarer modalities, which make greater contributions to the total variance across the dataset. However, this also means that very rare modalities are “overly influential for the determination of axes” (Le Roux & Rouanet, 2010, p. 61) and should be excluded from the analysis to avoid this undue influence. Le Roux and Rouanet (2010) define ‘very rare’ modalities somewhat flexibly, as being “say, of frequencies less than 5%” (p. 39). Compared to the typical applications of MCA in the analysis of larger-scale data from surveys and other questionnaires, the study described here uses a very small dataset of only 46 individuals. For this reason, the threshold for exclusion has been set at 10%, meaning that modalities representing four or fewer individuals are excluded. Although these modalities may represent a significant *percentage* of the individuals, they do not represent a significant *number* of individuals, and so ought not to be overrepresented in the results. Recall that MCA works by grouping together modalities which tend to co-occur – it would be inaccurate to make such claims about modalities which only occur a small number of times, as it is unclear whether this ‘tendency’ would carry across a (theoretical) larger dataset.

Variables and individual modalities were excluded from the analysis in a variety of ways. In most cases, variables were excluded because four or fewer individuals were marked for the *presence* of the variable. For binary variables, this logically implies that the *absence* modality meets the numerical threshold for inclusion; however, the *absence* modality was not retained for the analysis. All of the variables in the full dataset were defined on the basis of the qualitative analysis I had conducted previously; variables were identified in the first place because they were present in at least one typology. The *absence* modalities for each variable are, therefore, defined in relation to the *presence* modalities, but not vice versa. Therefore, *absence* modalities were not retained where their *present* counterparts were excluded.

In the case of the *sentiment analysis* variable, I followed the recommendation of Le Roux and Rouanet (2010) to pool similar variables or modalities together. Sentiment analysis has been described as a form of automated content analysis (e.g. Hasbullah et al., 2016), able to identify elements of content rapidly on larger-scale datasets. I therefore grouped together the two variables of *sentiment analysis* and *content analysis* under the latter label, marking typologies for the presence of one or the other, or for the absence of either.

Finally, two modalities were coded as *passive modalities*. This means that a single modality of a variable was excluded from the analysis, while retaining the rest of the variable's modalities. This included a modality of 'missing' values within the *sequentiality* variable (those typologies described as neither sequential nor non-sequential) and the *both* modality within the *mode* variable, referring to typologies which may apply to online or offline communities, of which there are only four. The exclusion of infrequent modalities in this way means that the analysis presented here is an example of *specific MCA* (Le Roux & Rouanet, 2010, p. 61). For illustrative purposes, a sample of the dataset, showing a limited number of typologies and variables, is shown in Table 2. The full dataset used for the MCA can be found in Appendix B. Appendix C contains the R script used to run the analysis.

Typology	Typology number	Expertise	Continua	Sequentiality	Mode
Ahmad, 2016	1	yes	no	yes	offline
Altier et al., 2022	2	no	no	both	offline
Benne and Sheats, 1948	3	no	no	no	offline
Brush et al., 2005 [1]	4	no	no	yes	online
Brush et al., 2005 [2]	5	no	no	no	online

Table 2: Sample of the dataset used for MCA. The full dataset can be found in Appendix B.

5.3.2 Findings

Table 3 below shows the five dimensions of variance identified by the MCA using the *soc.ca* package. We can see that the first two dimensions together account for a very large percentage (over 95%) of the variance across the entire dataset. Typically, the number of dimensions interpreted is the number required to account for 80% or more of the variance (Bühlmann et al., 2017; Gautier Morin & Rossier, 2021; Le Roux & Rouanet, 2010).

Therefore, I will interpret only the first two dimensions.

Dimension	Dim1	Dim2	Dim3	Dim4	Dim5
Eigenvalue	0.24	0.18	0.12	0.11	0.10
Variance (modified rate)	68.4	26.8	2.9	1.4	0.5
Cumulative variance %	68.4%	95.2%	98.1%	99.5%	100%

Table 3: Eigenvalues and variance for the five dimensions identified by MCA

Figure 4 shows the cloud of modalities arranged across these two dimensions. The coordinates of each modality on the two axes correspond to the extent to which it is associated with each dimension. The further from zero a modality's coordinate is on a given axis, the more strongly associated that modality is with the corresponding dimension.

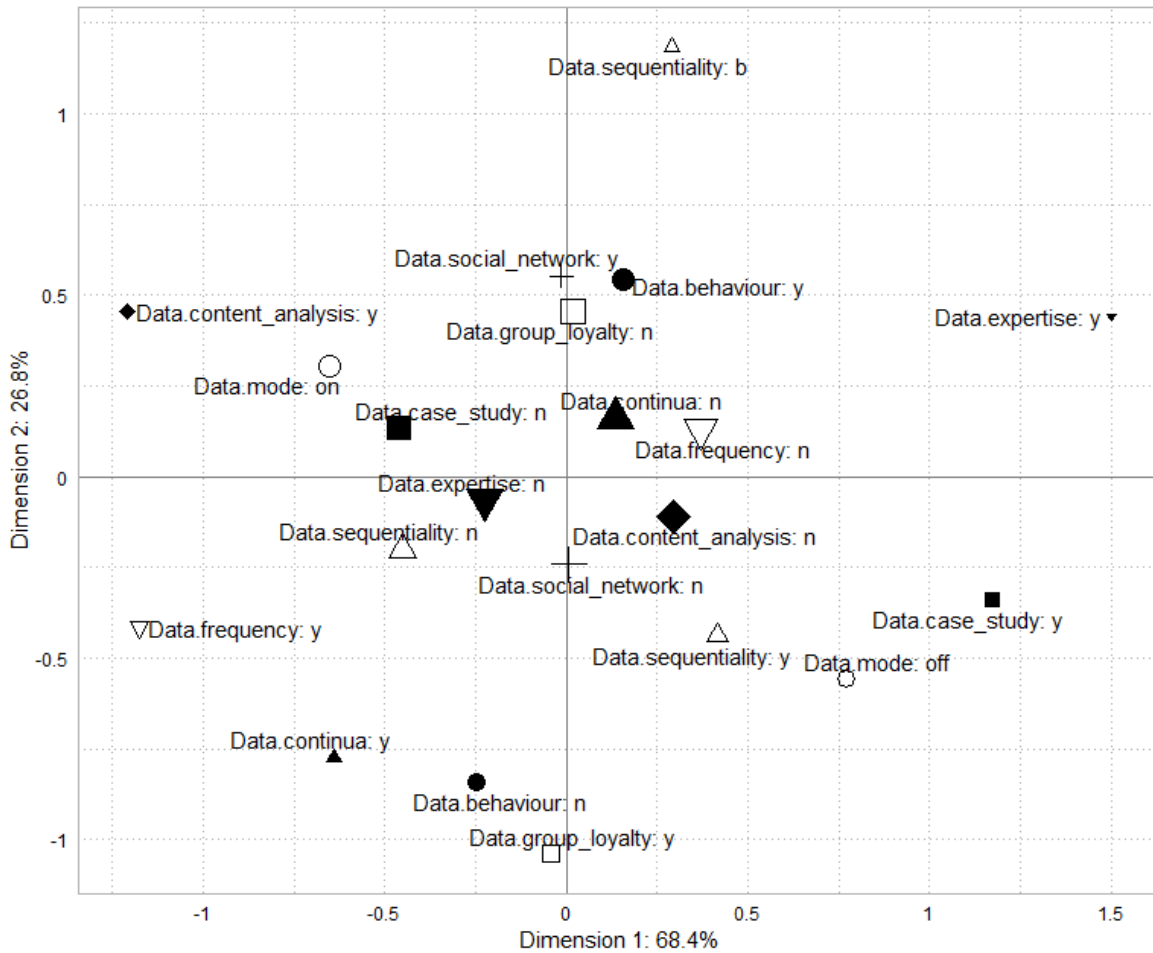


Figure 4: Cloud of modalities for Dimensions 1 and 2

The modalities which make above-average contributions to the two dimensions are detailed in Table 4 below. The average contribution is 4.35, as there are 23 active modalities ($100/23 = 4.35$). Only those modalities with an above-average contribution to the dimension, either on the positive or the negative side, are included in the interpretation of that dimension “as these represent the most distinguishing patterns of variation” (Clarke, 2019, p. 67) in the dataset.

Dimension 1					
Positive side			Negative side		
Modality	Contr	Coord	Modality	Contr	Coord
Case study: Y	16.1	1.17	Frequency: Y	13.7	-1.18
Expertise: Y	12.2	1.50	Content analysis: Y	11.9	-1.21
Mode: offline	8.6	0.77	Mode: online	10.0	-0.65
			Case study: N	6.23	-0.46
Dimension 2					
Positive side			Negative side		
Modality	Contr	Coord	Modality	Contr	Coord
Sequentiality: both	13.2	1.18	Group loyalty: Y	17.8	-1.04
Behaviour: Y	9.7	0.54	Behaviour: N	15.0	-0.84
Group loyalty: N	7.8	0.46	Mode: offline	5.9	-0.56
Social network: Y	5.0	0.55	Continua: Y	5.6	-0.77

Table 4: Modalities with above-average contribution to each dimension

5.3.2.1 Interpretation of Dimension 1

The first dimension distinguishes between typologies which describe *offline* communities, and those which describe *online* communities. The clearest indication of this interpretation is that the two modalities of the *mode* variable, *offline* and *online*, are present on opposing sides of the dimension. This finding is perhaps unsurprising, as these labels characterise the two main bodies of literature included in this dataset, social movement and internet studies. However, these results are nonetheless interesting from the point of view of systematically understanding these literatures, offering an insight into the approaches that have been taken – and have *not* been taken – to membership categorisation in the two contexts.

5.3.2.1.1 Positive side

As Table 4 shows, the positive side of the first dimension is associated with the modality *offline* of the variable *mode*. The influence of this on my interpretation of this dimension is clear. While showing lower contributions than other modalities, it is notable that complementary modalities of this variable are present on both sides of the dimension, showing that this variable is an important one to the interpretation of the dimension. The other modalities which contribute to the dimension also support this interpretation.

Another modality associated with the positive side of the dimension, which has its counterpart on the negative side, is the presence of a *case study* method. The case study method as I have defined it here (§5.2.2.2), is based on general and holistic observation of the focal community. It is an inductive method, with the researcher's observations guiding them in determining how to identify emergent categories of membership. I contend that this method is likely more suited to offline research contexts than online ones. Online, the ways in which participants can engage with the group, and what researchers can observe about the group and its members, will always be constrained by the affordances of the platform. These affordances (such as users' posting or replying behaviours; design of user profile; and

time- and date-stamping of user activity) provide a clear framework for the researcher's analysis, an obvious starting point guiding their identification of member or career types. Therefore, researchers working in online environments are more likely to be constrained and consider similar elements when designing their typologies. Researchers of offline communities are much less likely to be constrained in this way, and may, therefore, take the more holistic, inductive approaches I have categorised as *case studies*.

The final modality on the positive side of Dimension 1 is the consideration of *expertise*. The appearance of expertise on this side of the dimension may be explained with reference to the ways in which expertise is determined in the typologies included in this study. As I describe in §5.2.2.12, the definition of expertise I use here is somewhat loose and vague, best summarised by McAuley and Leskovec's (2013, p. 897) description of it as "some quality that users gain over time" and with practice. Papers in this study which offer specific accounts of how expertise is gained describe members' participation in training workshops, courses and camps (Ahmad, 2016; Introvigne, 1999; Perliger et al., 2016). These methods, particularly training camps, are more likely to take place in communities which are at least partially offline. This may account, in part, for the association of expertise with offline careers. More generally, the community-specific nature of expertise may also lend itself to a more holistic approach, as for the case study typologies. This is not to say that expertise could not be determined in ways that are suited to online studies, simply that the constraints associated with online studies may favour alternative approaches.

5.3.2.1.2 Negative side

Complementary to the *offline* modality appearing on the positive side of Dimension 1, the *online* modality appears here. Likewise, while the modality indicating the presence of the *case study* variable appears on the positive side of the dimension, the absence of this method appears on the negative side. On this side, the absence of the *case study* method contrasts with the association of two other methods, the use of *content analysis* and the consideration of *frequency*, with the negative end of the dimension. These methods arguably result from the online environment, which, as I argued above, constrains the nature of members' engagement with the community and, consequently, the types of engagement researchers are able to take into account.

The largely text-based nature of interaction afforded by online platforms means that studies focusing on users' written posts are common in research on online communities. In the context of these membership categorisation studies, this is often done through content or sentiment analysis (recall that these were pooled together for the purposes of the MCA; see §5.3.1). In the present dataset, every typology which used content analysis concerned an

online community. While content analysis of transcribed in-person speech is also possible, this method is doubtless more suited to online contexts where entire posts can be easily scraped and analysed. Because online platforms often only afford textual interaction, the researcher's choice to focus on text does not represent an inductive research design, contrasting with the case study methods found on the other end of Dimension 1.³

The final modality associated with the negative side of Dimension 1 is the consideration of *frequency*. Again, this represents a contrast from the less guided, more emergent approaches associated with the offline studies of the positive end of the dimension. Studying frequency of participation lends itself to online studies, where frequency across the entire career is much easier to determine through time- and date-stamps. Offline, fully accurate data about frequency of participation would require long periods of careful observation, and could only be determined for the duration of the observation. Saunders et al.'s (2012) approach asks activists to self-report the frequency of their own participation, although self-reporting may not generate accurate results.

5.3.2.2 Interpretation of Dimension 2

The second dimension contrasts typologies which focus on *behaviour*, the nature of members' participation in the community, with typologies focusing on members' *loyalty*, the amount of time they have spent engaging with the group. As for the first dimension, this interpretation is primarily supported by the fact that complementary modalities of the variables *behaviour* and *group loyalty* appear respectively on each end of the dimension.

5.3.2.2.1 Positive side

The complementary nature of the *group loyalty* and *behaviour* variables makes this a very robust distinction on this dimension. On the positive side, we find an association with typologies focusing on *behaviour*, and typologies which do *not* focus on *group loyalty*.

Although this is the most important contrast overall, the greatest contribution to this side of the dimension is actually typologies marked as having both sequential and non-sequential elements (i.e. the *both* modality of the *sequentiality* variable). This is something of a puzzle. This is the only modality of this variable which makes an above-average contribution on either side of Dimension 2 (Table 4). The absence of the *sequential* and *non-sequential* modalities is understandable in light of the discussions of the *behaviour* and *group loyalty* variables above (§5.2.2.9 and §5.2.2.10, respectively). The *group loyalty* variable encompasses both *duration*, which is more quantifiable and sequential (i.e. an individual must logically be a new user before they can be a long-term user), and *persistence*, which is

³ Although content analysis may involve emergent, rather than a priori, coding methods.

retrospective and non-sequential (the individual may *either* have chosen to remain as part of their community, *or* to disengage, but these are understood as discrete options rather than a sequence). Meanwhile, *behaviour* may also be non-sequential (i.e. different behaviours apparently chosen due to personal preference, as in Altier et al., 2022; Golder & Donath, 2004; Turner et al., 2005; Waters & Gasson, 2005) or sequential (as new community-specific behavioural resources are acquired over time; see Grant & MacLeod, 2020; Lave & Wenger, 1991; Wenger, 1998). It is not clear, however, why typologies with *both* characteristics might be associated so strongly with *behaviour* and disassociated with *group loyalty*. Nevertheless, these findings reflect that either behavioural or loyalty-type typologies may be either sequential or not, which is consistent with the theoretical framework I have laid out here. Empirically, but perhaps not necessarily, behavioural typologies are more likely to have elements of both.

The final modality contributing to the interpretation of this side of the dimension is the presence of social network analysis. This is consistent with the interpretation of this dimension as being largely defined by the consideration of behaviour. As highlighted above (§5.2.2.9), the category of 'behaviour' is quite broad for the purposes of this study, with typologies considering many kinds of behaviour grouped together for the purposes of creating a suitable MCA variable. Many of the types of behaviour identified by researchers as relevant for categorising members concern the nature and extent of individuals' interactions with other members (this is particularly the case in typologies concerning online communities). It is, therefore, unsurprising that some of these typologies would also concern themselves with interaction more generally, explicitly mapping out interactions and interactants using a social network analysis.

5.3.2.2.2 Negative side

On the negative side of Dimension 2 we find the opposite *behaviour* and *group loyalty* modalities to those found on the positive side. This side is therefore associated with typologies taking into account *group loyalty*, and *not* taking into account *behaviour*. Unlike on the positive side, the contributions of these modalities on this side of the dimension are very high, both among the highest contributions to either of the dimensions reported here (see Table 4). This makes it even clearer that these two variables are the most important to the interpretation of this dimension.

The remaining two modalities contributing to this side of the dimension, the *offline* mode and the use of *continua* rather than discrete career stages, are problematic for the interpretation of the dimension, as they do not clearly support, nor contradict, the interpretation I have made. Indeed, it is somewhat unclear mathematically why these modalities appear at this

side of the dimension. Both co-occur with the *behaviour* and *group loyalty* modalities on this side of the dimension at around the same rates as they co-occur with the corresponding modalities at the opposite side, as Table 5 shows.

Rate of co-occurrence between...		
Mode: offline	Group loyalty: Y	7 co-occurrences
	Group loyalty: N	9 co-occurrences
	Behaviour: N	8 co-occurrences
	Behaviour: Y	8 co-occurrences
Continua: Y	Group loyalty: Y	3 co-occurrences
	Group loyalty: N	5 co-occurrences
	Behaviour: N	4 co-occurrences
	Behaviour: Y	4 co-occurrences

Table 5: Rate of co-occurrence between Mode: offline and Continua: Y with other key modalities in Dimension 2

Looking at the other modalities associated with the positive side of the first dimension suggests that these two sets of modalities may ‘repel’ one another. The modality *continua: y* is the only modality which never co-occurs with *sequentiality: both*, and co-occurs only once with *social network: Y* (and 7 times with *social network: N*). Meanwhile, *mode: offline* co-occurs only twice with *sequentiality: both* and four times with *social network: Y* (12 times with *social network: N*).

While this offers a statistical explanation for the appearance of *continua: Y* and *mode: offline* on the negative side of the dimension and *sequentiality: both* and *social network: Y* on the positive, it remains unclear how these might contribute to an interpretation of the meaning of Dimension 2. However, *continua: Y* and *mode: offline* represent some of the lowest contributions to either of the two dimensions interpreted here, and may therefore legitimately be given less weight in the interpretation of the dimension, especially compared with the weight of evidence behind the behaviour vs. loyalty interpretation.

5.4 Discussion

To summarise, by using MCA to systematise the interdisciplinary literature on career roles and stages, I have identified two important dimensions of variation, which can be understood as the key organising principles underpinning this literature. Dimension 1 captures the distinction between typologies focusing on online and offline communities, respectively. This result is somewhat unsurprising, reflecting the two main types of literature in the study, namely social movement studies and internet studies. Dimension 2 primarily distinguishes between *behaviour* and *group loyalty* as the two principal measures used to identify different categories of member within a group. As I have outlined in §5.2.2.9, *behaviour* is construed broadly here, reflecting the highly disparate nature of the communities studied in this sample of typologies. The types of behaviour individuals engage in differ from community to

community. Most obviously, the kinds of activities members of online communities can take part in differs from those in offline communities. However, more specific properties, such as the goals of the community, are also relevant in determining the nature of the available behavioural resources.

The discussion above is clear that the interpretations of the dimensions I have offered are not always a perfect fit for the quantitative results produced by the MCA. This is particularly the case for Dimension 2. Nonetheless, I am satisfied that these are the best possible interpretations of the MCA results. For both dimensions, I have given greater weight to those variables of which complementary modalities appear at the two sides of the dimension, as these are clearly key variables to the calculation of the dimension and therefore to its interpretation. Fortunately, in the more contentious Dimension 2, these paired modalities account for some of the larger dimension contributions across the results. This further legitimises my decision to give more weight to these variables in my interpretation. More generally, like any statistical method, MCA is something of a black box. The results it gives are based only on co-occurrence of modalities and are not designed or manipulated in any way to provide answers that are neatly interpretable by researchers. The job of the researcher is, therefore, not to report a perfect explanation, but to offer the best possible interpretation for as many of the numerical results as possible. In addition, MCA is typically carried out on larger datasets, comprising more individuals than the present one. In this smaller dataset, it is feasible that some of the modalities co-occur simply by coincidence, with no theoretical explanation *per se*.

The size of the dataset relative to those used for other MCA studies may be seen as a limitation for this study, impacting the generalisability of these results to the study of careers across contexts. However, the study is not intended to be exhaustive. As mentioned above, the disparate nature of the literature on which it draws means that it was difficult to systematise the identification of typologies, and it is probable that similar works exist which do not appear in the present study. Nevertheless, I have identified some salient properties of studies of careers from a range of relevant papers, which I can use to guide my own study of the career in a white nationalist forum, and which may guide other researchers in their studies of careers elsewhere. It is also notable that both the main study reported here and the pilot study mentioned in §5.2.1 produced similar results. This offers some evidence that scaling up the study using a theoretical larger dataset may also replicate these findings, suggesting that they are fairly robust despite the relatively small dataset.

Overall, then, despite some potential limitations of this method and my interpretation of the results, I believe they are justifiable. I will, therefore, use the findings presented in this

chapter to inform my study of the white nationalist forum career in subsequent empirical chapters by using them as a guide to operationalising the abstract concept of the 'career' as described in the sociological literature. Dimension 1 is less obviously helpful for my purposes here, as the community I am studying is online with no need to consider career elements relevant to offline careers. Dimension 2, however, reveals key measures which can be used to recognise and define careers in either environment, namely a consideration of the *loyalty* shown by users to the community (i.e. the amount of time dedicated to the individual by the community, primarily in terms of the duration of their commitment; see §5.2.2.10), and of the *behaviours* exhibited at different points in the career.

Returning to Dimension 1 allows us to consider in greater detail how this can be achieved in online contexts specifically. As Table 4 shows, this dimension highlights the use of *content analysis* and the consideration of *frequency* of engagement as key techniques for studying careers online. These correspond neatly to the two sides of Dimension 2. As I have outlined previously, the text-based nature of much online discourse means that the analysis of language, including the content analysis employed by the scholars whose work is included in this study, is fundamental to understanding online behaviour. Meanwhile, frequency can be seen as similar to group loyalty in that both concern the amount of time the individual has dedicated to their career. For example, two members' careers may be of the same duration, but involve very different levels of engagement, perhaps signalling greater commitment on the part of one of these members. It may therefore be fruitful to consider both duration *and* frequency of engagement to more fully capture the extent of users' engagement with the community. Together, I will refer to these as the *temporal* elements of the career.

In summary, in this chapter, I have used the results of the MCA to identify two overall considerations for the study of careers in online spaces. These will set the agenda for the remaining studies in this thesis. In Chapter 6, I analyse the *temporal* element of careers in a white nationalist forum, using sequence analysis to identify the most frequent patterns of engagement over time. Chapters 7 and 8 pick up the *behavioural* element of careers. I carry out a corpus-assisted discourse analysis to investigate how individuals use language across and within the career types identified in Chapter 6. The analysis of temporal elements of the careers precedes that of behavioural elements for logistical rather than theoretical reasons. In Chapter 6 I identify typical career types before using these to facilitate the design of the linguistic study in terms of selection of users to contribute to comparative subcorpora.

Chapter 6 The temporal career: A sequence analysis

6.1 Introduction

In the previous chapter, I used a multiple correspondence analysis to identify key elements for the empirical study of careers in an online context. These were *temporal* elements (the duration and frequency of engagement with the community) and *behavioural* elements (how the individual engages with the community, with this engagement carried out – and therefore observable – through language). This chapter concerns the former, while Chapters 7 and 8 deal with the latter. The chapter opens with a brief review of how temporal elements of the career have been described in the literature. The details of the sequence analysis carried out here are laid out (see also §4.4.2), with a particular focus on the structure of the dataset required. The findings and interpretation of this computational analysis are then provided, including the implications of these for the remainder of the thesis.

As I discussed in §3.3.2, the literature on careers leaves us with few clear expectations regarding how careers might unfold over time. Centrally, scholars disagree on whether we might expect to see a single typical career path (Becker, 1973; Goffman, 1961), or a highly individualistic situation producing an “irreducible heterogeneity” of career paths (Fillieule, 2010, p. 3). Fillieule’s position leaves us with few expectations beyond a large and heterogenous set of careers, representing a range of patterns of frequency and engagement. It is also difficult to make predictions on the basis of Goffman and Becker’s model, as the size of the present dataset (see §4.3) likely precludes the discovery of a single clear career – compare Becker’s 50 informants to the more than 120,000 data subjects in this study. Beyond this foundational literature, a multitude of approaches have taken to the task of identifying career roles and stages by researchers working in a variety of fields. Typically, typologies are one of two types. The first (e.g. Benne & Sheats, 1948; Turner et al., 2005; Waters & Gasson, 2005) suggest a number of distinct roles, implying a distinct set of temporal career paths associated with each role, but do not take a diachronic perspective, and therefore provide no information on how an individual might be expected to progress in their role over time. The second (e.g. Dennen, 2008; McAuley & Leskovec, 2013; Singh, 2012; Smith Risser & Bottoms, 2014) appear to assume that every individual follows the same career path, progressing through a fixed series of stages over a period of time. However, no suggested timescale is given for this progression. Overall, the wide range of approaches taken and the lack of focus on timescales make it difficult to know how posting careers might unfold in the present research context. Working in the context of other right-

wing extremist fora, Scrivens et al. (2022) may offer the clearest expectations for this study. They find a small number of typical career types in each of their two fora. In both cases, the most common career type is short and marked by low engagement, while less common types show increasing duration and engagement levels.

In this chapter, I employ sequence analysis (Gabadinho et al., 2011; Gabadinho et al., 2010) to visualise white nationalist forum users' posting careers over time. Careers are presented as state sequences, i.e. as sequences of states over time. States are defined here in terms of posting rates per calendar year, as I describe in more detail later in this chapter (§6.2).

6.2 Method

I will employ sequence analysis to analyse the temporal aspects of careers in this forum. Sequence analysis here refers to a means of visualising sequences of events or states as they occur over time, allowing for comparative analysis of different paths which exist in a dataset. I carried out my sequence analysis using the R package *TraMineR* (Gabadinho et al., 2011; Gabadinho et al., 2010), which was developed specifically for social science researchers to identify and compare sequences of biographical data such as careers. *TraMineR* allows the researcher to create a list of all sequences in the dataset arranged by frequency, which will allow me to identify and further analyse the most typical careers.

Sequence analysis works by first identifying a set of possible states from the data. The word *states* refers to the possible 'values' of a sequence. To illustrate, in a simple representation of an occupational career, we might find states such as *student*, *employed*, and *retired*. The set of possible states is referred to as the *alphabet*. For this study, the alphabet is comprised of different levels of engagement with the white nationalist forum, defined by number of posts per calendar year. Sequence analysis will then reveal how engagement changes over time, allowing for a relatively detailed view of the temporal elements of posting careers in this forum. Before describing the alphabet and analysis process in more detail, I first outline the raw dataset used for this study.

6.2.1 Data and data restructuring

The data used for this study is a subset of the same larger dataset described above (see §4.3), taken from a large and long-standing white nationalist forum. This particular dataset disregards the content of the collected posts and is comprised of metadata, namely the number of posts per calendar year from each named user account. The dataset is comprehensive, including all users (n=122,993) who posted at any point between the years 2001 and 2018.

The dataset as I received it (Appendix D) was not immediately suitable for the analysis of career sequences. It was received in a *vertical time-stamped event* format (Gabadinho et al., 2010, p. 30) whereby the number of posts made by each user were listed alongside the year in which they were made. Using R, I restructured the dataset to a *states-sequence* format, “one of the most intuitive and common way[s] of representing a sequence” (Gabadinho et al., 2010, p. 28) as well as the internal format used by *TraMineR*. The two formats are exemplified below in Tables 6 and 7.

Member name	Year	Number of posts
User A	2007	1
User B	2009	1
User C	2002	43
User C	2003	51
User C	2004	1
User D	2007	3

Table 6: Anonymised sample of data as originally received, in the vertical time-stamped-event format

Member name	2002	2003	2004	2005	2006	2007	2008	2009
User A						1		
User B								1
User C	43	51	1					
User D						3		

Table 7: Anonymised sample of data in the restructured state-sequence format. For reasons of space, this sample displays only the years 2002-2009

6.2.1.1 Selection of categorical variables

TraMineR is designed to mine and visualise sequences of categorical variables. This made it necessary to recode the continuous variable, number of posts per year, as a categorical variable. Creating a small number of categories allows for similar career paths to be compiled into a single sequence, and thus for an easier and more meaningful large-scale comparison between sequences, as extremely fine-grained differences are discounted. I chose a set of three categories: 1-10, 11-100, and 101+ posts per year, along with the missing value *NA*, representing zero posts per year. This choice of categories is informed by the aim of the research to meaningfully identify and compare any distinct career paths, as well as by the data itself and by existing literature making use of participation frequency to categorise user types.

The three-category approach I have chosen is consistent with related research. Scrivens et al. (2020, 2022) also employ a three-way distinction between *low*, *moderate* and *high* rates of forum posting, terms I adopt in this thesis for ease of expression. In these papers, the labels are used relatively to the level of activity across the fora; the thresholds for a *low*, *moderate*, and *high* posting rate vary from forum to forum, depending on the overall level of traffic and engagement with the specific forum. For example, a rate of fifteen posts per

month may be a low rate in a very popular forum, and a high rate in a lower-traffic forum. I have followed this principle for the present study. The uppermost of my three categories, a high posting rate of 101+ posts per year, represents a very large range of posting frequencies, as an exceptional few posters make over 10,000 posts in some years. However, it is relatively uncommon for posters to make upwards of 100 posts in a year; taking into account every individual year of posting by every user in the dataset, only 9.18% of 'posting years' show a rate of more than 100 posts. I therefore considered it unnecessary to further divide this category. The purpose of this category is to identify careers with periods of very frequent posting relative to the forum as a whole, which the boundary of 101+ achieves. On the other end of the scale, the lowest rate of 1-10 posts per year represents a very large proportion of cells in the dataframe (70.61%). However, I decided not to further divide this category, again on the basis that the aim of the study is to identify common posting patterns. Whether a user posts 10 times or only once in a year, this represents a (relatively) low rate of posting. The final category – moderate posting rates of 11-100 posts per year – followed from the demarcation of the high and low categories, while maintaining a manageable number of categories. While I believe this selection of categories is well-informed and justified, I recognise that the results reported below would be somewhat different if I had employed a different number of categories or chosen different category boundaries.

Following the creation of categorical variables, the dataframe was pivoted from the vertical style seen in Table 6, to the horizontal style of Table 7. I also arranged the columns created by the pivot in chronological order.

6.2.1.2 *Creation of process time axis*

The next requirement for my purposes was to create a *process time axis*, whereby the origin point of each sequence is set at "the date of a starting event" (Gabadinho et al., 2010, p. 27). Here, the starting event is the first posting activity in the user's career. Put simply, the beginning of each career sequence is said to begin in Year 1, regardless of the actual calendar year of the first post. This was combined with the deletion of missing values to the left and right of the sequence (that is, before the first year of posting, and after the last). Missing values within the sequence were retained, signifying a gap in the posting career. The process time axis contrasts with a *calendar time axis*, whereby the origin for each individual sequence is the same fixed point in time, here the year 2001, when the dataset begins. The two types are illustrated in Figures 5 and 6, which show the 10 most common sequences in the dataset when using each time axis. In both graphs, sequences are presented in ascending order of frequency, with the most common at the bottom.

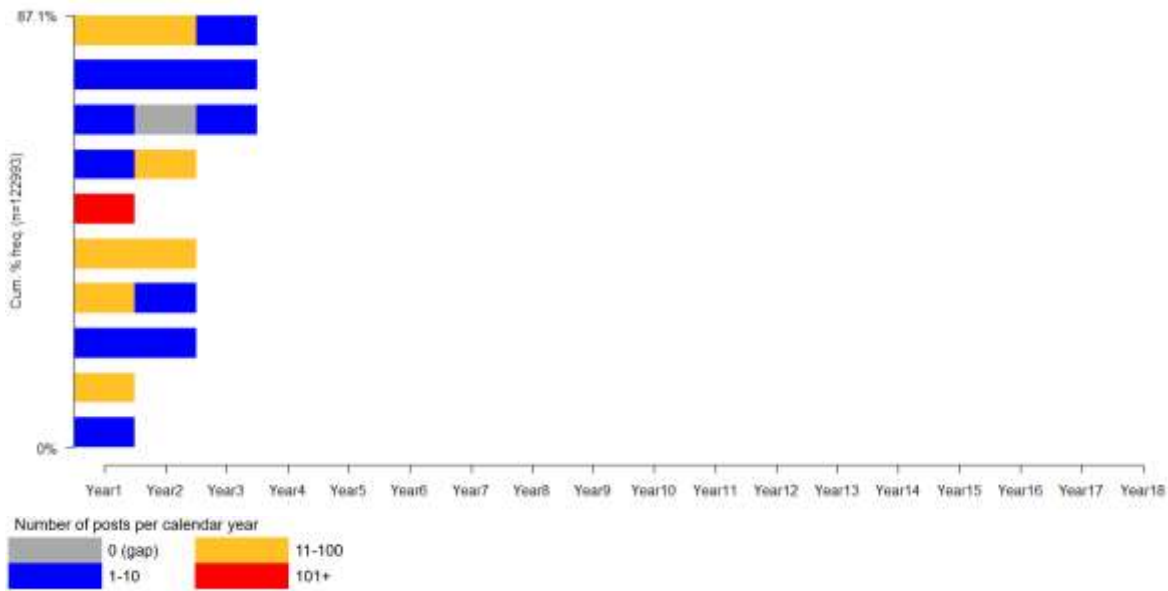


Figure 5: Ten most frequent sequences (process time axis)

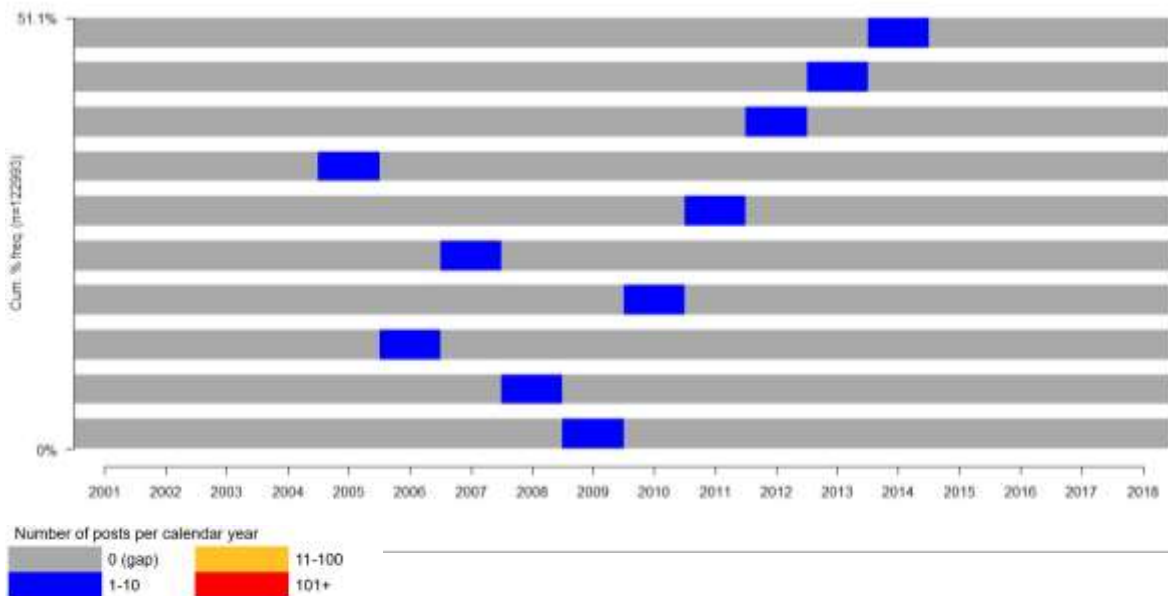


Figure 6: Ten most frequent sequences (calendar time axis)

In Figure 6, all 10 sequences are of the same type, distinguished only by the calendar year in which they occur. For my purpose here, which is to investigate duration and intensity of engagement within the career, these sequences can be considered identical. Figure 6 is therefore an inappropriate way of representing this information, as it redundantly separates identical careers. Meanwhile, Figure 5 combines all this information into a single sequence, the most frequent in the dataset. The process time axis is therefore more suitable for this

research, allowing the most common career paths to be more clearly identified and compared.

The code used to restructure the data and, and to produce the *TraMineR* sequence visualisations seen in Figures 5 and 6 and in subsequent sections, can be found in Appendix E.

6.3 Findings

TraMineR identified 4,225 distinct sequences in the dataset. These sequences are not spread evenly between the 122,993 users in the dataset, with the majority of users following only a small number of relatively homogeneous sequences.

To illustrate this, the 25 most frequent sequences are presented below in both graphical and table form. As usual with *TraMineR* visualisations, the most frequent sequence appears at the bottom of the graph; however, the table of frequencies is presented traditionally, with the highest frequency on top. A complete version of Table 8, with frequency data for all 4,225 sequences, can be found in Appendix F.

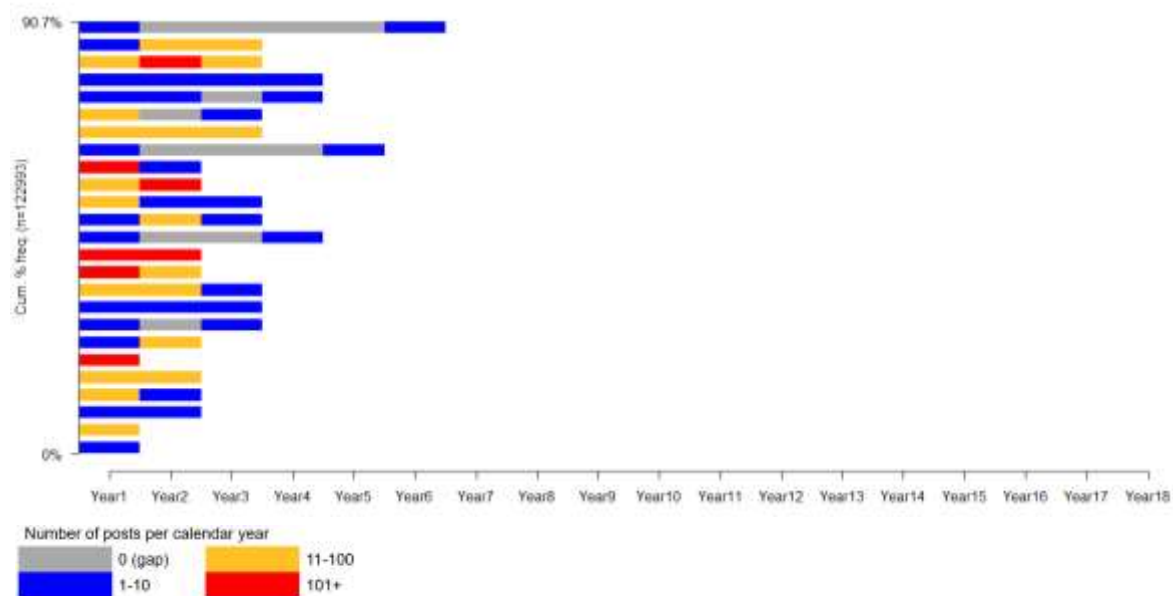


Figure 7: Twenty-five most common sequences

Rank	Frequency	Percentage of users (2 d.p.)	Cumulative percentage of users (2 d.p.)
1	84,149	68.42%	68.42%
2	8,466	6.88%	75.3%
3	6,138	4.99%	80.29%
4	2,281	1.85%	82.14%
5	1,467	1.19%	83.33%
6	1,180	0.96%	84.29%
7	1,059	0.86%	85.15%
8	998	0.81%	85.96%
9	891	0.72%	86.68%
10	514	0.42%	87.1%
11	507	0.41%	87.51%
12	455	0.37%	87.88%
13	448	0.36%	88.24%
14	371	0.3%	88.54%
15	362	0.29%	88.83%
16	335	0.27%	89.1%
17	296	0.24%	89.34%
18	264	0.21%	89.55%
19	224	0.18%	89.73%
20	223	0.18%	89.91%
21	209	0.17%	90.08%
22	200	0.16%	90.24%
23	167	0.14%	90.38%
24	166	0.13%	90.51%
25	158	0.13%	90.64%

Table 8: Twenty-five most common sequences

Among the 25 most common sequences, only one is longer than five years. This is career #25, observed in 158 individuals (0.13% of all users). This career is six years long, but contains a gap of four years, sandwiched between two years of low engagement with the community.⁴

Among the top 25 sequences, low posting rates are very common, followed by moderate. Only six of these sequences contain any periods of high engagement, with only one containing high engagement in more than one year (sequence #12, two years of high-intensity posting, observed in 455 or 0.37% of individuals). While the two most common careers comprise one year of low and moderate intensity posting (observed in 84,149 or 68.42% of users, and 8,466 or 6.88% of users respectively), the equivalent high-intensity career ranks at #6, observed in just under 1% of users (0.96%; n=1,180). The intervening careers are all of only two years' duration, made up of low (#3, 4.99% or n=6,138), moderate (#5, 1.19% or n=1,467), or mixed engagement (#4, 1.85% or n=2,281); in other words, these

⁴ Recall that *engagement* here can refer only to posting, and not other types of 'silent' engagement, which are not captured in this dataset.

slightly longer careers with lower engagement are more common than single-year, high-intensity careers, emphasising the relative rarity of high-intensity posting.

Sequences of three or more years begin at #8 (0.81%, n=998), and those of four or more at #13 (0.36%, n=448). There is only one career of five years' duration (#18, 0.21% or n=264) and one of six (#25, 0.13% or n=158). The longest career is the least frequent among these 25. Career gaps are fairly frequent in careers of three or more years, appearing in six of the top 25 (out of 14 'eligible' careers of three or more years). Beyond the top 25, the most frequent career of each length is always of the same 'shape' as #25, described above: a gap of varying length, sandwiched between a single year of low engagement at each end of the career. In the top 25, only one career of three or more years' duration contains a period of high engagement (#23, 0.14% or n=167).

The first 25 sequences alone account for just over 90% of all users. This top 25, and the dataset as a whole, are dominated by the single most frequent sequence, which comprises a single year of low-intensity posting and accounts for 68.42% of users (n=84,149). Even the second most frequent sequence, one year of moderate-intensity posting, is observed in only 6.88% of users (n=8,466), just over a tenth of those observed for the first. As Table 9 shows, only five sequences are observed in over 1% of users each, and only seven in 1,000 or more users.

Number of sequences observed in:		% of seqs.	Total number of users:	% of users
1 user	3,270	77.4%	3,270	2.66%
2 users	378	8.95%	756	0.61%
3 users	144	3.41%	432	0.35%
4 users	77	1.82%	308	0.25%
5-10 users	165	3.91%	1,103	0.9%
11-50 users	134	3.17%	2,818	2.29%
51-100 users	24	0.57%	1,748	1.42%
101-1,000 users	26	0.62%	7,818	6.36%
1,000+ users	7	0.17%	104,740	85.16%

Table 9: Number of sequences observed in certain numbers of users

Of the 4,225 distinct career sequences, 77.4% (n=3,270) are only observed in a single individual. This means that while 122,993 users follow 4,225 career paths, 119,723 of these individuals follow just 955 paths. A further 599 sequences (14.18% of the total) are observed in between two and four users. This leaves 356 career paths (8.43%) which are observed in five or more individuals; these 356 are shared between 118,227 users. We have already seen that a single career path (#1) is shared by 84,149 users.

Career length	Number of users	Percentage of users	Cumulative percentage
1 year	93,795	76.26%	76.26%
2 years	12,661	10.29%	86.55%
3 years	5,266	4.28%	90.83%
4 years	3,111	2.53%	93.36%
5 years	2,118	1.72%	95.08%
6 years	1,600	1.3%	96.38%
7 years	1,086	0.88%	97.26%
8 years	838	0.68%	97.94%
9 years	647	0.53%	98.47%
10 years	523	0.43%	98.90%
11 years	423	0.34%	99.24%
12 years	285	0.23%	99.47%
13 years	208	0.17%	99.64%
14 years	177	0.14%	99.78%
15 years	125	0.1%	99.88%
16 years	69	0.06%	99.94%
17 years	35	0.03%	99.97%
18 years	26	0.02%	99.99%

Table 10: Number of users following sequences of each length

Table 10 shows the number of users following careers of each possible duration (i.e. between one and 18 years). This table, which summarises the entire dataset, unsurprisingly shows very similar patterns to those identified in the top 25 sequences. Given the prevalence of sequence #1, it is also unsurprising that 76.26% of users (n=93,795) have careers of only a single year. The second highest frequency career type, the two-year career, drops sharply to only 10.29% of users (n=12,661). This pattern holds across the dataset, with the number of users decreasing with each increase in career length. While two-year careers alone represent around 10% of users, to reach the next 10% of users, careers of between three and six years' duration must be combined (9.83% or n=12,095). Meanwhile, the remainder of the dataset – careers of seven or more years – account for only around 4% of all users (3.61% or n=4,442). Only a very small percentage – 0.02%, or 26 individuals – posted in both 2001 and 2018, representing the longest possible careers, and even fewer engaged every year for that period. As Figure 8 shows, 15 of these careers show at least one gap of at least one year. Almost all of these individuals follow distinct career paths over their 18 years, with the exception of two individuals (marked #9 and #17 here), who share the same path of a single year of moderate posting intensity followed by 17 years of high intensity.

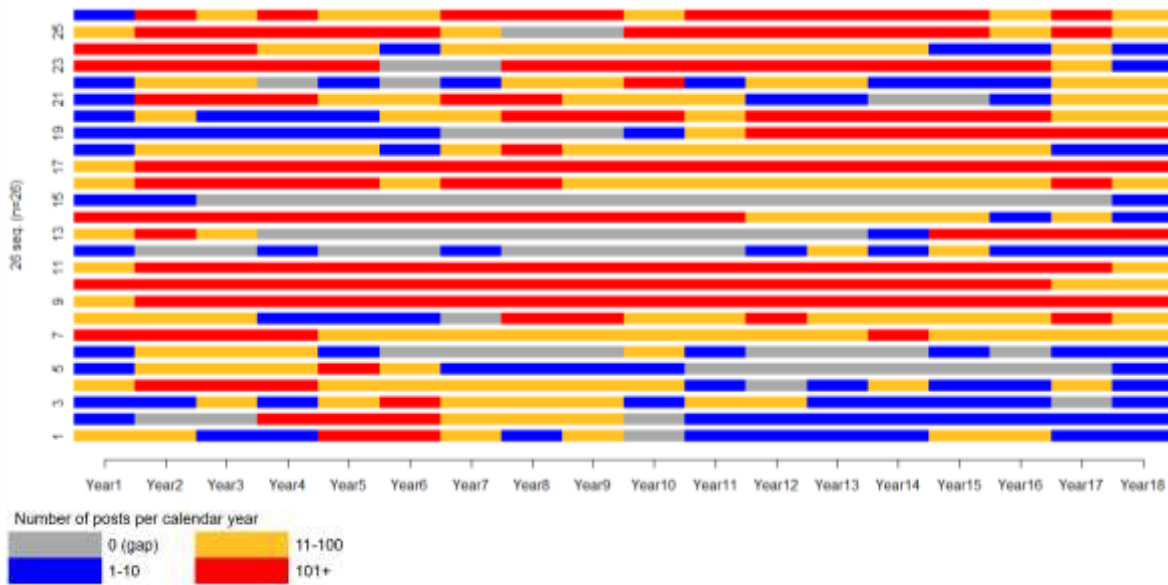


Figure 8: All 26 eighteen-year career sequences

The following observations are based on the most frequent 102 sequences in the dataset, representing 94.16% of users (n=115,806). This smaller sample is selected to allow for more thorough investigation and description of a wider range of career paths, including longer careers than those found in the top 25. This would be unfeasible across the entire sample of 4,225 sequences. Sequences #101 and #102 have the same frequency as #100 (23 users), so all three were included to avoid excluding them arbitrarily.

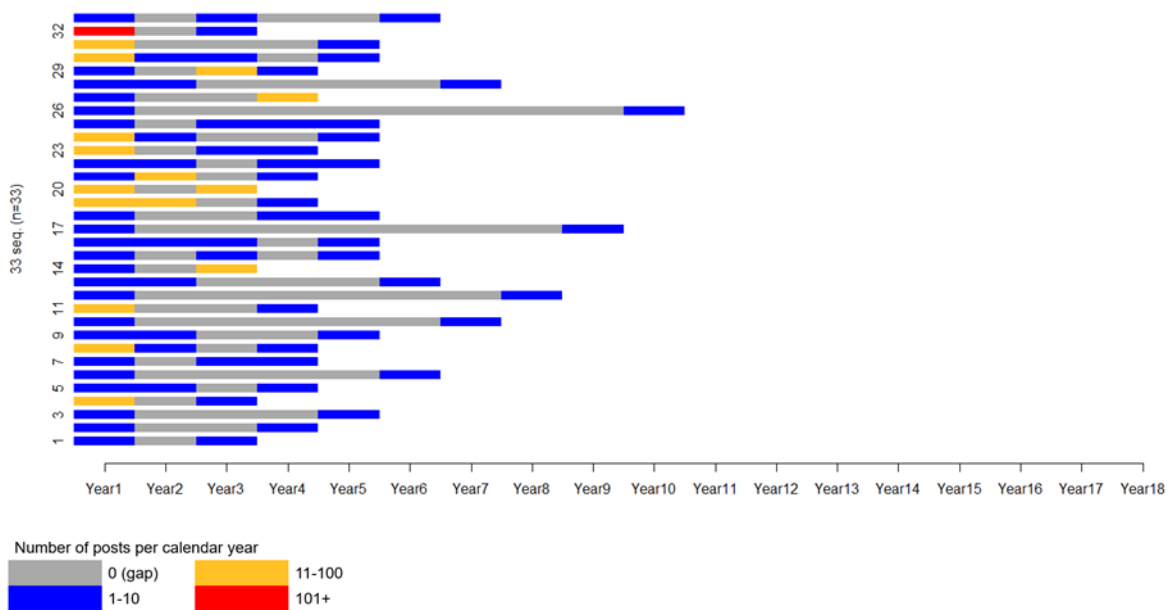


Figure 9: All careers with gaps in the most frequent 102 sequences

Career gaps are fairly frequent in this sample. Out of 91 eligible sequences (of three or more years), 36.26% (n=33) contained a gap. This represents 39.49% (n=3,692) of the 9,350 individuals who follow one of these eligible careers. When all 102 sequences (eligible and ineligible) are included, this figure naturally drops, with 3.19% of the 115,806 individuals taking a career gap of one calendar year or longer. The most frequent pattern of career gaps is a single gap of one year, found in 16 of the top 102 careers and representing 2,100 individuals. Fifteen careers (1,505 users) display posting gaps of two or more years, ranging up to 8 years (marked #17 in Figure 9; 83rd most frequent overall, 31 users). Careers with multiple gaps are much rarer in this sample, represented in only two careers (those marked #15 and #33, respectively 52nd and 98th most frequent, totalling 87 users). Generally, careers with gaps in this sample show low-to-moderate engagement even in active years. Only one career, #97 (observed in 26 users) has a period of high intensity posting. Upon returning from a career gap, most individuals show low engagement for the remainder of the career (exceptions are the careers marked #14, #20, #27, and #29 in Figure 9, totalling 167 individuals).

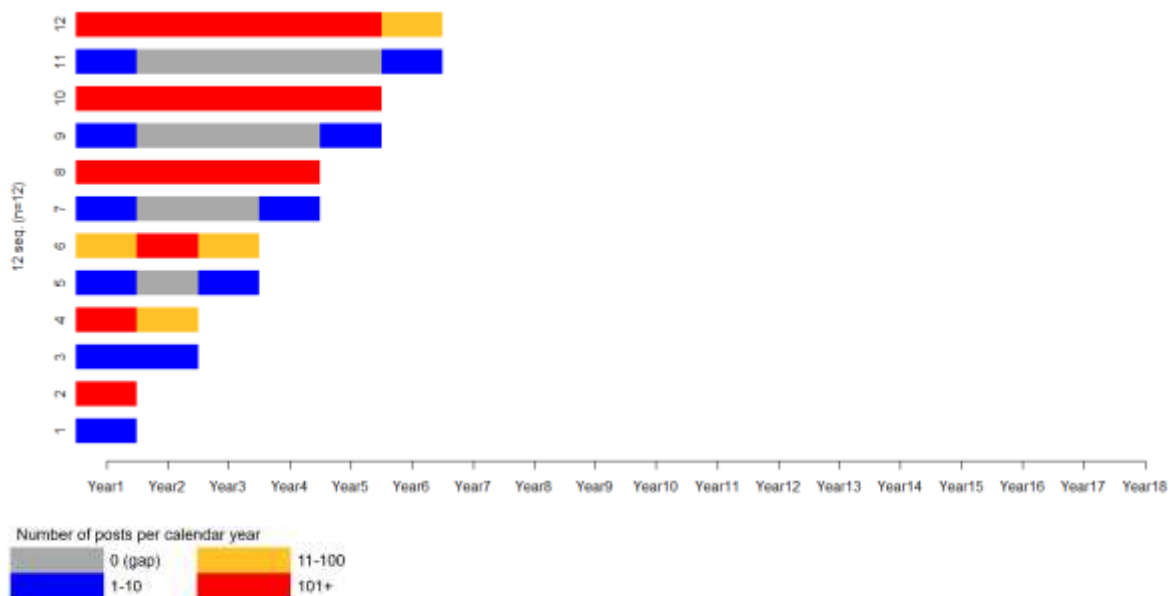


Figure 10: Subsample of the most frequent 102 sequences, showing a) the most frequent sequences of each length and b) the most frequent of each length containing at least one year of high-intensity posting

Figure 10 reveals that, for every career length above two years, the most frequent sequence contains a gap of at least one year. This offers further support for two of the above findings: first, gaps are frequent in the dataset at all eligible career lengths; second, career gaps tend to be accompanied by low engagement. However, Figure 10 also shows a complementary finding. While periods of low or no engagement tend to co-occur in the same careers, periods of high engagement also seem to co-occur with one another. This is the case for the four-, five-, and six-career careers in this sample.

I also undertook a simple qualitative analysis of trajectories, inspired by Scrivens et al.'s (2022) computational trajectory modelling. Here, I distinguish *consistent* careers, where the rate of posting is within the same bracket for the entire career; *increasing* and *decreasing* careers, where the posting rate increases or decreases as the career unfolds over time; and *mixed* trajectories, non-linear careers with periods of both increased and decreased engagement. These observations are naturally somewhat obscured by the set of sequence states I have selected for this analysis. A career marked as comprising three years of moderate engagement would be identified as following a consistent trajectory here, while in reality an individual may have posted 100, 50, and 11 times over the three years respectively. This is of course a limitation, but some general patterns may still be observed. Single-year careers are excluded from these observations, as they cannot be said to follow a trajectory over time. This leaves 20,011 members across the remaining 99 career sequences.

Trajectory	Number of users
Consistent	9,776
Length of career	
2 years	8,060
3 years	1,263
4 years	328
5 years	102
6 years	23
7+ years	0
Intensity of career	
Low	7,294
Moderate	1,756
High	726
Increasing	2,023
Low to moderate	1,389
Moderate to high	483
Low to high	160
Low to moderate to high	0
Decreasing	5,103
High to moderate	860
Moderate to low	3,620
High to low	461
High to moderate to low	162
Mixed	5,100
Excluding careers with gaps	1,408

Table 11: Number of users following different career trajectories in the top 102 sequences

By far the most common trajectory is the consistent, which is observed in 9,776 individuals. Among these consistent career paths, the most frequent is unsurprisingly the two-year, low-intensity careers, accounting for 6,138 users alone. The number of users following each consistent career path decreases as both duration and engagement increase. The next most frequent trajectory, observed in 5,103 users, is the decreasing trajectory. Most users follow a

gradual pattern of decreasing engagement, although a certain number (n=461) step directly down from high to low engagement. Almost equally as common is a mixed trajectory (n=5,100). This group includes careers containing gaps, which I have already explored in more detail above. A smaller number of users (n=1,468) follow a mixed trajectory with no gaps. The least common trajectory is the increasing type, observed in 2,032 users. Again, upstepping is typically gradual, although a small number of users (n=160) step directly from low to high engagement.

6.4 Discussion

One of the most striking findings from this study is the high level of homogeneity among careers in this environment. On the whole, forum members follow roughly the same typical career paths. This is broadly consistent with the view taken by Becker (1973) and Goffman (1961). Typically, careers in this forum are short, often just a single year of posting, with generally lower engagement; this is similar to the findings reported by Scrivens et al. (2022). This is particularly notable in the case of the most common career, a single year characterised by low engagement. This career alone dominates the dataset, representing ten times more users than even the second most common. Although distant runners up, the next few most common careers share similarities with the most frequent. Among the top seven – the only careers shared by at least 1,000 users – all but #5 share either their one-year duration, or a period of low intensity, with #1. All in all, over 85% of users follow one of this set of similar paths.

In particular, careers that can be considered both long in duration *and* high in intensity are rare. Among the top 102 most frequent sequences, the longest careers with even a single period of high engagement are #99 and #102, six-year careers observed in only 25 and 23 individuals respectively. Searching for a career of seven or more years with a period of high engagement takes us to #120, observed in 19 users. Interestingly, all three of these careers are made up entirely or almost entirely of high engagement. It seems that many users fall into an 'all or nothing' pattern of engagement in these slightly longer careers, whereby users who post with high intensity in even one year of their career are likely to do so for more extended periods.

These findings are optimistic, suggesting that on the whole users do not assimilate deeply into the community and its white nationalist collective identity. There are of course limitations to this assumption. Based on this dataset, we cannot determine whether an individual goes on to, or has already, become heavily involved in other white nationalist groups or organisations elsewhere online or offline. Nonetheless, the generally low rates of engagement over time are a promising sign. The smaller pool of individuals who follow

longer career paths is a useful place to start with the investigation of behavioural elements of the career, as they presumably show greater behavioural change across their longer careers.

Although I have identified that there is a typical career path in this forum, this path is not what the literature which describes a typical career might lead us to expect (§3.3.2).

Although timescales are not mentioned in this literature, the claim of multiple consecutive stages in papers such as Dennen (2008), McAuley and Leskovec (2013), Singh (2012), and Smith Risser and Bottoms (2014) implies a career taking place over a longer span. This contrasts with the very short typical career I have identified here. Some researchers who describe typical careers do recognise that individuals may end their career prematurely, never progressing to later stages. Becker (1973) describes this phenomenon in the deviant career of the marijuana user, noting that the most common career path in his study involves remaining at a stage of casual use rather than progressing along the path to become a habitual user. However, this is a matter less of duration and more of frequency of engagement, and still implies sustained participation over a long period. Preece and Shneiderman (2009, p. 16) note that “a decreasing number of people” are likely to progress to each subsequent stage in their framework, logically making the earliest stage likely to be the most populous. However, they give no sense of early career cessation on the scale that we see in this study. The results here are, however, consistent with Scrivens et al. (2022) – whose work also concerns right-wing extremist fora – both in identifying short, low-intensity careers as typical, and in recognising and describing the existence of other, less common careers.

While I have shown that a typical career path does exist in this forum, variation in careers is also significant. There are a very high number of distinct career paths, with many of these observed in only a few users or even a single individual. Non-linear career paths – those with fluctuating engagement over time, including career gaps – are fairly common, although less so overall than linear careers. However, the possibility of non-linear careers is relatively underrepresented in the literature. Scrivens et al. (2022), whose data and research questions are similar to my own and who produce some similar results to those reported here, find that rapidly decreasing posting rates are extremely common. This finding is replicated in my own results in that the end of the career – the cessation of posting – naturally represents a decrease in posting rate. However, the same researchers find that consistent posting over time – what they call a chronic trajectory – is completely absent in one forum, and accounts for just under a quarter of posters in the other. Contrary to this, almost 45% of careers of two or more years in this set of 102 careers feature consistent rates of posting throughout. This rate drops among longer careers, which tend to show more

variation in posting rate. Note again that my analysis of trajectories was based on a smaller, more manageable sample of careers, so there are limitations to the conclusions I can draw. However, this sample accounted for over 94% of all users' careers. Looking at careers beyond the top 102 sequences we notice that longer careers do tend to have more gaps and fluctuations. This is exemplified in Figure 8, which shows the longest careers in the dataset, those of 18 years' duration. Despite the incompleteness of the sub-dataset of 102 careers, it was promising to find that relatively few users follow increasing trajectories. This lends further support to the notion that only smaller numbers of users become further assimilated into a white nationalist collective identity.

The prevalence of variation in this dataset leads me to conclude that it is not appropriate in the case of this forum to focus only, or even primarily, on the typical career. As I have already detailed, I am making the assumption that those who follow the typical short career path do not assimilate towards a collective identity. Therefore, in order to achieve my overarching goal of exploring the white nationalist collective identity of this forum and improving understanding of the white nationalist threat, I must also investigate the rarer, longer careers. In order to more manageably handle the significant variation in career sequences, I have identified four main career types based on duration, which are summarised in the table and figure below.

Career type	Career duration	Number of users	Percentage of users
Mayfly	1 year	93,795	76.26%
Dragonfly	2 years	12,661	10.29%
Mantis	3-6 years	12,095	9.83%
Cicada	7+ years	4,442	3.61%

Table 12: Four career types

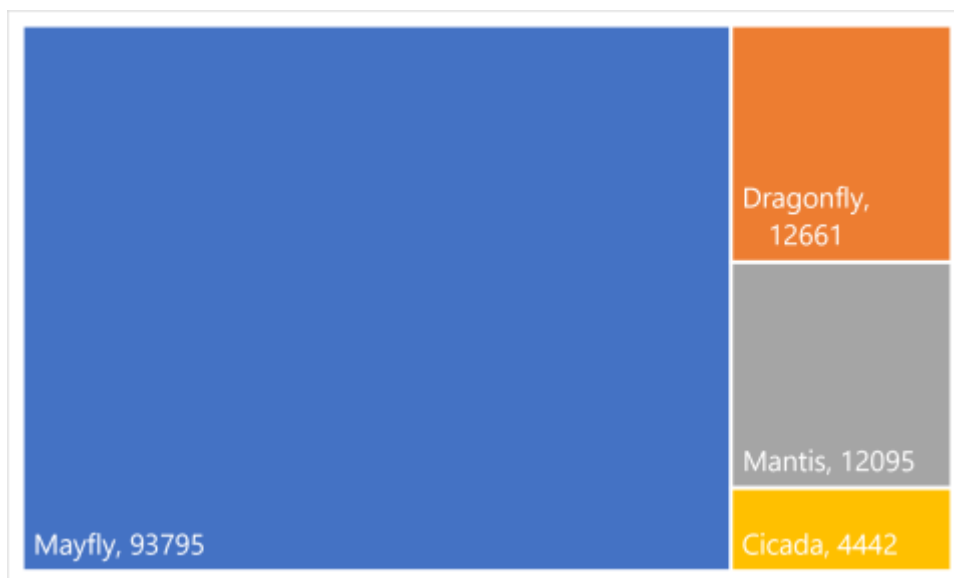


Figure 11: Four career types

The first type, the *mayfly* career, is the typical career type, representing over three quarters of all users. Named for the mayfly's fleeting, hours-long lifespan, this is the shortest career type, taking place within a single calendar year. As we have already seen, this often involves only a low rate of posting in that time. The next two types, *dragonfly* and *mantis* careers, each account for 10% of users. Like the lifespan of the dragonfly, measured in weeks or even days, the second career type is longer than that of the mayfly, representing careers of two years. Likewise, the mantis, which may live for several months, lends its name to the third career type, comprising careers of between three and six years. The remainder of the careers, those of seven years or longer, are named for the *cicada*, which may live for close to 20 years. This is the rarest career type, representing around 4% of users. In subsequent chapters, I use these four types – the mayfly, the dragonfly, the mantis, and the cicada – to further investigate users' assimilation towards a collective identity in this forum. This involves returning to behavioural aspects of careers, as identified in Chapter 5, in order to examine *how* individuals following different career path types engage with the community.

The findings reported here do not offer a clear solution to the central disagreement found in the foundational sociological literature on the career. The number of different career paths, and especially the number of paths that are observed in only one user, lends credence to Fillieule's (2001, 2010) idea that careers are often unique to the individual. However, we do not find the "irreducible heterogeneity" of careers described by Fillieule (2010, p. 3). As Goffman (1961) and Becker (1973) predict, despite a degree of individual variation, the findings clearly show a single, typical career path. Arguably, both views – of the heterogeneity and the homogeneity of the career, respectively – are legitimate interpretations of these findings. In order to resolve this tension, then, I propose to take the middle ground between the two positions by investigating the selection of four career types identified above – the mayfly, dragonfly, mantis, and cicada careers – which represent both typical and less typical sequences. This approach incorporates investigation of the typical career with a recognition of the significant variation we also see. It also takes into account the practicalities of working with large datasets, which do not allow for the entirely individual approach recommended by Fillieule.

In this chapter, I have identified four major career types in this white nationalist forum on the basis of career duration. In the following chapters, I will return to the behavioural or linguistic element of the career, using the four types identified here as the basis for a corpus-assisted analysis of discourse across and within careers.

Chapter 7 The behavioural career I: A quantitative linguistic analysis

7.1 Introduction

In Chapter 5, I employed multiple correspondence analysis to identify key dimensions of variation in the academic literature on careers. I found that the first of these two elements reflects the level of temporal commitment to the group, mostly in terms of career duration but also the frequency of engagement. I explored this in more detail in Chapter 6. On the basis of career length, I identified four distinct groups of users, which I called mayflies (1 year of engagement; the most populous group), dragonflies (2 years), mantises (3-6 years), and cicadas (7+ years of engagement; the least populous group).

Across the next two chapters, I will return to the second element, which reflects the behaviours that members exhibit as part of the community. As previously discussed in Chapter 5, the idea of *behaviour* was used to capture a number of elements and was broadly defined as concerning “the nature of participation which members engage in” (§5.2.2.9). For the purpose of this study, the behaviour under investigation is limited to linguistic behaviour, i.e. linguistic expression. This derives from the nature of the community as an online environment where members’ participation is largely text-based and linguistic (Seargeant & Tagg, 2014; Tagg, 2015). In addition, language is, as I have established in §3.8.2, an often-overlooked element of collective identity creation and career progression in the sociological literature, despite being widely recognised elsewhere as a central resource in the construction of identity.

The literature on language across what we might understand as the career (§3.4) offers some expectations regarding the findings here. This literature is primarily qualitative and pragmatic in nature, and rarely identifies specific linguistic constructions associated with different career stages, which is my goal in this chapter (a qualitative approach to this data can be found in Chapter 8). Several researchers find that the use of first person singular pronouns is common in the early career, while first person plural pronouns are more frequent in the later career (Angouri & Sanderson, 2016; Danescu-Niculescu-Mizil et al., 2013; Nguyen & Rosé, 2011; Rudolf von Rohr et al., 2019). This suggests that singular pronouns may be more associated with short careers, and plural with long careers. Benwell and Stokoe (2006) and Golder and Donath (2004) also report that modal verbs and other hedging devices are more common among new users as a negative politeness strategy (Brown & Levinson, 2006). On the lexical level, several researchers (e.g. Angouri &

Sanderson, 2016; Benwell & Stokoe, 2006; Chiang, 2018; Graham, 2016) find that new users often make explicit reference to their newness, while longer-standing members use more topic-specific vocabulary (Danescu-Niculescu-Mizil et al., 2013; Nguyen & Rosé, 2011; Peuronen, 2011). Newcomers may demonstrate respect for the community using apologies (Golder & Donath, 2004) and thanks (Nguyen & Rosé, 2011), while later-career members may take a gatekeeping role by welcoming them to the community (Angouri & Sanderson, 2016).

The present chapter takes a higher-level quantitative approach. First, I describe an exploratory corpus analysis (§7.2), using keyness analysis to identify features which appear to be associated with the career types I identified in the previous chapter. This is followed by a more detailed exploration of correlations between these features and career types (§7.3). Finally, I investigate how these correlations with user type relate to the language use over time of a handful of individual long-term *cicada* users (§7.4). In the following chapter, I investigate these selected users in more depth to gain a better qualitative understanding of their language use throughout their individual careers.

7.2 Exploratory corpus analysis

7.2.1 Data and method

A subcorpus was created for each of the four user types. To allow for more meaningful comparisons between the four subcorpora, each was designed to have approximately the same number of words. By calculating the average number of posts per user, I was able to ascertain how many users of each type would be required to create subcorpora of approximately equal size. The size of each subcorpus was matched to that of the dragonfly group, who produced the fewest total posts. A language detection script was run on each of the subcorpora to exclude posts in languages other than English. The four subcorpora are summarised in the table below. They can be found in Appendix G.

Group	Number of posts	Number of words	Number of members
Mayflies	396,433	34,893,454	70,325
Dragonflies	399,415	29,135,826	11,455
Mantises	396,639	26,454,993	5,105
Cicadas	392,719	26,542,958	2,931
<i>Total</i>	<i>1,585,206</i>	<i>117,027,231</i>	<i>89,816</i>

Table 13: Summary of the four subcorpora

I used Sketch Engine to analyse the corpus due to its ability to process very large datasets. To explore the corpus, I used Sketch Engine's key n-grams tool, using the remainder of the corpus as the reference corpus for each focus subcorpus in turn. Experimentation showed that 3-5 grams were the most useful range, as bigrams did not provide enough information

about the n-gram and its potential meaning(s), while the inclusion of 6-grams resulted only in repetition of strings already found in the 3-5grams.

Sketch Engine uses the simple maths method (Kilgarriff, 2009) to calculate keyness scores. This method allows the researcher to set the N-value for the keyness calculation. The N-value controls the frequency of the n-grams identified by the keyness analysis. A lower N-value (which is standard) shows lower frequency n-grams which are more strongly identifying of the focus corpus compared to the reference corpus and therefore have higher keyness scores. A higher N-value instead shows higher frequency n-grams which are still relatively distinctive between the two corpora. As a result, the keyness scores are closer to 1 than for lower N-values. I chose a relatively high N-value of 1,000, which does not show lower-frequency, more distinctive topic-related n-grams, but instead higher-frequency, somewhat less distinctive n-grams which reveal more structural elements of the focus subcorpus. Losing these topical elements is desirable here, as I assume that topic is controlled for across the four subcorpora, at least to the extent that interests me here. The present chapter seeks not to identify what ideas the four types of users express, but rather how those different groups express themselves linguistically.

7.2.2 Findings

I examined the first 100 key n-grams for the mayfly and cicada subcorpora. Details of these, along with their frequencies and keyness scores, can be found in Appendix H. These two subcorpora were selected for reasons of efficiency, as I assume that the difference between these two groups is maximal. I manually analysed the key n-gram lists and identified patterns in the types of key linguistic expressions across the two subcorpora. As this stage of the research was intended to be exploratory and to identify themes for later further investigation, this analysis was largely based on the key n-grams as they appear in the list (that is, without context). However, I also consulted a small selection of concordance lines for more ambiguous n-grams, in order to confirm my interpretations were representative of the reality of the data.

7.2.2.1 Mayflies

The mayfly subcorpus is largely characterized by key n-grams which appear to relate to the presentation and introduction of posters' identities. Many of the n-grams signal explicit claims about the self, using first person singular pronouns:

*I am a; I am not; I live in; my name is; and I am; that I am; I have been;
that I am; this is my; I am from; because I am; I am an; I was born; am not
a; live in a*

While the prevalence of first person singular pronouns shows mayflies' interest in making claims about their own individual identities, a second set of n-grams reveals that they also orient to their perceived social or group identities, and emphasise this as a key part of their identity:

proud to be; be proud of; to be proud; proud of my

Consulting the concordance lines for these n-grams confirms that mayflies are claiming they are 'proud to be white' or 'proud of [their] white heritage'. This shows a concern with introducing both individual and group identities, as users emphasise their enthusiastic identification with the central identity category on which the forum is based.

Mayflies also perform their (individual) identity within the wider group of forum members through orientation to their newness:

my name is; I am new; new to this; my first post; new to the

This is achieved through explicit mentions of newness and through introductory posts wherein members present themselves to the community. Both means of performing newness are in line with other literature on newcomers in online spaces (e.g. Arguello et al., 2006; Chiang, 2018; Golder & Donath, 2004; Graham, 2016; Nguyen & Rosé, 2011).

Two other sets of n-grams may also relate to the performance of newness. The first is the expression of goals as the user looks forward to their planned forum career ahead and their planned activities as a forum member.

would like to; I would like; I want to; I would like to; look forward to; just want to; just wanted to

Concordance lines show that major themes surrounding these n-grams include suggestions, plans, seeking information and opinions ('would like to know'), and hopes for the future. In particular, 'look forward to' shows users' plans for their use of and engagement with the forum and its users. The presence of 'just' in 'just want/wanted to' reveals the use of negative politeness strategies (Brown & Levinson, 2006) by mayflies, signalling a desire not to offend or unduly intrude in an environment to which they are newcomers. Again, this reflects existing research on newcomer behaviour in online spaces (Benwell & Stokoe, 2006; Golder & Donath, 2004).

We also see references to the site itself, which may link to the expression of newcomer identity as users engage directly with an environment which is new and different for them.

on this site; on this forum; and welcome to [site]; to be here; people on this

Indeed, collocation lines here show that major themes surrounding these n-grams include users' impressions of the site and its users, both positive and negative. 'To be here' is largely used to express that the user is 'glad/pleased to be here'. On the contrary, some users also perform a less traditional newcomer identity, as the n-gram 'and welcome to [site]' shows; this string performs a gatekeeper's function typically associated with longer-standing members. However, the prevalence of other related n-grams such as 'from NW [U.S. state] and welcome' and '[U.S. state] and welcome to [site]' shows that many 'welcome' tokens originate from a single user. This suggests that this gatekeeping function is likely not characteristic of mayflies in general.

At a higher level, we can remark that many of the n-grams mentioned so far are expressed through a small number of similar verb types which might generically be described as 'stative' verbs. A more thorough discussion of the meaning of 'stative' will follow (section §7.3.1.3); in brief, it can be understood as referring to verbs which describe a state of being, rather than an action or activity. Such verbs are extremely prevalent in this key n-grams list:

I am a; I am not; I live in; would like to; my name is; I would like; and I am; what is your; I want to; I would like to; you want to; is your favorite; I am new; that I am; proud to be; do you think; be proud of; to be white; white people are; and I have; live in a; black people are; to be here; this is my; you guys are; look forward to; I am white; that there are; I am from; to be proud; because I am; that you are; just want to; just wanted to; I am an; people who are; to live in; I was born; am not a; if you want

Other notable features in the n-gram list include direct engagement with other users, primarily achieved through second person pronouns.

you don't; what is your; some of you; why do you; you want to; is your favourite; many of you; and welcome to [site]; what is your favourite; do you think; what do you; how do you; brothers and sisters; most of you; all of you; you guys are; that you are; you can't; if you want; thank you for

This category includes 'and welcome to [site]', although as discussed above this may result from a single unusual user. Inspection of concordance lines for 'brothers and sisters' reveals that this phrase is often used by posters to greet or otherwise engage with their fellow forum members. These 'brothers and sisters' are often further specified as 'white' or 'WN [white nationalist] brothers and sisters', tying back into mayflies' emphatic reference to and pride in their own identification with the shared social identity of whiteness. Another example of this data following expectations from the literature is found in 'thank you for'; newcomers

commonly thank existing members for their contributions (Chiang, 2018; Nguyen & Rosé, 2011).

Finally, despite my modification of the keyness calculation to favour discursive-structural elements rather than topics, the topic of race nevertheless emerges in the key n-grams list:

the white race; reparations reparations reparations; to be white; white people are; black people are; a black man; of the white; of other races; I am white; a black person; a white nationalist; a white person

We may hypothesise that these n-grams, particularly those relating to whiteness and white nationalism, are part of the same phenomenon as the 'pride' category, and reflect the introduction and presentation of self with an emphatic expression of white identity. Nevertheless, this was an unexpected finding from a key n-gram list in a forum that concerns itself with race, and it is interesting to note that newcomers potentially discuss such issues relatively more frequently than longer-standing users.

7.2.2.2 Cicadas

Contrary to the mayfly group which is largely focused on issues of identity, the cicada group instead appear to make assertions not about the self, but about external phenomena. This is reflected in a large number of key n-grams which apparently express categorical assertions, very few of which contain a first person pronoun of any kind:

it is a; is going to; it will be; would have been; are going to; will have to; won't be; it was a; they don't; he didn't; that is the; we don't; there is no; was going to; there will be; he doesn't; they can't; will be a; would be a; this is a; going to be; need to be; he is a; has to be; it isn't; we have a; didn't have; needs to be; that is a; it doesn't; n't going to; going to have; he would have; we have to; it won't; doesn't have; he was a; will not be; I would have; you can get; they won't; don't need; that is what; we can't; will be the; it is the; they didn't; I wouldn't; should have been; n't have to; they need to; they will be; and it is; and that is; that is what

Although modals are frequent in this category, meaning that these n-grams do not strictly speaking show bare or categorical assertions, these are 'modals of certainty' (Biber et al., 1999), which increase the epistemic commitment expressed compared with 'modals of likelihood' (compare *will, would* with *may, might*). These constructions therefore apparently reflect strong and assertive claims. Of course, any of these expressions may be modified with hedging constructions, thus reducing the author's epistemic commitment to the proposition. Nonetheless, it is noteworthy that the more categorical or assertive elements of

cicadas' propositions, rather than any hedging elements present in the co-text, are present in the key n-grams list.

A subsection of these assertion n-grams, especially those which contain modal verbs, reveal that cicadas make predictions about the future and other hypothetical situations:

is going to; it will be; would have been; are going to; will have to; won't be; there will be; will be a; would be a; going to be; if it is; n't going to; going to have; he would have; it won't; will not be; I would have; they won't; will be the; they will be

This contributes to a sense of certainty and knowledge of how the external world and other actors function. On the other side of this, we also see a number of n-grams showing reflection on the past:

a few years ago; back in the; few years ago; over the years; at this point; a long time

This is perhaps unsurprising from longer-standing members, who we may expect to reflect on their time as a group member and how the world, and their situation, has changed in that time. Cicadas may be using these reflections to position themselves as 'veterans' (Newsome & Grant, in press) – members with significant experience and knowledge gained over a long tenure as a group member – and to claim the social power that derives from this.

Where claims about the self rather than the external world are made by cicadas, these seem to concern users' thoughts and actions, rather than their identity.

I wonder if; is a good; I wonder how; reminds me of; I would have; it looks like; looks like a; I wouldn't; I think he; needs to be

Indeed, even here reference to the self is often not overt; while some of these n-grams include first person singular reference, others show that authors encode their perceptions and stance in impersonal ways. The insertion of authorial voice and stance suggests an interest among cicadas in presenting and claiming their own viewpoints and assessments, and in this way appears to complement the prevalence of categorical assertions. Both categories of n-gram suggest that cicadas are concerned with presenting themselves as knowledgeable. Perhaps therefore relatedly, we also see a category of n-grams which appear to reflect the use of examples in order to illustrate an argument:

a couple of; one of the; one of those; one of these; in this case; more than a; this sort of

These n-grams again suggest that cicadas are constructing arguments and expressing their own point of view.

Similarly to the mayflies, cicadas show a selection of key n-grams which are apparently used to engage directly with other users in the forum community. While second person pronouns were very common among the mayfly n-grams, cicadas also achieve this effect in other ways:

welcome to [site]; to have you; you can get; in this thread

While the mayfly n-grams suggest that they engage with others mostly by asking questions, the cicada n-grams suggest a range of different functions. While both groups perform a welcoming function, there is evidence in the concordance lines that this sentiment is expressed by many different cicadas, which does not appear to be the case for the mayfly group (see above). Welcoming newcomers to the community is commonly associated with later-career users (Angouri & Sanderson, 2016). Cicadas also offer suggestions and/or information to other users ('you can get'), and explicitly engage with and evaluate the content of other posts in a thread ('in this thread').

Also concerning pronoun use, it is interesting to note that the first person singular pronouns which are common among mayflies are almost absent in the key cicada n-grams. Among the cicadas, we instead find a category of n-grams containing first person *plural* pronouns:

of our people; we don't; we have a; we have to; we can't

These perhaps hint at the development of a collective identity and/or a greater awareness of the self as a member of some community in the later career, as Danescu-Niculescu-Mizil et al. (2013) and Nguyen and Rosé (2011) suggest.

Finally, much like for the mayflies, a topical element reveals itself in the key n-gram list despite my primary interest in structural features. Rather than race, the cicadas show themselves to be interested in discussion of political institutions and organisations, both mainstream and white nationalist:

the white house; the federal government; the republican party; the federal reserve

the national alliance; knights of the; the ku klux klan; the council of conservative citizens

This focus on mainstream politician institutions is reminiscent of the findings of my previous work (Booth, 2023) where I found that, despite the site's claimed focus on racial issues,

political outsiders actually receive more attention, and may be perceived as a greater threat, than racial outsiders. Meanwhile, the mention of white nationalist organisations may align with the greater prevalence among cicadas of n-grams containing first person plural pronouns, suggesting a greater focus on and understanding of the wider white nationalist community.

7.2.3 Discussion

These exploratory results provide a preliminary understanding of the differences in linguistic behaviour which exist between the shortest- and longest-standing members of the forum community. The most striking difference comes from the mayflies' focus on the self and the cicadas' on the external world. Mayflies are apparently primarily concerned with presenting their own identity, particularly their individual identity, evidenced by the prevalence of key n-grams containing first person singular pronouns. On the contrary, these pronouns are rare among the cicadas' key n-grams, where categorical assertions about the external world (signalled by third person pronouns and especially the impersonal pronoun *it*) dominate. Where first person pronouns do appear, these are plural. Cicadas' focus has apparently shifted from the individual to the collective, suggesting a greater consciousness of themselves as members of a group.

The two groups also differ in their relation to knowledge and authority. Mayflies directly engage with others, largely through second person pronouns; this apparently has a largely information- or advice-seeking function (e.g. 'what is your', 'why do you'). On the other hand, cicadas use multiple resources to make claims, construct arguments, and present their own perspective, suggesting a concern with appearing knowledgeable or even authoritative.

Finally, while the purpose of the present exploratory work was to identify structural features of the two groups' discourse, both key n-gram lists revealed that actors external to the site are a key topic of conversation. For mayflies, this focus is racial; they emphasise their own white identity, and offset this with the mention of other races, highlighting the importance of this distinction. For cicadas, political institutions and organisations are key. These include both insiders to the white nationalist movement (e.g. 'the national alliance', 'ku klux klan'), and mainstream actors more distant from this white nationalist centre (e.g. 'the white house', 'the federal reserve').

However, these findings are naturally limited and exploratory. They are based only on interpretation of the meaning of key n-grams, with only limited reference to what these phrases 'really' show in their true discourse contexts as revealed by the concordance for each term. Nonetheless, the findings are promising. They show that there are, in fact, qualitative differences between the language used by mayflies and cicadas respectively,

supporting the hypothesis that linguistic behaviour changes across the career. They are also broadly aligned with existing literature on newcomer and expert linguistic behaviour in online spaces (e.g. Benwell & Stokoe, 2006; Golder & Donath, 2004; Newsome & Grant, in press), suggesting that the patterns I have observed do result from the distinct career types rather than from chance differences. Although these results are only preliminary in terms of what they can tell us about the differential behaviour of career types, they are useful in that they offer a “way in” (Baker, 2012, p. 248) to thinking about the concrete linguistic ways in which career types may differ, which I can explore in more quantitative and qualitative detail. The next part of this chapter details the linguistic features I have selected for further analysis on the basis of the findings presented here, and present quantitative results on the prevalence of these features across career types.

7.3 Correlation of linguistic features with career type

7.3.1 Selection of features

7.3.1.1 *Epistemic commitment and stance*

The first feature selected for closer analysis across the corpus is *epistemic commitment*, defined by Simpson (1993, p. 48) as being “concerned with the speaker’s confidence or lack of confidence in the truth of a proposition expressed.” Epistemic commitment exists on a continuum from strong commitment (the proposition *must* or *must not* be true) to weak commitment (the proposition *may* or *may not* be true).

Epistemic commitment is one element of a set of systems of linguistic expression which have variously been referred to as *stance*, *evaluation*, *appraisal*, and other terms. As this variety of terms suggest, these systems have been defined in slightly different ways by different researchers. However, one frequently used definition of this phenomenon is that given by Biber et al. (1999, p. 966), which states that stance is an expression of “personal feelings, attitudes, value judgements, or assessments” of people, things, situations, or ideas. I adopt Biber’s terminology and definition here.

Stance encompasses a number of different linguistic functions. Most saliently, Martin and White (2005) distinguish between *attitude* – our feelings, emotional reactions, and evaluation of people and things – and *engagement*, the adoption of “a position with respect to propositions” (p. 40). The latter, which encompasses Simpson’s (1993) definition of epistemic commitment, is my focus here. Epistemic commitment, and engagement more generally, can be expressed in a variety of ways. Indeed, Martin and White’s (2005) highly functional *appraisal* approach posits that there is no one-on-one relationship between form and function, and therefore that almost any linguistic form may, in the right context, express

engagement. Simpson (1993), as well as Biber (2006) and colleagues (1999), take a distinctly more formal approach, outlining those linguistic forms which are most commonly associated with the expression of engagement. Despite their more restricted approach to identifying engagement, these forms are still numerous: modal verbs, adverbs, and adjectives, nouns, and verbs of stance.

I have chosen to focus on epistemic commitment for three main reasons. The first is simply that such elements were captured in the key n-gram lists discussed in the previous section. In particular, the expression of epistemic commitment appears to be relevant in characterising the linguistic behaviour of cicadas, as they employ bare assertions and modal verbs showing strong epistemic commitment, as well as overt insertions of the authorial voice in order to explicitly evaluate propositions. The exploration of epistemic commitment as a facet of stance is also theoretically fitting. Expression of stance has long since been tightly linked to identity performance, and in fact the earliest understandings of the link between language and identity centred around stance or *footing* (Goffman, 1981), through which individuals position themselves in relation to their audience(s), the content of their talk, and to existing discourses around the chosen topics (Bakhtin, 1981). This is therefore well suited to a study of (collective) identity. Finally, as I am assuming that the topics and opinions found on the special interest forum I am studying are generally controlled for, my interest is not in these views per se, but in how these views are expressed linguistically. This is captured by a study of stance.

Biber (2006) provides an extensive list of epistemic stance resources. This includes constructions of both *certainty* and *likelihood*, which are at the two ends of epistemic commitment described by Simpson (1993). The former increases the author's commitment to the truth of the proposition, and consequently closes off to recognition of alternative views. The latter reduces commitment, and implicitly recognises that other views may be valid. This list forms the basis of my own study of stance, and is reproduced below. I have made only minor modifications, which are detailed in footnotes. This formal approach, while limited (e.g. Martin & White, 2005) is well-suited to corpus analysis, as search strings can easily be designed to capture a finite list of specific combinations of words and grammatical constructions.

The list of epistemic stance resources, adapted from Biber (2006), is as follows:

Modal and semi-modal verbs

- Of possibility, permission and ability
 - o *Can, could, may, might*
- Of prediction and volition

- *Will, would, shall, be going to*

Epistemic stance adverbs

- Of certainty
 - *Actually, always, certainly, definitely, indeed, inevitably, in fact, never, of course, obviously, really, undoubtedly, without doubt, no doubt*
- Of likelihood
 - *Apparently, evidently, kind of, in most cases/instances, perhaps, possibly, predictably, probably, roughly, sort of, maybe*

Stance complement clauses controlled by verbs

- Epistemic stance verb and *that*-clause⁵
 - Of certainty
 - *Conclude, demonstrate, determine, discover, find, know, learn, mean, notice, observe, prove, realis/ze, recognis/ze, remember, see, show, understand*
 - Of likelihood
 - *Assume, believe, doubt, gather, guess, hypothesis/ze, imagine, predict, presuppose, presume, reckon, seem, speculate, suppose, suspect, think*
- Epistemic stance verb and *to*-clause
 - Of probability (likelihood)
 - *Appear, happen, seem, tend*
 - Of cognition/perception (likelihood)⁶
 - *Assume, believe, consider, expect, find, imagine, judge, know, presume, suppose⁷*

Stance complement clauses controlled by adjectives

- Epistemic stance adjective and *that*-clause
 - Of certainty

⁵ For *that*-clauses, the corpus query was designed so as to capture both overt and zero-marked *that* (e.g. *I remember that he said...* and *I remember he said...*). I also created separate queries to identify *that*-clauses attributed to the first person singular, in order to capture the most explicit insertions of the authorial voice.

⁶ A handful of verbs suggested by Biber – *forget, learn, pretend, remember* – are excluded, as examination of my corpus shows that epistemic uses of this construction are rare or absent.

⁷ Examination of my corpus shows that *suppose + to* often has a deontic function alongside the epistemic. There is a general pattern in the data that tokens of *suppose + to + be* are epistemic, while other instances tend to be deontic. Therefore I only included *suppose + to + be* in this search.

- *Apparent, certain, clear, confident, convinced, correct, evident, false, impossible, inevitable, obvious, positive, right, sure, true, well-known*⁸
- Of likelihood
 - *Doubtful, likely, possible, probable, unlikely*
- Epistemic stance adjective and *to*-clause⁹
 - Of certainty
 - *Certain, guaranteed, sure*
 - Of likelihood¹⁰
 - *Apt, liable, likely, prone, unlikely*

Epistemic stance noun and *that*-clause

- Of certainty
 - *Assertion, conclusion, conviction, discovery, [no] doubt [that], fact, knowledge, observation, principle, realis/zation, result, statement*
- Of likelihood
 - *Assumption, belief, claim, contention, feeling, hypothesis, idea, implication, impression, notion, opinion, possibility, presumption, suggestion*

7.3.1.2 Pronouns

The second feature of interest for further analysis is personal pronouns. These pronouns are, clearly, a question of identity; understanding who is referred to by the pronouns *I, we, you, they*, etc shows how the author positions themselves in relation to various others. Due to this, the study of pronouns has been widely used in discourse analysis, especially critical discourse analysis (Fairclough, 2013). Critical discourse analysis aims to shed light on power relations, which, as Wodak and Meyer (2015) highlight, ultimately flow from the fundamental distinction between *us* and *them* – proximal and distal, from the point of view of the author. This is of particular relevance for discourse analysis of extremist groups and other racist discourses; creation and maintenance of boundaries between those who belong, and those who do not, are central in such contexts (§2.5).

In this study, I focus on first person pronouns (both singular and plural) and second person pronouns. While third person pronouns, mostly notably *they*, are recognised as important to the construction of identity through their position in the central dichotomy of *us* and *them*, these pronouns have been excluded from the present study for several reasons. While *they*

⁸ The string *well known* is also included.

⁹ I categorised these as expressions of certainty and likelihood myself, as Biber does not do so.

¹⁰ The construction *due to* was excluded, as it proved too difficult to only exclude tokens with the meaning 'because of'.

often has a boundary-drawing, even antagonistic, function in discourse, in practice it is also frequently used in a neutral way to refer to any third party, including those who are part of the in-group of interest. While first and second person pronouns have a limited range of possible referents, *they* may refer to any actor or even an inanimate object. This makes analysis of this pronoun impractical in large datasets, as significant qualitative analysis is required to determine the referent. *They* is also mostly absent from the key n-gram lists for mayflies and cicadas, further justifying its exclusion here.

On the contrary, the first and second pronouns show interesting contrasts between the two career types. The first person singular pronouns that are common for the mayflies is seemingly 'swapped' for the first person plural in the key n-gram list for the cicadas, suggesting a transition from perception of the self as an individual, to being part of a group of some kind. While the distinction between *us* and *them* has received significant attention from (critical) discourse analysts (e.g. Fairclough, 2013; Wodak & Meyer, 2015), that between *us* and *me* has barely been explored. Under what circumstances do authors position themselves as individuals, and when do they position themselves instead as members of a group? Turning to second person pronouns, while these occur in the key n-gram lists for both career types, they appear in more mayfly key terms. These pronouns also apparently tend to serve different functions across the two career types.

For completeness, the subject, object and possessive forms of all selected pronoun types are included in the relevant corpus queries.

7.3.1.3 *Dynamicity and stativity*

The final linguistic feature arising from the analysis of key n-gram lists is the distinction between stative and dynamic verbs. The mayfly list shows that mayflies frequently make claims about their identity and state of being, which are expressed using stative rather than dynamic verbs.

The literature on dynamic and stative verbs reveals that finding a suitable definition of the two terms is extremely difficult, and scholars have different perspectives on the details of how best to categorise verbs into the two types. However, some general properties appear across multiple authors' work. While statives can be described as denoting "states of affairs" (Biber et al., 1999, p. 458; Binnick, 1991, p. 183), *dynamic* verbs refer to events, acts and processes. The latter implies some end point or completion; dynamic situations do not inherently persist through time, and will come to an end unless effort is exerted (Binnick, 1991; Smith, 1991). Dynamic situations can therefore often be understood as voluntary, or involving some degree of agency (Binnick, 1991). On the contrary, statives imply continuity over a certain period of time, although this can be a short period. The state will persist

without effort, unless some external agency intervenes. The continuity of statives over time also extends to their static, unchanging nature throughout their duration; unlike dynamic situations, there is no change of state or internal phases or stages within this time (Binnick, 1991; Huddleston & Pullum, 2002; Smith, 1991). Statives can overall be understood to *hold* or *obtain*, while dynamic situations *occur*, *happen*, or *take place* (Huddleston & Pullum, 2002; Smith, 1991). As Smith neatly summarises, “[a]lthough states obtain in time they do not take time” (1991, p. 19). Statives are agreed to be by far the smaller of the two categories, with multiple subcategories of dynamic verbs existing (Dowty, 1979; Mourelatos, 1978; Smith, 1991), although the difference between these is beyond the scope of the current work. A sample of dynamic and stative verbs mentioned in the literature include:

Dynamic: *learn, listen, look, buy, run, walk, drive, notice, write, drink, bathe, sing, arrive, laugh, seek, repair, paint*

Stative: *know, appreciate, hear, consider, desire, love, believe, want, intend, understand, own, be, have, hope*

Matters are complicated by the fact that dynamicity and stativity are often properties of the sentence or clause rather than of the verb itself, and the same verb can have a dynamic or stative meaning in different contexts. Several researchers mention ‘serial states’ (Binnick, 1991; Huddleston & Pullum, 2002), also known as ‘habitual states’ (Smith, 1991), whereby a stative meaning is derived from dynamic verbs – for example, *he teaches English* or *she smokes*. In these examples, a series of habitual occurrences over time amounts to a state. Other elements of the discourse context can also change our understanding of the dynamicity of a verb, such as the following examples from (Smith, 1991, p. 18):

Bill knew the truth.

Suddenly Bill knew the truth.

While the former is stative – Bill is in the state of knowing the truth – the introduction of the adverb *suddenly* turns the latter into a dynamic verb, describing a moment of realisation. Inversely, Huddleston and Pullum (2002, p. 168) highlight how the same situation can be conceptualised in dynamic or stative ways:

Dynamic: *read a book*

Stative: *have one’s head buried in a book*

As this discussion demonstrates, there can be no exhaustive list of dynamic and stative verbs. This is disadvantageous for a corpus-based study, which requires some kind of formal

linguistic characteristic to search for within the corpus. To manage the problem of the variability of dynamicity from context to context, and to recognise that some verbs nevertheless tend to refer to one situation type more frequently than the other, Smith (1991) proposes that each verb has an *intrinsic* status which “depends on its value in a minimal sentence” (p. 177). In this view, the wider context of the sentence can instead endow the verb with a *derived* meaning which changes its situation type. This is the case for *Suddenly Bill knew the truth*, where *knew* gains a derived dynamic meaning through the addition of *suddenly*. Out of necessity, I assume that all verbs in the corpus take their intrinsic meaning, recognising that this introduces a degree of error into my quantitative findings. However, as we will see later in this chapter, I am interested at this stage in the relative frequencies of different linguistic features across my four subcorpora; as the level of error can be assumed to be consistent across all four, this is a reasonable compromise in order to be able to quantify dynamicity and stativity across career types.

In order to create a list of dynamic and stative verbs to search for in the corpus, I used the word list function in Sketch Engine to identify the 50 most frequent verbs across the entire corpus, then categorised these as dynamic or stative with reference to the literature on dynamicity. While no exhaustive list of either verb type can exist, individual papers often include smaller lists of common verbs used for illustrative purposes, which were used as a starting point for categorisation. I also applied the tests for stativity which are commonly found across the literature (e.g. Binnick, 1991; Dowty, 1979; Huddleston & Pullum, 2002; Lakoff, 1966; Smith, 1991), and the detailed definitions of different verb types, to categorise more ambiguous verbs. The categorisation process raised several issues. The first relates to the verb *do*; I ultimately decided to exclude it from the analysis, as it was too difficult to separate tokens of *do* as a lexical verb from its dummy and auxiliary uses. The other relates to the verb *start*. This verb is neither stative nor dynamic, but is instead a verb of aspect. Rather than excluding this verb, I decided instead to include *aspectuals* as a third category of verb, split into three categories of *starting point*, *end point*, and *progress*, containing verbs taken from Biber et al. (1999):

Starting point: *begin, start, commence, proceed*

End point: *finish, stop, complete, end, cease, quit*

Progression: *continue, keep, carry on, keep on, go on*¹¹

I made the choice to include these additional verbs, rather than exclude *start*, hypothesising that these may be more closely associated with the cicada group. This follows from the

¹¹ *Go on* is added to this list, following Huddleston and Pullum (2002).

finding from the key n-grams that cicadas are apparently more likely to reflect on the past and past events, and to position themselves as a veteran of the community.

From this discussion, it will be clear that my approach to operationalising dynamicity and stativity for a large-scale corpus analysis has limitations. However, by following the literature on these concepts, I have created list of verbs representing a reasonable proxy for these concepts.

7.3.2 Calculation of correlation coefficients

To quantitatively compare the linguistic behaviour associated with different career types, I conducted a series of searches in Sketch Engine. Each search corresponded to one of the features I described in the previous section. For each subcorpus, I recorded the relative frequencies (per million words) of every linguistic feature of interest. The results of these searches are presented in the following three tables.

Epistemic stance resource	Mayflies	Dragonflies	Mantises	Cicadas
Modals/semi-modals: possibility, permission, ability	5,567.81	5,610.4	5,702.1	5,849.5
Modals/semi-modals: prediction and volition	7,771.08	8,080.98	8,447.63	8,690.03
Adverbs: certainty	3,665.87	3,623.54	3,656.73	3,554.54
Adverbs: likelihood	1,221.53	1,246.3	1,281.03	1,304.76
Verb and <i>that</i> -clause: certainty	1,199.58	1,140.12	1,125.28	1,064.89
<i>I + verb + that-clause: certainty</i>	390.33	369.9	365.26	334.34
Verb and <i>that</i> -clause: likelihood	1,452.99	1,399.93	1,402.01	1,317.75
<i>I + verb + that-clause: likelihood</i>	833.5	837.91	851.91	794.51
Verb and <i>to</i> -clause: probability (likelihood)	491.1	462.24	492.18	459.82
Verb and <i>to</i> -clause: cognition/perception (likelihood)	72.77	74.56	77.99	75.33
Adjective and <i>that</i> -clause: certainty	195.94	191.44	193	193.06
<i>I am + adjective + that-clause: certainty</i>	42.11	42.35	42.51	38.75
Adjective and <i>that</i> -clause: likelihood	24.78	26.19	26.51	28.35
<i>I am + adjective + that-clause: likelihood</i>	1.55	1.96	2.38	1.87
Adjective and <i>to</i> -clause: certainty	18.47	21.63	23.91	24.68
Adjective and <i>to</i> -clause: likelihood	67.33	61.32	62.27	59.62
Noun and <i>that</i> -clause: certainty	266.34	239.12	223.99	212.38
<i>My + noun + that-clause: certainty</i>	6.53	7.5	7.02	8.33
Noun and <i>that</i> -clause: likelihood	173.36	152.62	132.95	140.21
<i>My + noun + that-clause: likelihood</i>	6.82	5.77	4.41	5.29

Table 14: Relative frequencies (per million words) of epistemic stance resources across the four career-type subcorpora

Pronoun type	Mayflies	Dragonflies	Mantises	Cicadas
1 st person singular	26,947.5	25,590.6	24,600.1	22,587.5
1 st person plural	7,385.16	7,754.52	7,631.38	7,979.75
2 nd person	14,979.9	12,960.3	12,344.5	11,763.7

Table 15: Relative frequencies (per million words) of pronoun types across the four career-type subcorpora

Verb type	Mayflies	Dragonflies	Mantises	Cicadas
Dynamic	24,440.31	24,717.35	25,013.51	24,720.03
Stative	58,606.84	56,626.84	56,071.47	55,405.86
Aspectual	675.52	718.06	740.37	739.87
Starting point	356.42	395.95	403.78	394.28
End point	165.73	155.88	160.26	165.97
Progression	153.37	166.23	176.33	179.87

Table 16: Relative frequencies (per million words) of verb types (dynamic, stative, and aspectual) across the four career-type subcorpora

Having collected the frequency per million words for each feature of interest, I measured the groupwise correlation coefficients for each category of features, using the *dplyr* package in R for data manipulation (Wickham et al., 2021). For the purposes of the calculation, each career type was assigned a number, whereby mayflies = 1, dragonflies = 2, etc. Rather than using the relative frequencies reported above, a ranking approach was taken, whereby the relative frequencies of a feature for each career type were replaced with a number from 1-4, with 1 denoting the lowest relative frequency among the four groups. This is to account for the wide range of relative frequencies represented across the dataset.

The reason for this is illustrated in the two following tables:

Pronoun type	Subtype name	Career type	Relative frequency
First person singular	<i>I</i>	1	19,426.5
First person singular	<i>I</i>	2	18,623.9
First person singular	<i>I</i>	3	18,279.1
First person singular	<i>I</i>	4	16,999.4
First person singular	<i>me</i>	1	2,854.78
First person singular	<i>me</i>	2	2,646.94
First person singular	<i>me</i>	3	2,390.15
First person singular	<i>me</i>	4	2,114.38
First person singular
Correlation coefficient:		-0.0711	

Table 17: Illustration of correlation coefficient calculated on the basis of relative frequency

Pronoun type	Subtype name	Career type	Ranked frequency
First person singular	<i>I</i>	1	4
First person singular	<i>I</i>	2	3
First person singular	<i>I</i>	3	2
First person singular	<i>I</i>	4	1
First person singular	<i>me</i>	1	4
First person singular	<i>me</i>	2	3
First person singular	<i>me</i>	3	2
First person singular	<i>me</i>	4	1
First person singular
Correlation coefficient:		-1	

Table 18: Illustration of correlation coefficient calculated on the basis of ranked frequency

Table 17 shows that the pronouns *I* and *me* within the group *first person singular pronouns* follow the same trend: that is, they decrease in usage across the four career types, suggesting a strong correlation between this feature and career type. However, because of the very different actual values involved in the calculation (c. 18,000 tokens for *I*, and 2,500 for *me*), the correlation coefficient for this calculation is very weak, only -0.0711. Where the same decreasing trend is observed for similar values, the correlation is much stronger, as can be seen in Table 18. I therefore chose to use the ranking approach exemplified in Table 18, to allow correlations of feature frequency with career type to be more clearly observed. This method results in some data loss, as the magnitude of difference in frequency across career types is lost. However, as I am using the correlation coefficients only as a kind of ‘screening procedure’ to identify features of interest for further analysis (see §4.4.3), rather than the coefficients being an end in themselves, I consider this level of data loss to be acceptable for my purposes here.

Having decided on a ranking approach and therefore creating an ordinal variable out of the relative frequencies, I nevertheless decided to employ the groupwise correlation method rather than the Spearman’s rho method, which is designed for ranked data. Again, this decision relates to my use of the correlation coefficients as a kind of screening process for further analysis. The groupwise method allowed for a more detailed insight into correlations, taking into account more datapoints from the feature list. It therefore resulted in a larger and more detailed range of correlation coefficients, and therefore a more discerning screening process for further investigation. For example, using the Spearman’s rho method, over three quarters of the epistemic stance features were shown to have a ‘very strong’ correlation with career type, a result which I felt did not reflect the true relative frequencies of these features across career types.

The correlation coefficients for the features identified in §7.3.1 are presented below. The results here are colour coded according to the strength of correlation, following Evans (1996):

Correlation strength	Coefficient
Very weak	0.00 – 0.19
Weak	0.20 – 0.39
Moderate	0.40 – 0.59
Strong	0.60 – 0.79
Very strong	0.80 – 1.00

Table 19: Correlation strength descriptors, following Evans (1996)

Feature category	Feature	Correlation coefficient
Pronouns	Second person pronouns	-1
Pronouns	First person singular pronouns	-0.7
Stance	I + verb + <i>that</i> -clause: certainty	-0.6353
Stance	Noun + <i>that</i> -clause: certainty	-0.5364
Stance	Verb + <i>that</i> -clause: certainty	-0.4471
Stance	Noun + <i>that</i> -clause: likelihood	-0.4286
Stance	Certainty (overall)	-0.3
Stance	Verb + <i>to</i> -clause: probability (likelihood)	-0.2
Dynamicity	Stative verbs	-0.2
Stance	Adverbs: certainty	-0.1714
Stance	My + noun + <i>that</i> -clause: likelihood	-0.1685
Stance	Verb + <i>to</i> -clause: cognition/perception (likelihood)	-0.1181
Stance	Adjective + <i>that</i> -clause: certainty	-0.1014
Stance	Verb + <i>that</i> -clause: likelihood	-0.1014
Stance	I am + adjective + <i>that</i> -clause: certainty	-0.0755
Stance	Likelihood (overall)	-0.06
Stance	Adjective + <i>to</i> -clause: likelihood	-0.04
Dynamicity	Dynamic verbs	0.025
Stance	My + noun + <i>that</i> -clause: certainty	0.0461
Stance	I + verb + <i>that</i> -clause: likelihood	0.1571
Stance	I am + adjective + <i>that</i> -clause	0.2
Stance	Adverbs: likelihood	0.2545
Dynamicity	Aspectual verbs: starting point	0.3
Dynamicity	Aspectual verbs: end point	0.333
Stance	Modals/semi-modals: prediction and volition	0.4
Stance	Adjective and <i>that</i> -clause: likelihood	0.4384
Dynamicity	Aspectual verbs: progression	0.52
Stance	Modals/semi-modals: possibility, permission, ability	0.7
Pronouns	First person plural pronouns	0.7
Stance	Adjective + <i>to</i> -clause: certainty	0.9333

Table 20: Strength of correlations between linguistic features and career type, colour-coded according to Evans (1996)

Note that negative correlations are more strongly associated with shorter careers, and positive correlations with longer careers.

On the whole, over half of the correlations are weak, and over a third very weak, showing that despite the indications of the key n-gram lists, not all of the proposed features correlate with career type. Nevertheless, there are some stronger correlations, with pronominal features in particular showing strong correlations with career type. Directionally, these mirror the findings from the analysis of key n-grams. First person plural pronouns are strongly

correlated with longer careers, while their singular counterpart shows the inverse correlation. Taken together, these findings may indicate a growing sense of collective identity over time, as users begin to identify and express themselves less as an individual and more as part of, and on behalf of, the group as a whole.

Of the three feature categories, dynamicity showed the weakest correlation with career length, with only aspectual verbs of progression reaching even a moderate correlation strength. These verbs are moderately correlated with longer careers, perhaps indicating, as hypothesised, that longer-standing members reflect on what they perceive to be ongoing issues or concerns. Elsewhere, the use of stative verbs shows a somewhat stronger (negative) correlation with career length than dynamic verbs, reflecting the prevalence of identity claims using *be* in the mayfly key n-gram list; this correlation is, however, weak overall.

Along with the correlation coefficients for individual epistemic stance features, I also calculated coefficients for expressions of certainty and likelihood as a whole. Overall, these showed only weak and very weak correlations with career length, respectively. However, of the four stance features associated with shorter careers, three express certainty (increased epistemic commitment) and only one likelihood (decreased commitment). The latter is the *noun + that-clause* construction, which is negatively correlated with career length in both its certainty and likelihood forms, indicating a moderate preference for this form among members with shorter careers. The two epistemic stance features which are more associated with longer careers are more evenly balanced, with one feature expressing certainty and one likelihood. This may suggest a minor tendency towards expression of certainty (upstepping epistemic commitment), rather than likelihood (downstepping epistemic commitment), among members with shorter careers.

Both categories of modal verbs are more strongly associated with longer careers. In the case of modals of prediction and volition, this finding is in line with my interpretation of the key n-grams, which appears to show that cicadas are more likely to make predictions about the future.

For further quantitative and qualitative analysis, I selected only those features which show at least a moderate correlation with career length. For ease of reference, I will refer to these as 'mayfly features' – those associated with shorter careers – and 'cicada features' – those associated with longer careers. The selected features are summarised below:

Mayfly features		Cicada features	
Feature	Coefficient	Feature	Coefficient
Second person pronouns	-1	Adjective + <i>to</i> -clause: certainty	0.9333
First person singular pronouns	-0.7	First person plural pronouns	0.7
I + verb + <i>that</i> -clause: certainty	-0.6353	Modals/semi-modals: possibility, permission, ability	0.7
Noun + <i>that</i> -clause: certainty	-0.5364	Aspect verbs: progression	0.52
Verb + <i>that</i> -clause: certainty	-0.4471	Adjective + <i>that</i> -clause: likelihood	0.4384
Noun + <i>that</i> -clause: likelihood	-0.4286	Modals/semi-modals: prediction, volition	0.4

Table 21: Features selected for further quantitative and qualitative analysis

7.4 Case study: Five individuals

7.4.1 Method

Having identified a set of features for further examination, I turn now to investigating the use of these features by individual forum users. The remainder of this chapter takes a quantitative approach, examining the frequency of usage of these features throughout these users' careers. The next chapter takes a qualitative approach, focusing on *how* features are used different at different stages of their careers.

In order to investigate the use of linguistic features throughout the career, I selected the five most prolific cicadas in the corpus and created a dataset for each, comprising all of each user's posts organised chronologically. These were uploaded to Sketch Engine for analysis. Details of the users and datasets can be seen below. The full datasets can be found in Appendix I.

User	Career length (years)	Years active	No. of posts	No. of words
User 1	14	2005-09; 2012-18	4840	336,479
User 2	12	2005-16	4089	114,180
User 3	9	2005-13	3843	356,742
User 4	15	2004-18	3405	116,678
User 5	7	2008-14	3120	276,865

Table 22: Details of users selected for further analysis

For each user corpus, I used Sketch Engine to calculate the relative frequency over time for each of the linguistic features selected for closer analysis. As posts in the corpora were arranged chronologically, I split the corpus into ten deciles and recorded the frequency and

relative frequency of each feature per decile. This allowed me to identify how usage of each feature changed throughout the users' careers.

7.4.2 Findings

In this section, I present and discuss a quantitative analysis of the selected linguistic features across the five users' careers. I begin by presenting these on the level of the feature, combining results from all five users, to examine whether these follow the expected patterns of usage over time. I then move to the level of the user, considering how each individual's linguistic behaviour changes throughout their careers. This is followed by a discussion, bringing these two angles together and commenting on points where the results align and differ from expectations.

7.4.2.1 Observations by feature

7.4.2.1.1 Mayfly features

7.4.2.1.1.1 Verb + that-clause: certainty ($r = -0.4471$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	105	103	98	108	72	67	92	95	87	87
2	74	81	74	54	74	135	101	94	88	81
3	127	102	111	109	118	120	155	132	95	125
4	218	156	102	116	27	95	122	88	75	75
5	53	39	39	62	77	71	39	71	59	68

Table 23: Relative frequency per million words of verb + that-clause: certainty per user, per decile

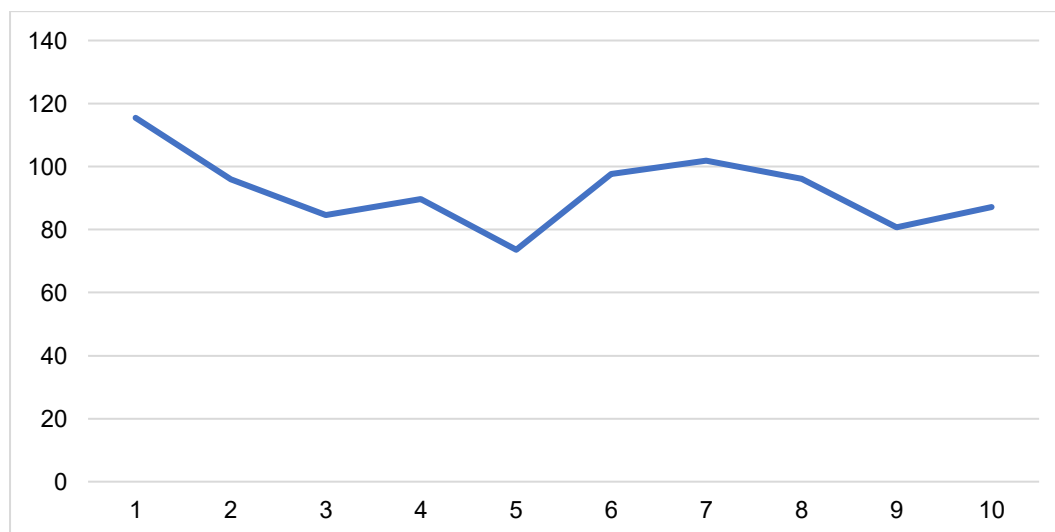


Figure 12: Average usage (relative frequency per million words) of verb + that-clause: certainty across all five users' careers

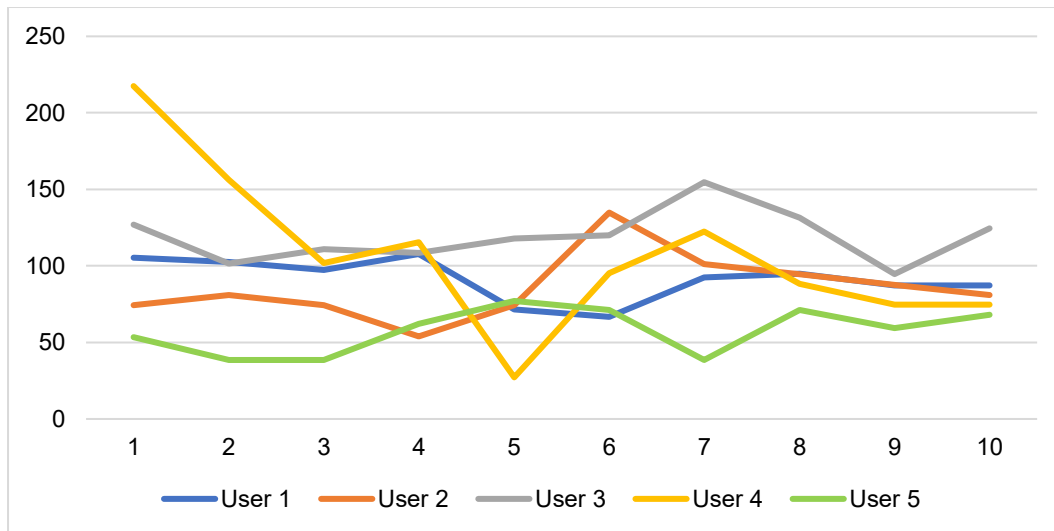


Figure 13: Usage (relative frequency per million words) of verb + that-clause: certainty by all five users across their careers

On average, usage of the feature *verb + that-clause: certainty* generally follows the expected pattern of decreasing usage as the career goes on, showing a general downwards trend throughout the career. However, there is an unexpected secondary peak of usage in the latter half of the career. User 4 is something of an outlier among the five users, showing higher usage in the early deciles and impacting the average usage for this feature in those deciles. Indeed, if User 4 is excluded, the trend across the remaining four users is much flatter throughout, with no general downwards trend. On the individual level, only Users 1 and 4 follow the expected usage pattern.

7.4.2.1.1.2 *I + verb + that-clause: certainty* ($r = -0.6353$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	28.23	28.23	12.83	20.53	23.1	15.4	15.4	15.4	10.27	20.53
2	47.2	47.2	47.2	20.23	47.2	74.18	33.72	40.46	60.69	47.2
3	53.11	34.64	43.87	34.64	43.87	36.95	48.49	41.57	18.47	32.33
4	54.38	33.99	47.58	33.99	0	40.78	47.58	27.19	20.39	6.8
5	8.9	8.9	0	17.79	11.86	8.9	2.97	20.76	5.93	8.9

Table 24: Relative frequency per million words of *I + verb + that-clause: certainty* per user, per decile

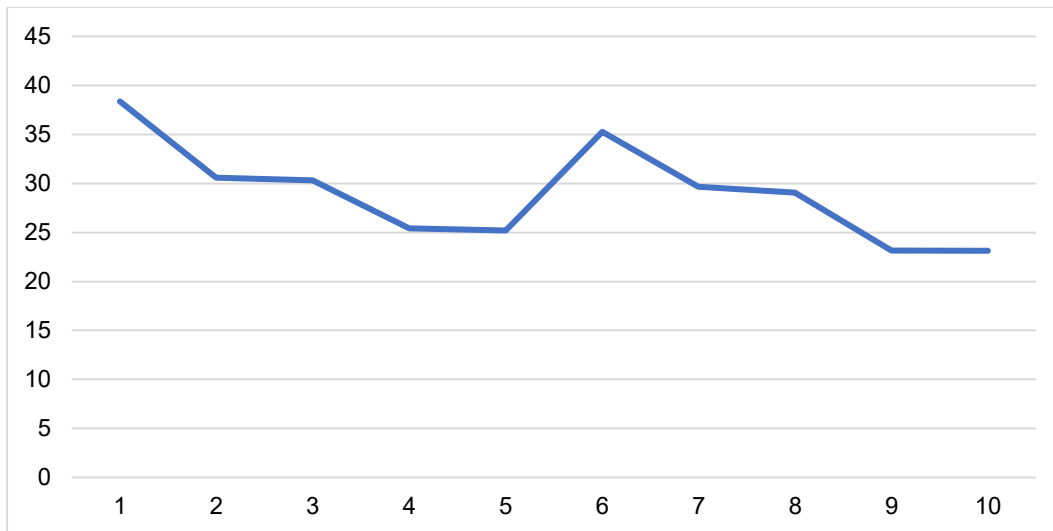


Figure 14: Average usage (relative frequency per million words) of *I + verb + that-clause: certainty* across all five users' careers

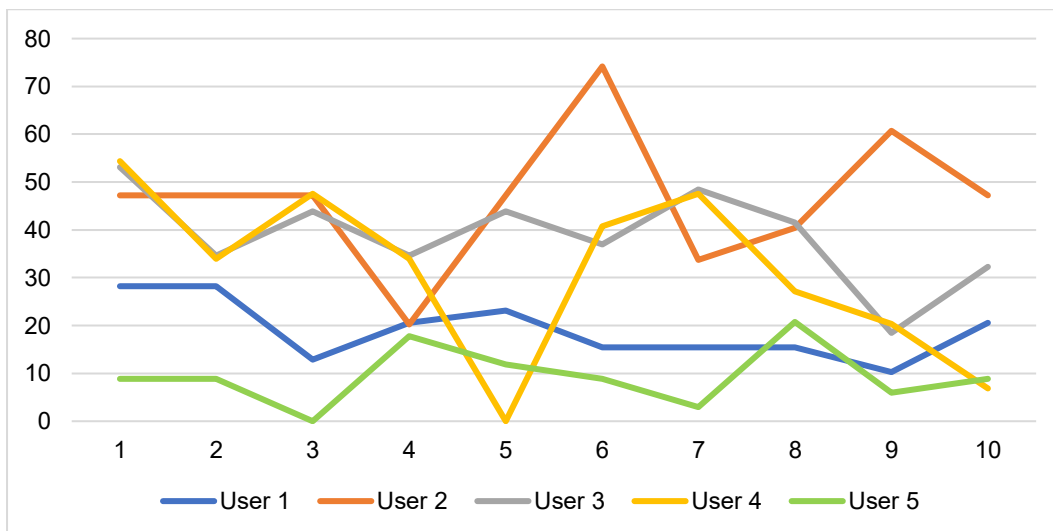


Figure 15: Usage (relative frequency per million words) of *I + verb + that-clause: certainty* by all five users across their careers

Much like the parent feature *verb + that-clause: certainty*, the first person iteration of the feature generally follows the expected downwards trend over time, albeit with a secondary peak in the later career. Unlike the parent feature, there is no clear outlier among the five users. However, most of these users do not clearly follow the expected pattern, except for User 1 and, to a lesser extent, User 3. This is unsurprising following the results shown above for *verb + that-clause: certainty*, despite the stronger overall correlation with career length found for this feature (§7.3.2).

7.4.2.1.1.3 Noun + that-clause: certainty ($r = -0.5364$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	15.4	15.4	23.1	38.49	17.96	23.1	23.1	25.66	17.96	10.27
2	6.74	0	0	6.74	20.23	13.49	26.97	13.49	6.74	0
3	11.55	20.78	23.09	13.86	23.09	16.16	20.78	4.62	18.47	11.55
4	6.8	6.8	6.8	0	0	0	6.8	6.8	0	0
5	2.97	11.86	2.97	11.86	17.79	20.76	26.69	17.79	8.9	11.86

Table 25: Relative frequency per million words of noun + that-clause: certainty per user, per decile

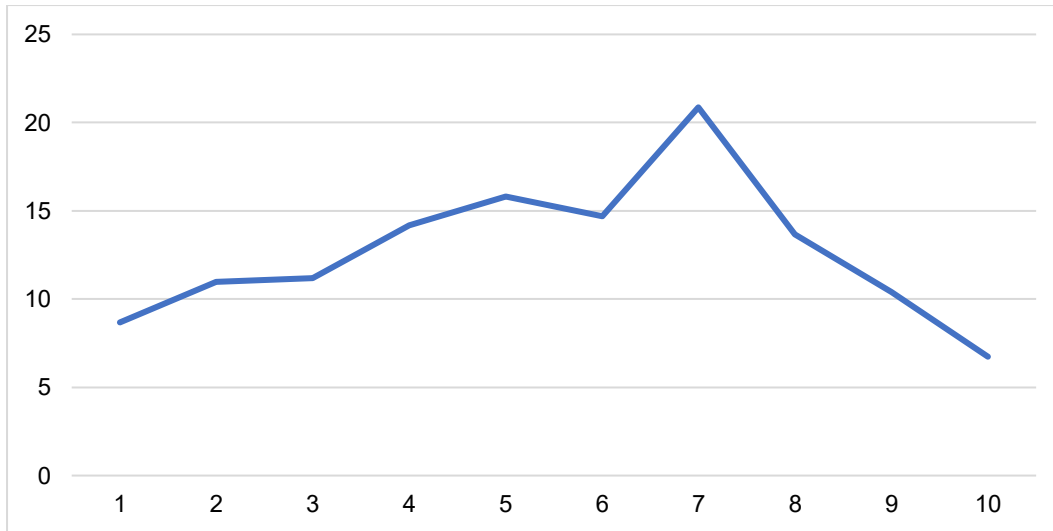


Figure 16: Average usage (relative frequency per million words) of noun + that-clause: certainty across all five users' careers

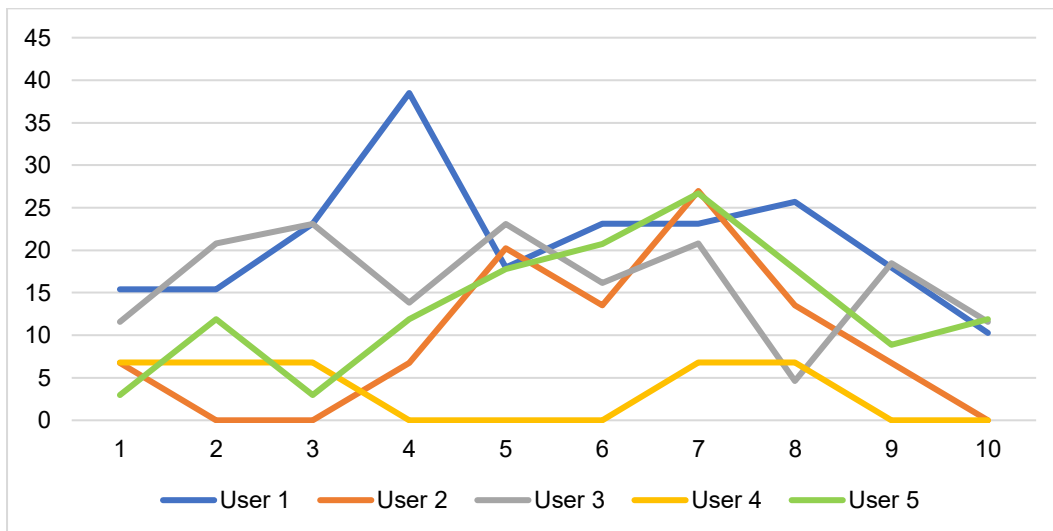


Figure 17: Usage (relative frequency per million words) of noun + that-clause: certainty by all five users across their careers

The average usage across the five users clearly does not follow the expected pattern here; rather than a downwards trend across the career, usage is dominated by a peak in the mid-late career. However, usage of this feature is fairly low across all five users, with the peak

average usage at only 21 tokens per million words, so no patterns relating to this feature are strong. User 3 comes closest to following the expected pattern, although the downwards trend is weak. Users 2 and 5 show the opposite pattern, with usage generally rising through their careers.

7.4.2.1.1.4 Noun + that-clause: likelihood ($r = -0.4286$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	12.83	10.27	10.27	28.23	12.83	7.7	5.13	15.4	17.96	20.53
2	0	6.74	0	0	13.49	0	26.97	13.49	6.74	0
3	6.93	18.47	18.47	23.09	34.64	2.31	16.16	30.02	13.86	13.86
4	6.8	0	6.8	20.39	6.8	6.8	0	0	0	0
5	2.97	5.93	8.9	5.93	0	14.83	5.93	8.9	2.97	0

Table 26: Relative frequency per million words of noun + that-clause: likelihood per user, per decile

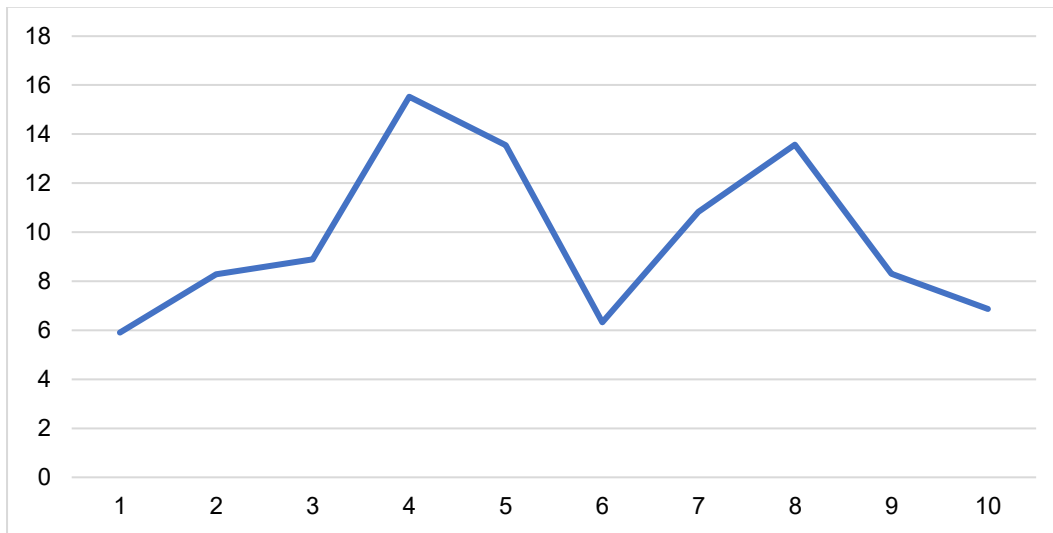


Figure 18: Average usage (relative frequency per million words) of noun + that-clause: likelihood across all five users' careers

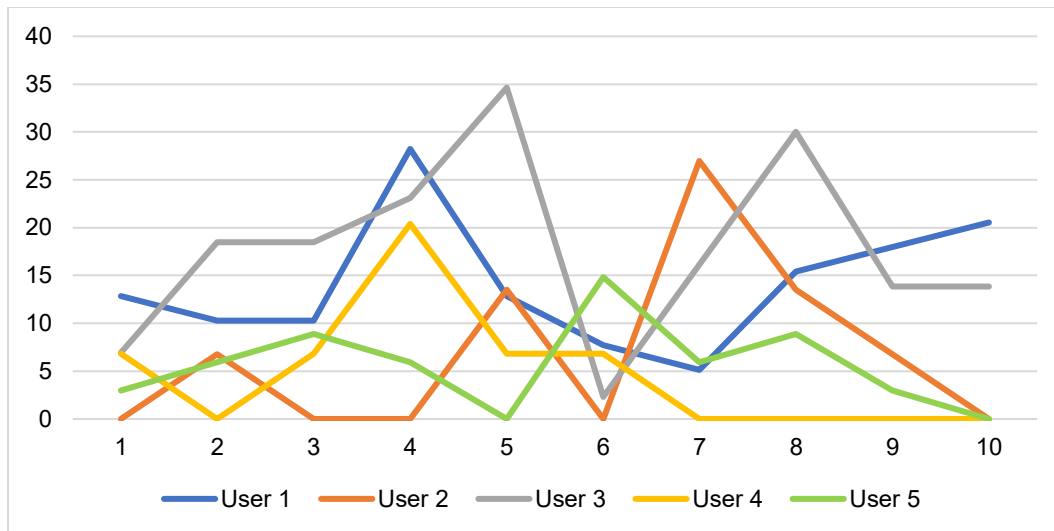


Figure 19: Usage (relative frequency per million words) of noun + that-clause: likelihood by all five users across their careers

Similar to the other *noun + that-clause* construction, this feature does not follow the expected downwards trend over time. Indeed, the average usage does not show a clear pattern of any kind, with notable fluctuations in both directions throughout the career. Again, however, usage of this feature is low across all five users, making it difficult to draw conclusions about its usage across the career. Within this low-usage range, each of the five users show distinct usage, with peaks of usage at different points; all five users converge in low usage at the sixth decile, but again with such low frequencies it is difficult to ascribe any significance to this. User 4 once again comes closest to showing the expected pattern, with barely any tokens in the latter half of their career.

7.4.2.1.1.5 First person singular pronouns ($r = -0.7$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	1,250	1,314	1,383	1,229	1,404	1,299	1,222	1,193	1,422	1,481
2	3,729	3,891	3,898	2,198	2,711	2,927	2,448	3,736	3,918	3,116
3	2,709	2,635	2,542	2,330	2,459	2,355	2,129	1,838	1,358	1,342
4	2,039	3,419	2,094	2,229	1,964	2,087	2,039	1,971	2,080	1,849
5	2,443	2,479	1,942	1,560	1,240	1,231	1,320	934	1,477	1,254

Table 27: Relative frequency per million words of first person singular pronouns per user, per decile

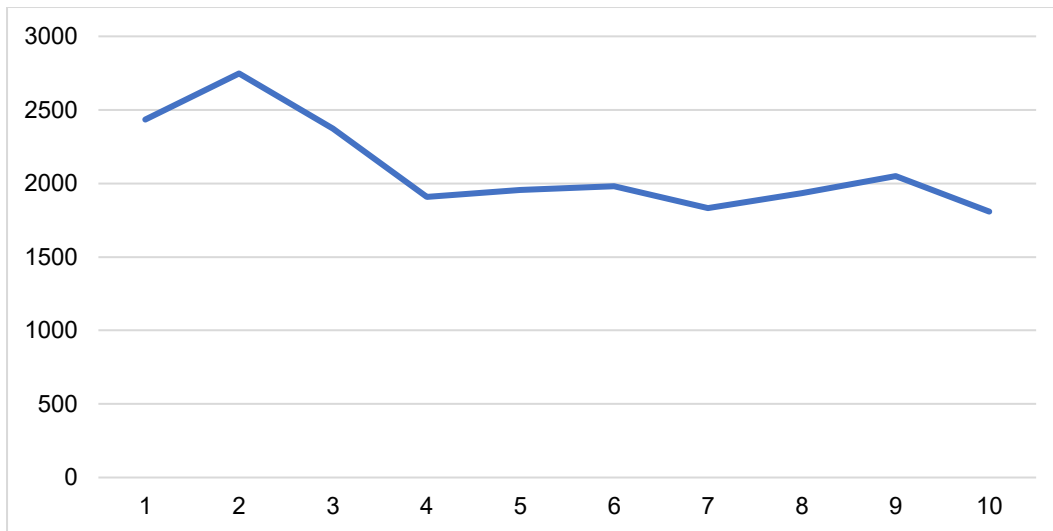


Figure 20: Average usage (relative frequency per million words) of first person singular pronouns across all five users' careers

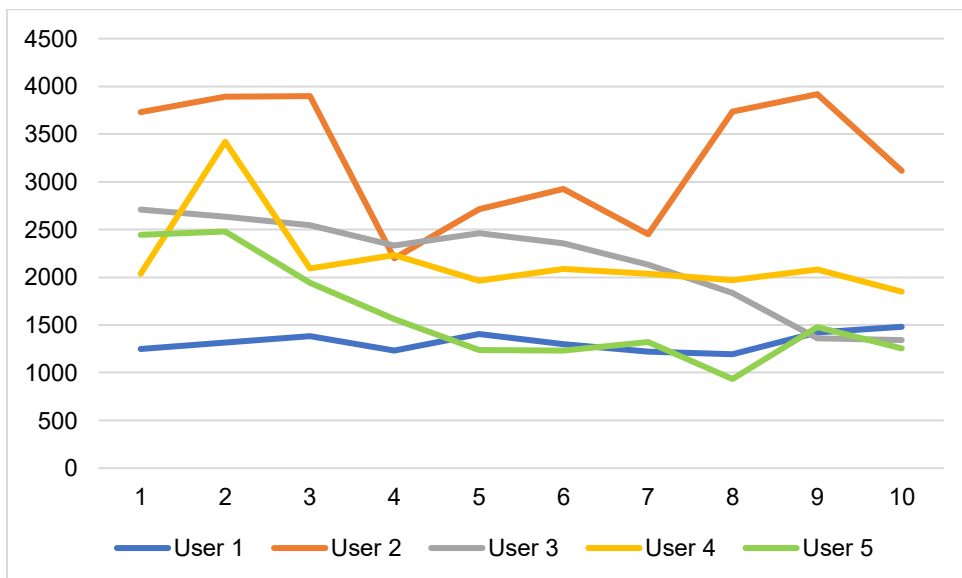


Figure 21: Usage (relative frequency per million words) of first person singular pronouns by all five users across their careers

Usage of this feature clearly follows the expected pattern, with a clear downwards trend throughout the career, reflecting the strong correlation identified between first person singular pronoun usage and career length. On the individual level, most users (Users 3, 4, and 5) also clearly follow this pattern, although User 1 shows the opposite pattern (while usage is generally consistent throughout the career, there is a very slight upwards trend throughout the career).

7.4.2.1.1.6 Second person pronouns ($r = -1$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	711	831	482	403	403	565	603	644	508	516
2	802	1,045	614	789	722	944	762	600	573	809
3	1,313	755	1,073	1,097	903	884	928	1,226	880	1,101
4	1,291	843	653	591	761	931	918	822	822	720
5	1,014	934	833	771	807	599	593	706	780	697

Table 28: Relative frequency per million words of second person pronouns per user, per decile

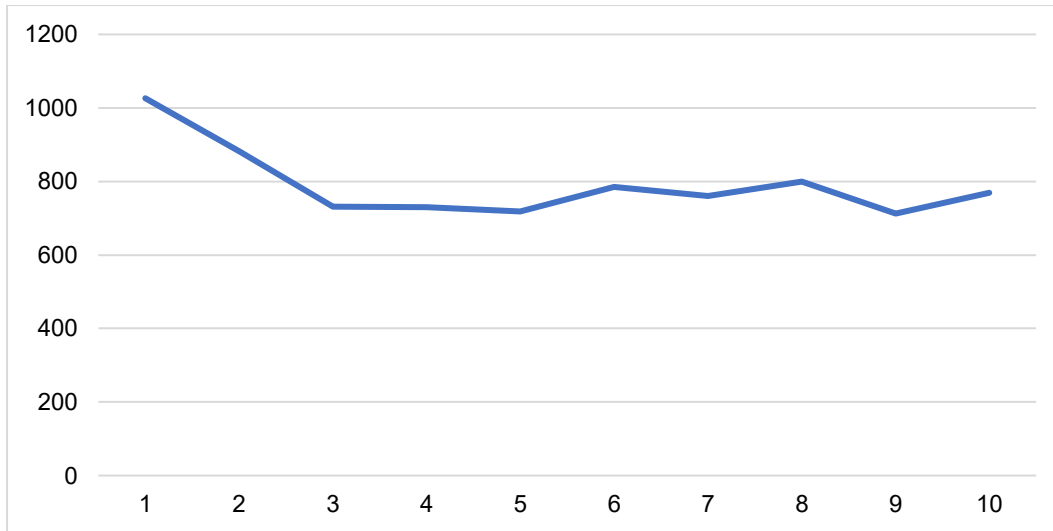


Figure 22: Average usage (relative frequency per million words) of second person pronouns across all five users' careers

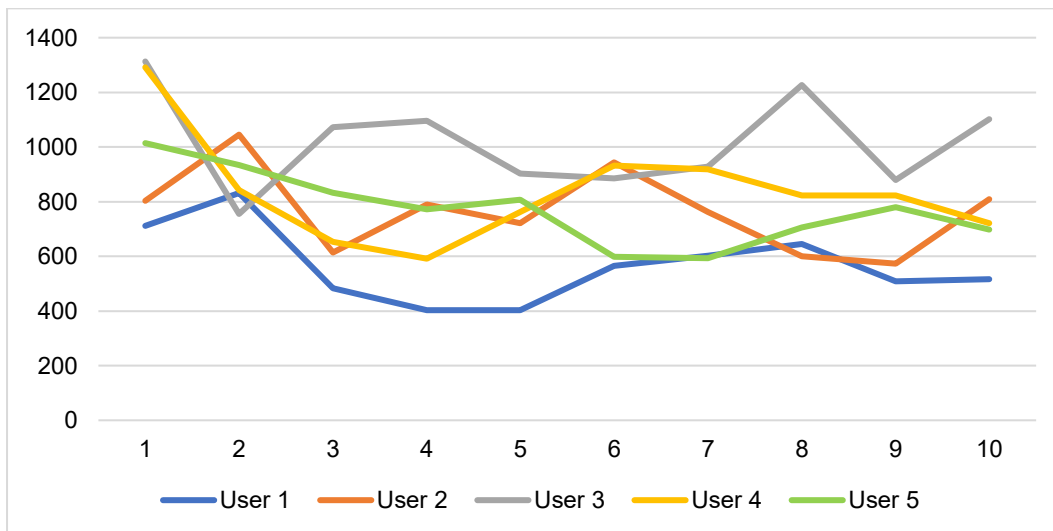


Figure 23: Usage (relative frequency per million words) of second person pronouns by all five users across their careers

Like the other pronominal mayfly feature, second person pronoun usage follows the expected downwards trend throughout the career, decreasing from a clear peak in the first decile. This reflects the feature's very strong correlation with career length identified

previously. On the individual level, all users except User 3 broadly follow the expected pattern here, although some of these (particularly Users 1 and 4) show a sharp initial decrease in usage followed by a slight increase in the mid-career. Nonetheless, the trend for most users is generally downwards as the career progresses.

7.4.2.1.2 Cicada features

7.4.2.1.2.1 Modals and semi-modals: possibility, permission, and ability ($r = 0.7$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	693	690	562	603	557	541	583	649	783	662
2	519	344	371	303	364	398	371	276	391	324
3	427	395	473	515	483	466	487	390	374	365
4	585	571	619	551	544	530	673	625	571	632
5	721	718	727	682	676	741	569	661	620	667

Table 29: Relative frequency per million words of modals and semi modals: possibility, permission, and ability per user, per decile

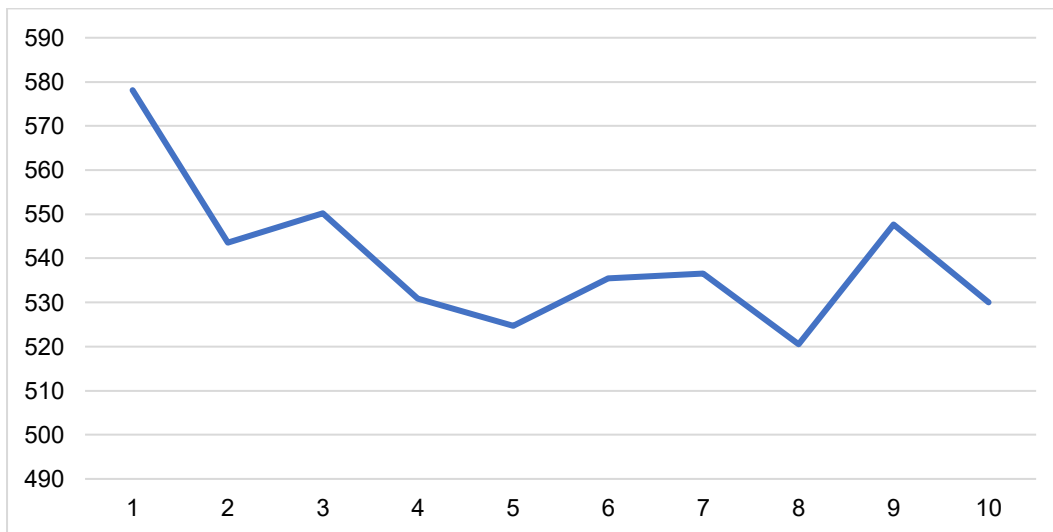


Figure 24: Average usage (relative frequency per million words) of modals and semi-modals: possibility, permission, and ability across all five users' careers

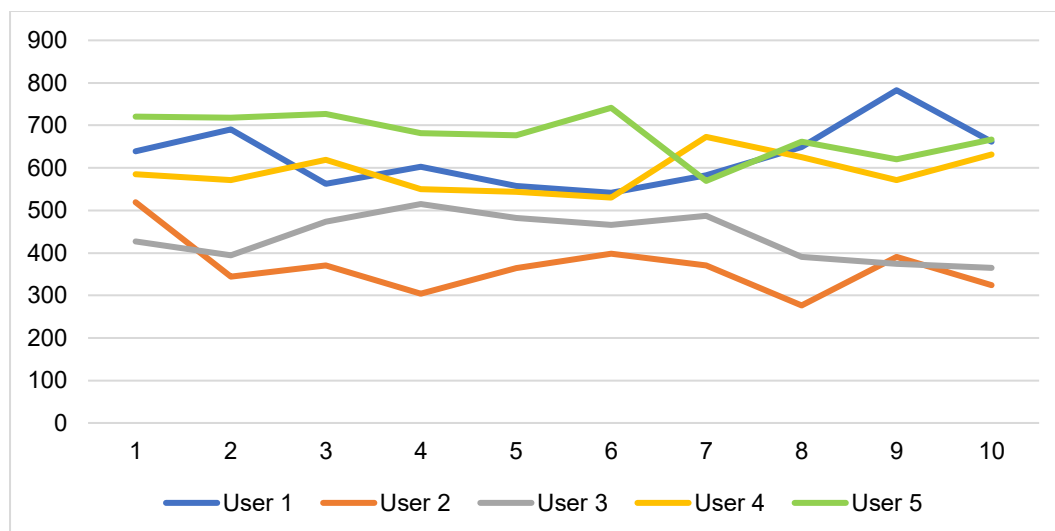


Figure 25: Usage (relative frequency per million words) of modals and semi-modals: possibility, permission, and ability by all five users across their careers

The use of modals and semi-modals of possibility, permission, and ability does not follow the expected pattern of increased usage as the career progresses. This is somewhat surprising given the strong positive correlation identified between this feature and career length. Figure 24 in fact shows the opposite pattern of a general downward trend, albeit with fluctuations throughout. However, taking the scale of the y-axis into account, we see that the average rate of usage remains fairly stable throughout the career, showing only a small range from 578 tokens per million words at its peak to a minimum of 521. This is reflected at the individual level, where all five users show broadly similar rates throughout their respective careers. While Users 2 and 5 follow the overall pattern showing a slight decrease in usage over time, Users 1 and 4 stand in contrast to this, following the expected pattern of a slight upward trend through the career.

7.4.2.1.2.2 Modals and semi-modals: prediction and volition ($r = 0.4$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	1,003	1,050	1,178	1,119	1,037	962	1,009	1,021	1,196	1,044
2	641	614	647	539	647	769	789	829	647	573
3	949	1,071	1,035	991	956	921	963	806	674	727
4	1,149	999	999	958	945	1,013	1,047	1,006	999	938
5	958	691	1,251	1,103	1,105	925	940	904	943	958

Table 30: Relative frequency per million words of modals and semi modals: prediction and volition per user, per decile

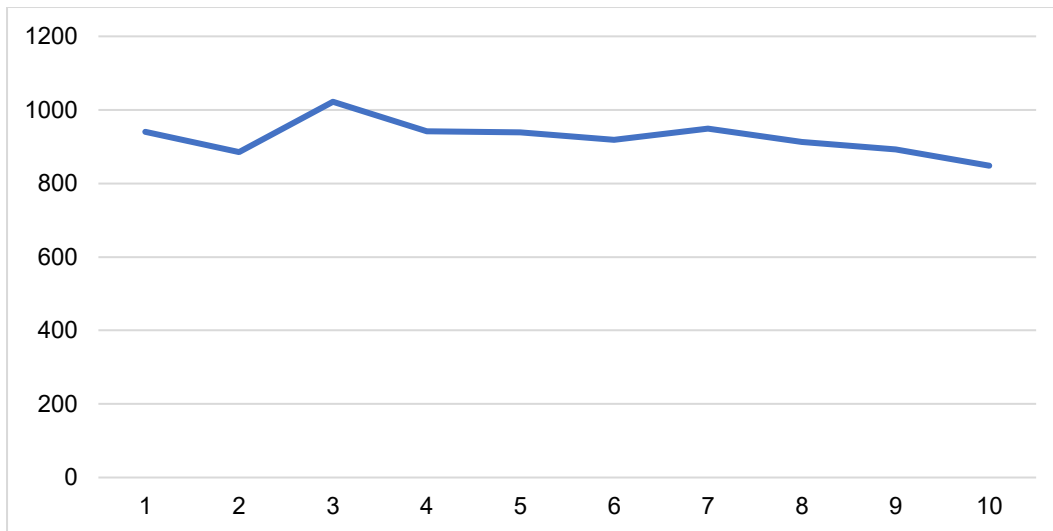


Figure 26: Average usage (relative frequency per million words) of modals and semi-modals: prediction and volition across all five users' careers

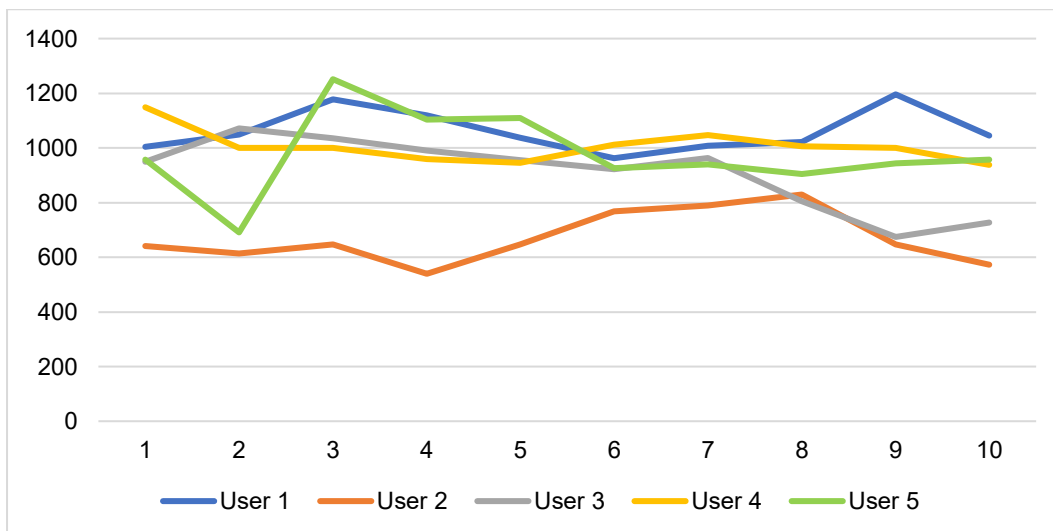


Figure 27: Usage (relative frequency per million words) of modals and semi-modals: prediction and volition by all five users across their careers

Usage of this feature is broadly consistent across the career, although there is a small peak at the third decile and a slight downward trend from that point onwards, which is contrary to expectations for this feature. This feature only shows a moderate correlation with career length, which may account for this unexpected result. Unsurprisingly, given the overall pattern, most of the five users show generally consistent rates across their careers. User 3 is contrary to this, showing a somewhat more marked downwards trend. No user clearly shows the expected upward trend over time.

7.4.2.1.2.3 Adjective + that-clause: likelihood ($r = 0.4384$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	2.57	0	0
2	0	0	0	0	0	0	0	0	0	0
3	4.62	2.31	4.62	4.62	0	2.31	0	2.31	0	2.31
4	0	0	0	0	0	0	0	0	0	0
5	2.97	0	0	8.9	0	2.97	5.93	11.86	2.97	5.93

Table 31: Relative frequency per million words of adjective + that-clause: likelihood per user, per decile

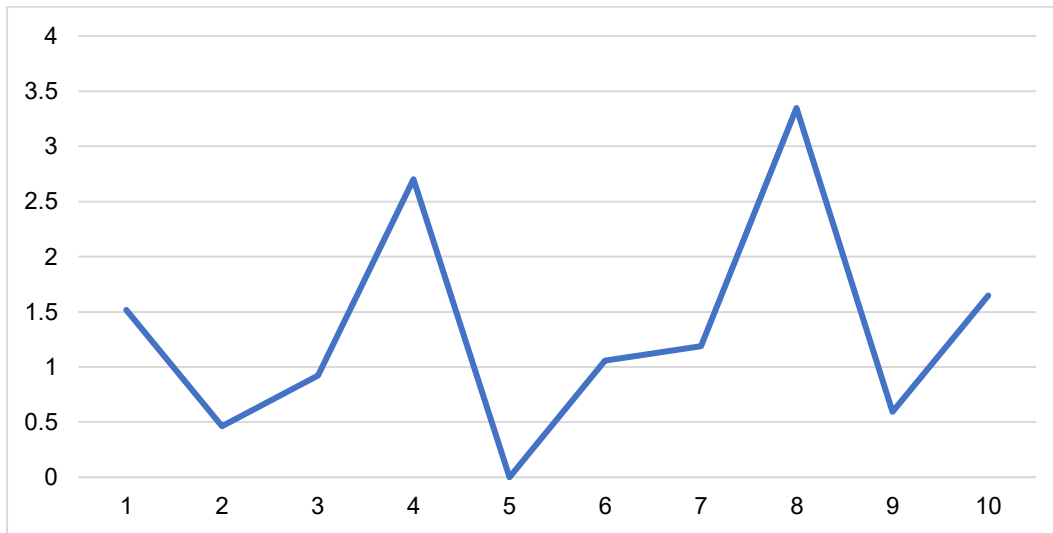


Figure 28: Average usage (relative frequency per million words) of adjective + that-clause: likelihood across all five users' careers

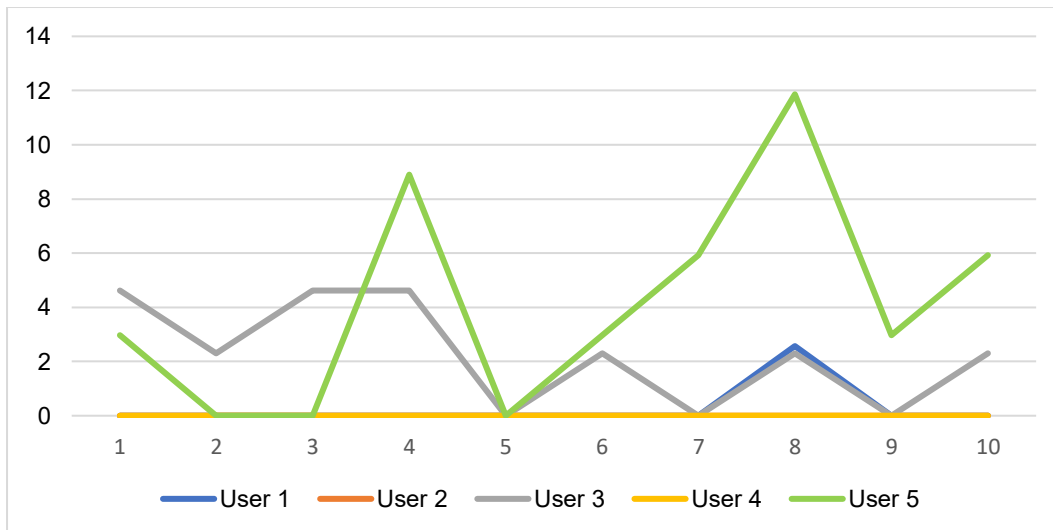


Figure 29: Usage (relative frequency per million words) of adjective + that-clause: likelihood by all five users across their careers

This feature is a very low frequency feature among the five users, with a peak average relative frequency of just 3.35 tokens per million words, making it very difficult to draw general conclusions about the usage of this feature over time. Unsurprisingly, the overall

pattern largely reflects the usage of User 5, the most frequent user of this feature; even this user shows no tokens in three out of ten career deciles. Two users, Users 2 and 4, do not use this construction at all throughout their career.

7.4.2.1.2.4 Adjective + to-clause: certainty ($r = 0.9333$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	0	0	2.57	5.13	2.57	7.7	10.27	2.57	2.57	7.7
2	0	6.74	13.49	0	6.74	0	6.74	0	0	0
3	0	0	6.93	0	2.31	4.62	2.31	0	2.31	4.62
4	0	0	0	0	0	0	0	0	0	0
5	2.97	5.93	0	2.97	0	0	0	2.97	2.97	0

Table 32: Relative frequency per million words of adjective + to-clause: certainty per user, per decile

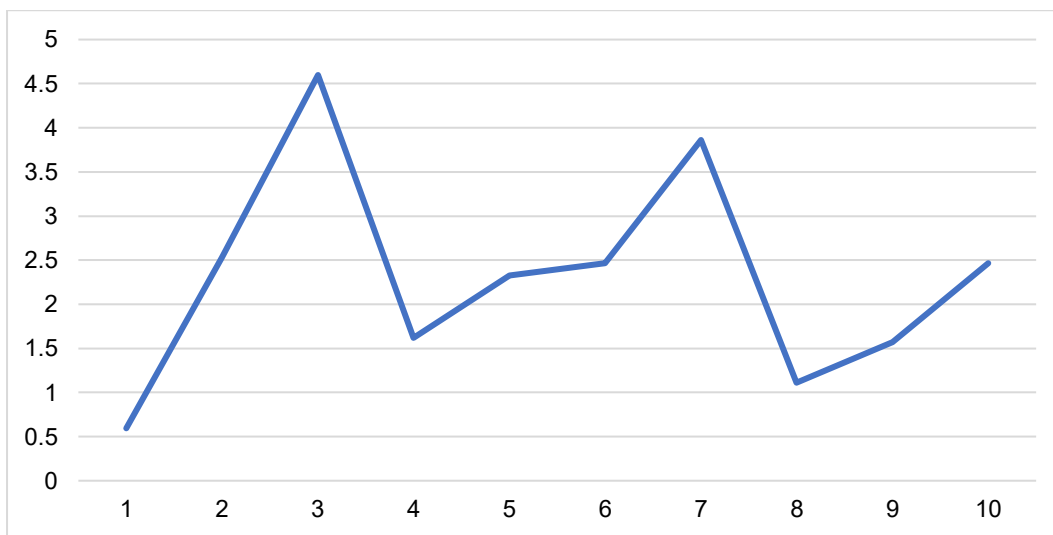


Figure 30: Average usage (relative frequency per million words) of adjective + to-clause: certainty across all five users' careers

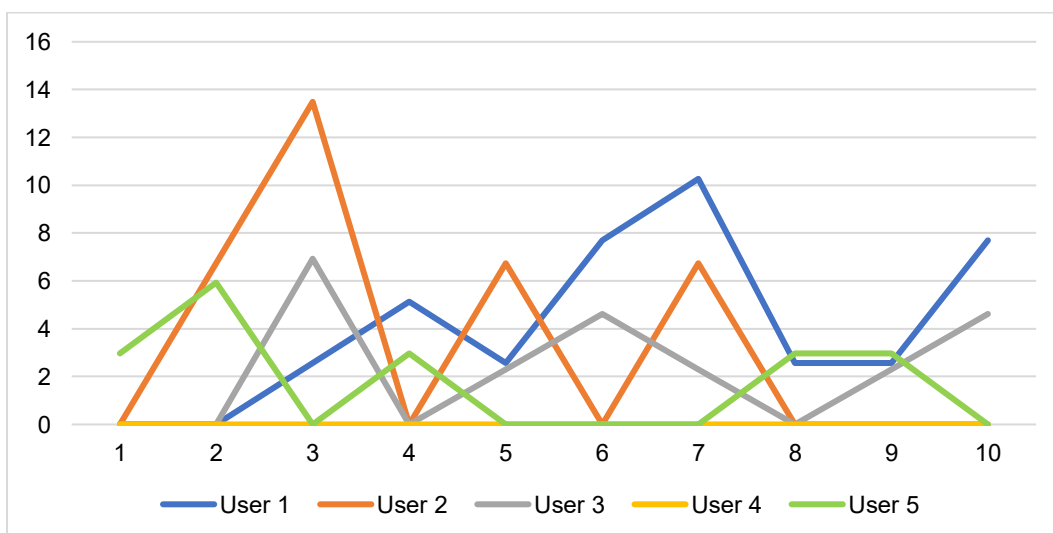


Figure 31: Usage (relative frequency per million words) of adjective + to-clause: certainty by all five users across their careers

Much like the other adjectival stance feature, *adjective + to-clause: certainty* is a very low-frequency feature among all five users, making it difficult to draw general conclusions. Unlike *adjective + that-clause: likelihood*, in the present case this is despite the feature showing a very strong correlation with career length. Within the low frequency range, each of the five users displays a distinct pattern of usage.

7.4.2.1.2.5 First person plural pronouns ($r = 0.7$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	1,037	1,340	1,029	985	829	1,147	1,609	1,632	1,563	1,486
2	357	378	452	722	708	553	337	196	378	492
3	903	974	887	794	695	681	1,048	630	533	665
4	2,454	1,618	1,495	1,074	904	999	727	884	945	727
5	727	744	433	566	534	436	463	498	418	466

Table 33: Relative frequency per million words of first person plural pronouns per user, per decile

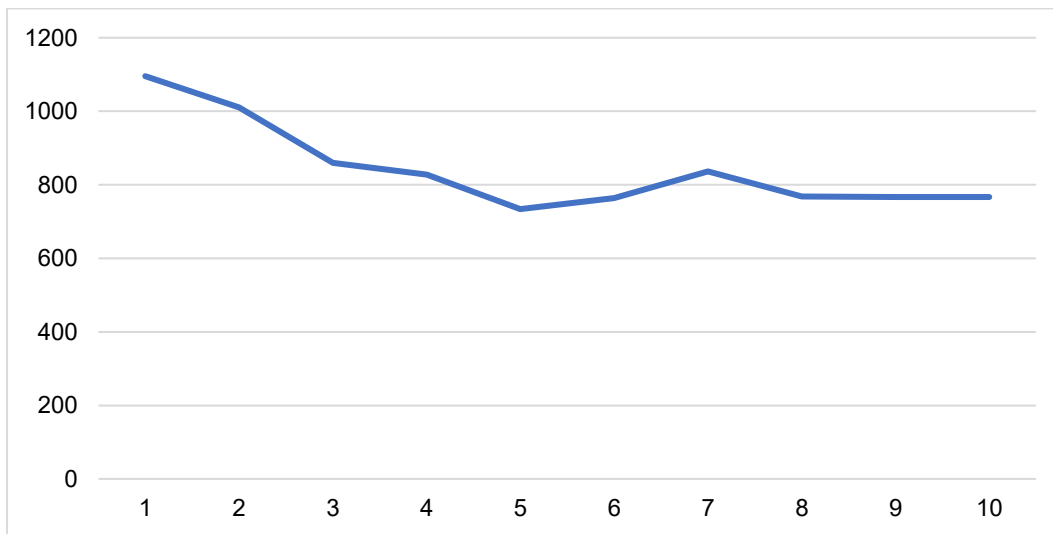


Figure 32: Average usage (relative frequency per million words) of first person plural pronouns across all five users' careers

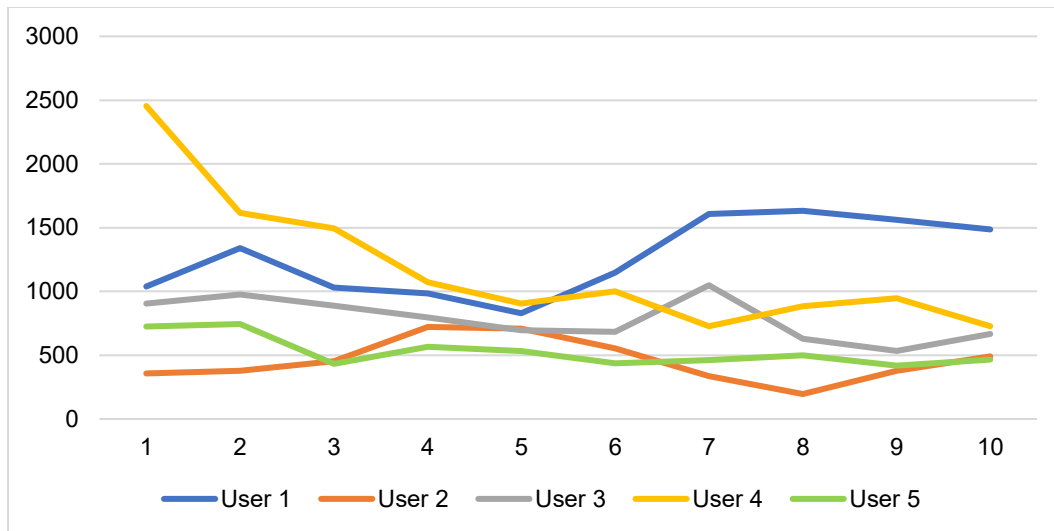


Figure 33: Usage (relative frequency per million words) of first person plural pronouns by all five users across their careers

Like the other pronominal features included in this study, but in this case contrary to expectations, use of first person plural pronouns on average decreases over time. This is surprising considering the strong positive correlation between this feature and career length identified previously. On the individual level, only one user, User 4, shows a marked decrease in the usage of these pronouns over time, with an extremely high peak at the first decile compared with the other four users. While other users – Users 3 and 5 – also show this general decrease in use throughout the career, this pattern is not so marked as for User 4. Despite the general pattern identified from the five users, User 1 nonetheless shows the expected pattern, with a general increase in usage of these pronouns as the career goes on.

7.4.2.1.2.6 Aspectual verbs: progression ($r = 0.52$)

User	Decile									
	1	2	3	4	5	6	7	8	9	10
1	38.49	35.93	30.8	41.08	38.49	20.53	48.76	51.33	35.93	41.06
2	0	6.74	6.74	33.72	6.74	33.72	0	26.97	6.74	20.23
3	32.33	20.78	39.26	25.4	11.55	18.47	9.24	16.16	16.16	18.47
4	6.8	0	33.99	20.39	20.39	13.59	6.8	20.39	20.39	27.19
5	5.93	17.79	8.9	5.93	23.72	2.97	0	11.86	14.83	8.9

Table 34: Relative frequency per million words of aspectual verbs: progression per user, per decile

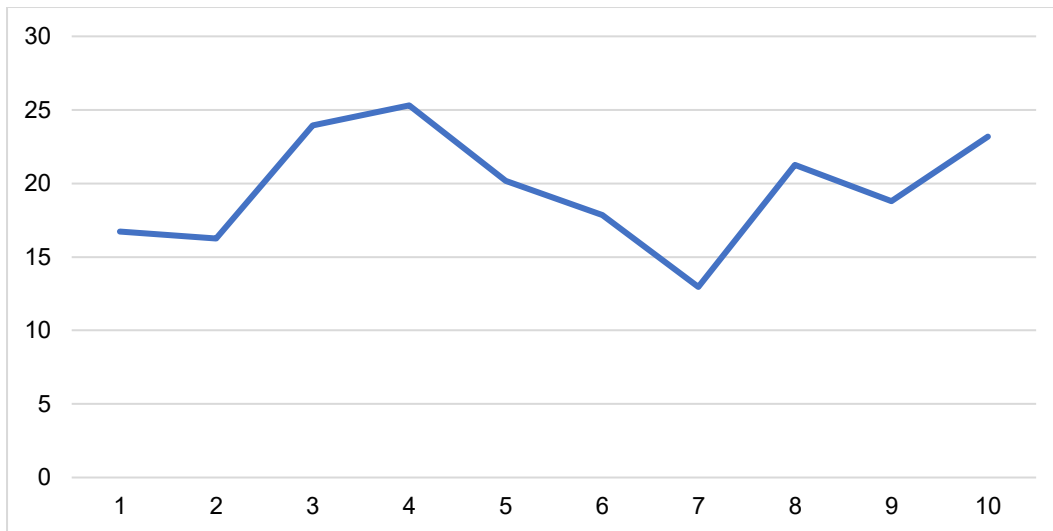


Figure 34: Average usage (relative frequency per million words) of aspectual verbs: progression across all five users' careers

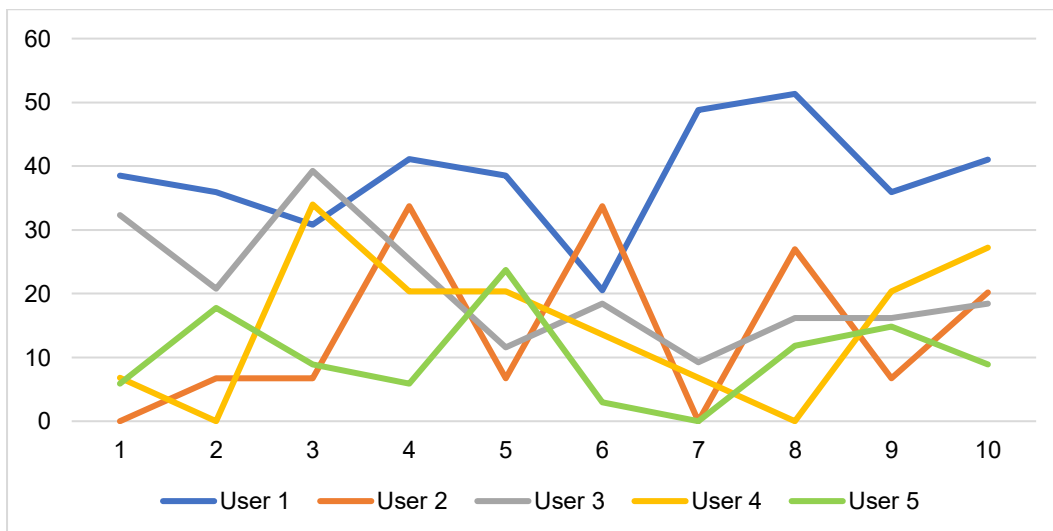


Figure 35: Usage (relative frequency per million words) of aspectual verbs: progression by all five users across their careers

Unlike most of the other cicada features outlined here, the *aspectual verbs: progression* feature aligns with expectations insofar as the rate of usage at the end of the career is higher than that at the beginning. However, the average usage across time is marked by fluctuation, rather than a clear increase throughout the career. Interpretation of the pattern shown for this feature with any certainty is difficult as it is a relatively low-frequency feature across these five users. As Figure 35 shows, within this relatively low range of frequencies each user exhibits a different pattern of usage across the career. Only User 1 approaches the expected clear increase in usage of this feature over time, although there is some fluctuation throughout the career. On the other hand, User 3 shows a relatively smooth downwards trend in usage, the opposite of the expected pattern.

7.4.2.2 Observations by user

The following sections in part reproduce the information presented above, but this time organised by user. To ensure the graphs are readable, two graphs are presented for each user, one showing usage over time of higher-frequency features, the other of lower-frequency features. On both graphs, mayfly features are represented by blue lines, and cicada features by orange lines. This is to help the reader identify at a glance those features of which usage is expected to increase or decrease over time, to allow easier comparison with the user's actual behaviour. For reasons of space, the findings for each user will begin, in each case, on the next full page.

7.4.2.2.1 User 1

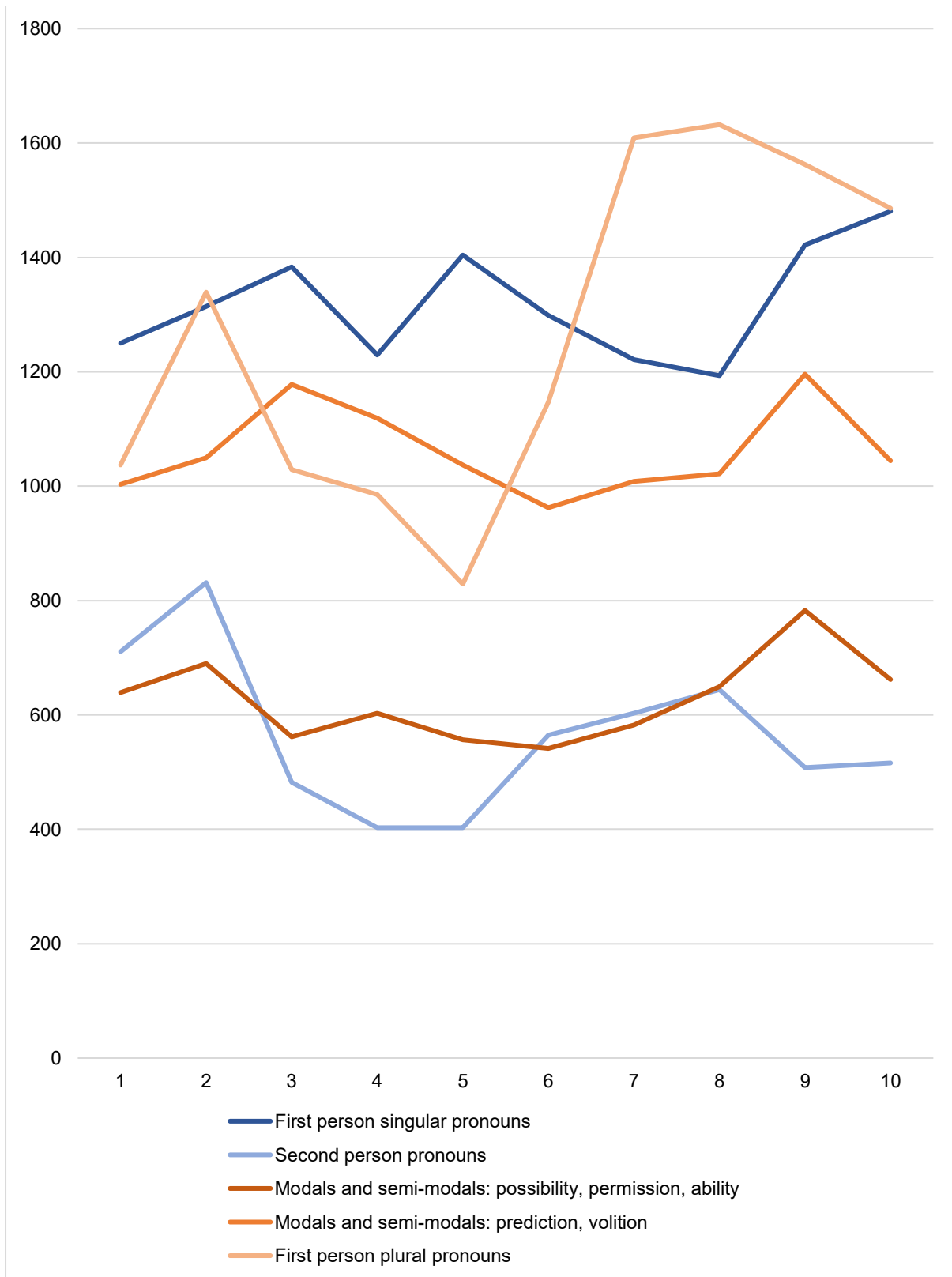


Figure 36: Use of higher-frequency features (pmw) across User 1's career. Mayfly features are represented in blue, and cicada features in orange

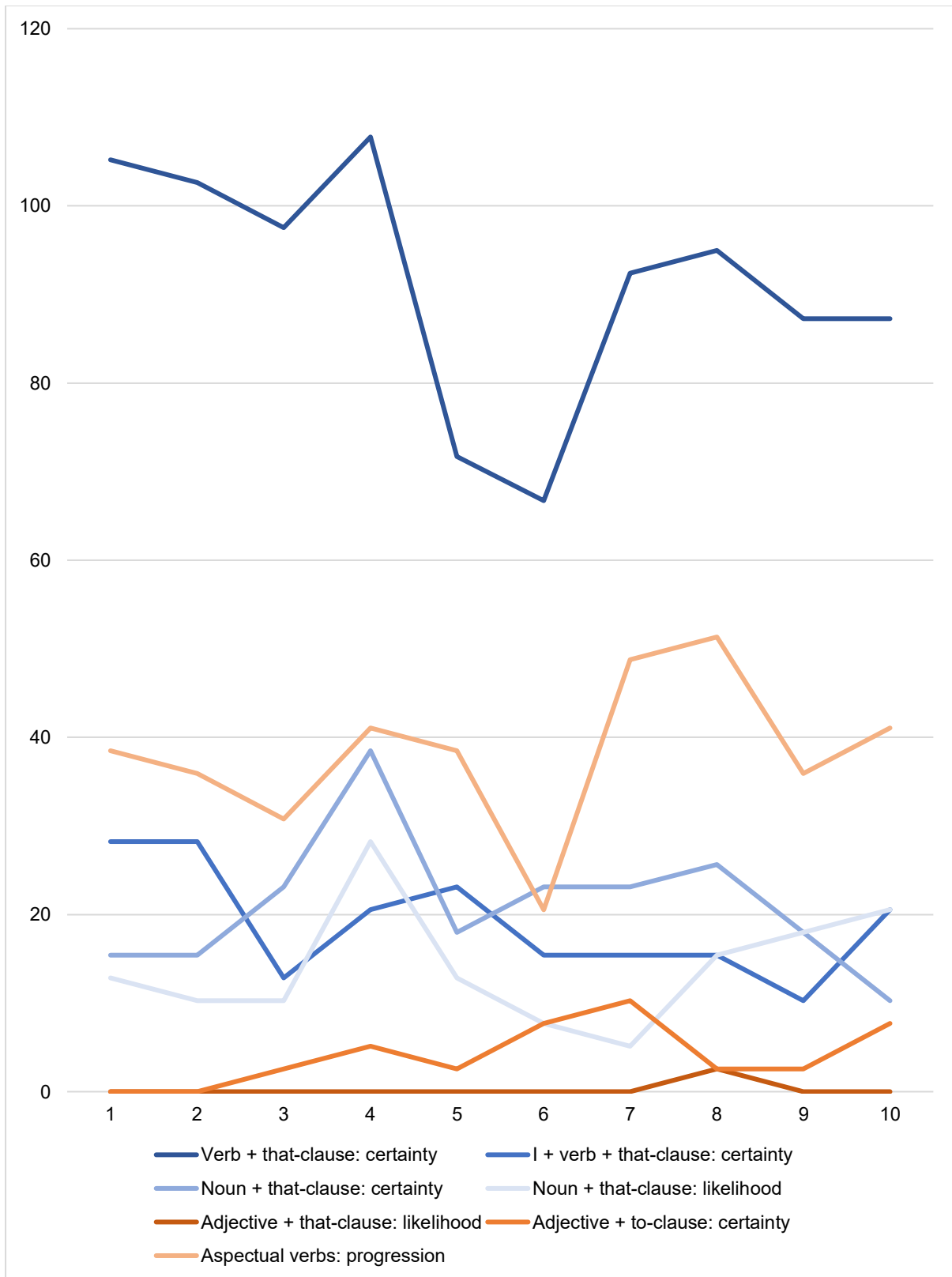


Figure 37: Use of lower-frequency features (pmw) across User 1's career. Mayfly features are represented in blue, and cicada features in orange

Of all the users investigated here, User 1 conforms the most to expected patterns, following expectations to a greater or lesser extent for seven of the twelve features (*verb + that-clause: certainty*, *I + verb + that-clause: certainty*, *second person pronouns*, *modals and semi-modals: possibility, permission, ability*, *adjective + to-clause: certainty*, *first person plural pronouns*, and *aspectual verbs: progression*). Note that for the feature *adjective + to-clause: certainty*, only a small number of tokens (n = 16) are identified across the entire career; while not a strong finding by itself, this nonetheless contributes to the wider finding that User 1 is the most conformist of the five users. Uniquely among the five users, there are no features for which User 1 displays the opposite of the expected pattern (i.e. a noticeably downward trend for a cicada feature, or vice versa). Despite the general trends of usage for these features following the expected downward and upward trends respectively, User 1 does not show a straightforward linear pattern of usage across the career for any of these features. The graphs for all these features reveal some kind of mid-career fluctuation, or, in the case of the modals and semi-modals of possibility, permission, and ability, a U-shaped usage pattern.

User 1 shows a preference for the use of aspectual verbs of progression, showing the highest rate of all the users at eight of the ten deciles (Figure 35, §7.4.2.1.2.6). As noted above, this pattern of usage also follows the expected pattern of increase over time; User 1 is the only user to exhibit the expected pattern for this feature. This is also true for two of the other features, *adj + to-clause: certainty* and *first person plural pronouns*. This finding is particularly notable for the first person plural pronouns, as not only do the other four users *not* conform to the expected pattern, they in fact follow the opposite pattern (i.e. decreasing usage over time).

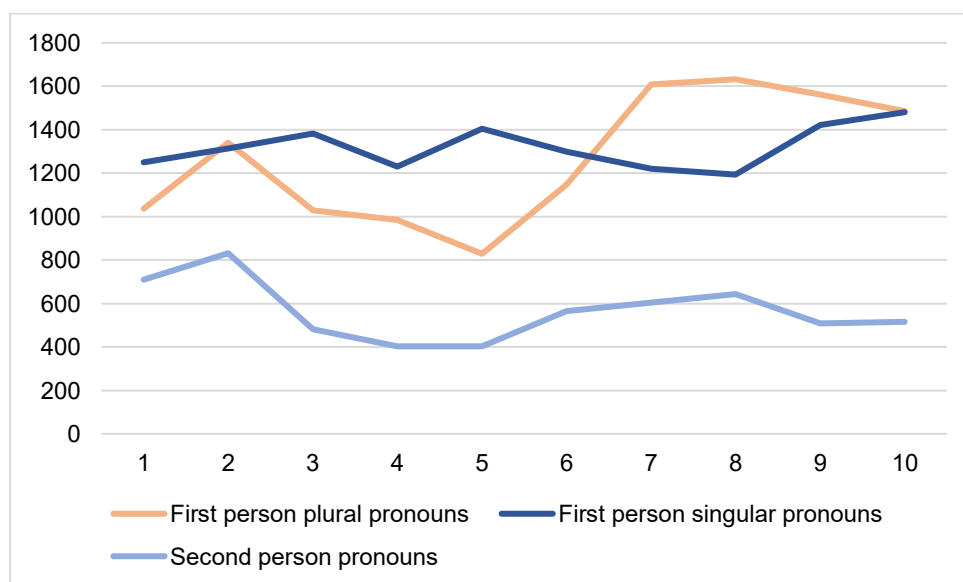


Figure 38: Usage of the three pronominal features (pmw) across User 1's career

User 1's usage of the three pronominal features is unusual among the five users in several ways. As Figures 41, 45, 48, and 51 below illustrate, for most users, first person singular is generally the most frequently used pronoun type throughout the career by some margin; first person plural and second person pronouns are typically used at a similar, lower rate. User 1 is the exception to this, with both the first person singular and plural pronouns showing higher usage relative to the second person. Given this, it is unsurprising that User 1 presents one of the lowest rates of first person singular pronouns overall (Figure 21, §7.4.2.1.1.5). Their usage is at a similar level to that of Users 3 and 5's late-career usage, but only after these users' respective decreases in usage of these pronouns, from the ninth and fifth deciles respectively. Although all users' usage of second person pronouns is relatively low compared to the other pronominal features, User 1 also presents the lowest usage of these pronouns in seven of the ten deciles (Figure 23, §7.4.2.1.1.6). On the other hand, User 1 is one of the most frequent users of first person plural pronouns throughout their career (Figure 33, §7.4.2.1.2.5). This may suggest that User 1 shows the greatest assimilation to a collective identity.

This hypothesis is also supported by the comparative use of first person singular and plural pronouns by User 1, which follows the expected pattern. The singular is more frequent in the early career, but is overtaken by the plural in the latter half of the career. This may indicate an increased sense of collective identity as User 1's career progresses. Notice, however, that the rates of usage of the two pronouns reconverge in the final decile. Perhaps this signals a diminished commitment to the group in the final stage of the career, forecasting the user's exit from the forum.

7.4.2.2.2 User 2

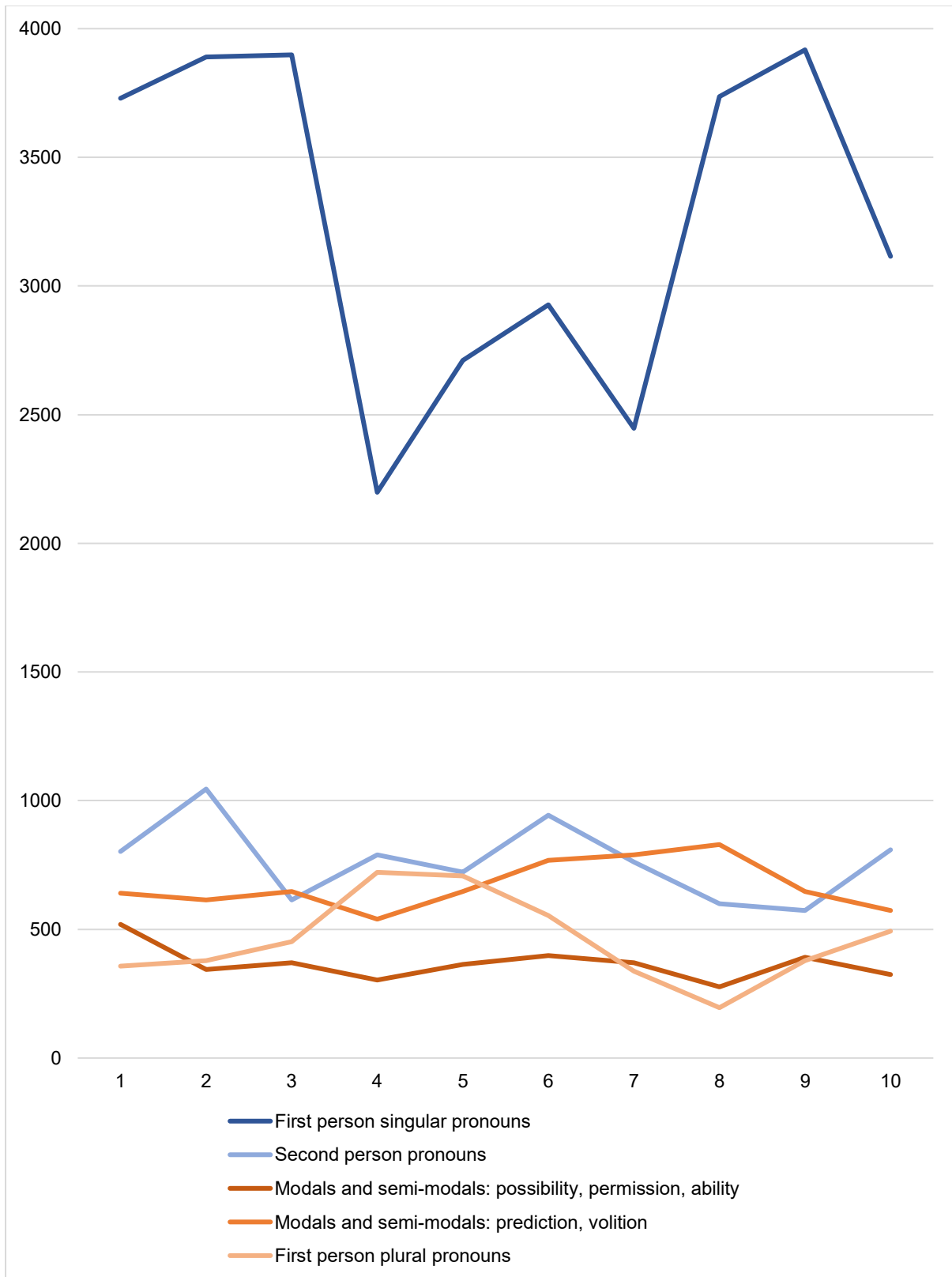


Figure 39: Use of higher-frequency features (pmw) across User 2's career. Mayfly features are represented in blue, and cicada features in orange

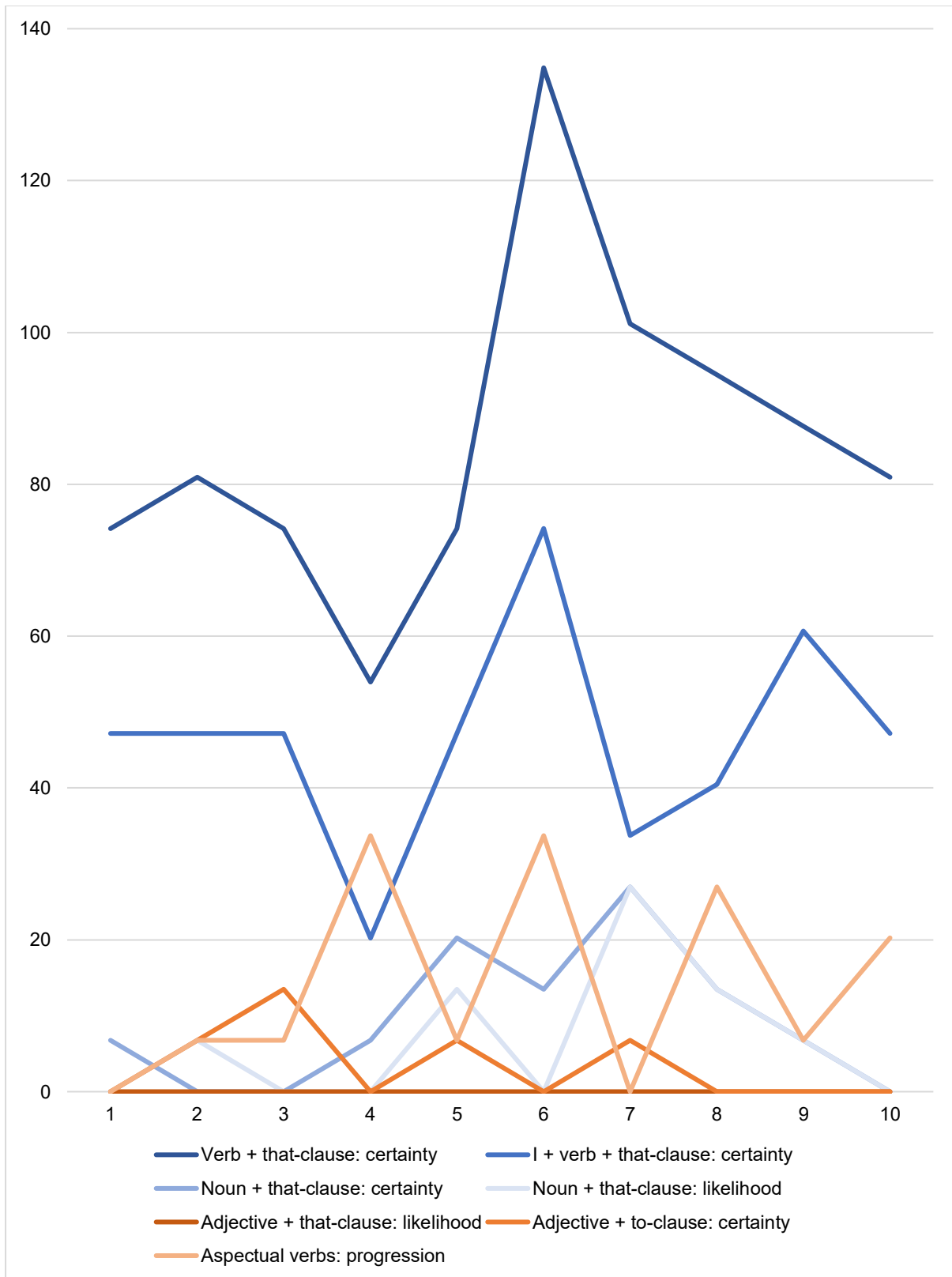


Figure 40: Use of lower-frequency features (pmw) across User 2's career. Mayfly features are represented in blue, and cicada features in orange

Following the user who conforms the *most* to expected patterns of usage, we now turn to the *least* conformist of the five users. User 2 follows the expected pattern for only two features – *second person pronouns* and *modals and semi-modals: prediction and volition*. They are the only user to follow this pattern for the latter, although as Figure 39 shows, they do so only loosely. On the other hand, the user exhibits the inverse of the expected pattern for seven features (*verb + that-clause: certainty* and *I + verb + that-clause: certainty, noun + that-clause: certainty* and *noun + that-clause: likelihood, modals and semi-modals: possibility, permission and ability, adjective + to-clause: certainty, and first person plural pronouns*), which is the joint highest number (shared with User 5) of any user.

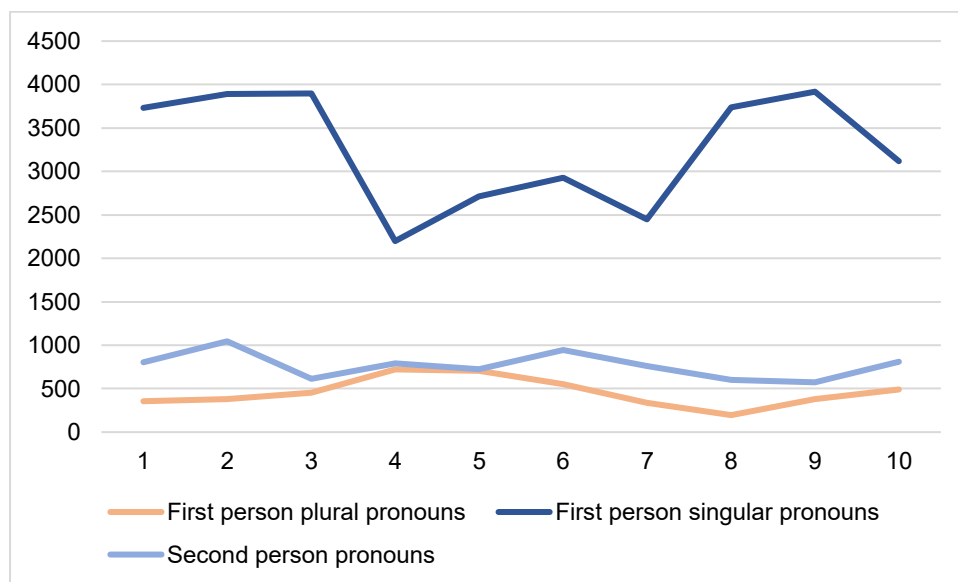


Figure 41: Usage of the three pronominal features (pmw) across User 2's career

User 2 shows the typical pattern of similar rates of usage for first person plural and second person pronouns, with more frequent usage of the first person singular. However, this user is unusual in the overwhelming frequency of their usage of the latter pronouns. This is by far their most frequently used of all the twelve features (Figure 39), and they are overall the most frequent user of this pronoun type by some margin (Figure 21, §7.4.2.1.1.5). As both graphs show, this feature shows an unusual pattern of extremely high usage at both the beginning and end of the career, with a sharp mid-career decrease.

Given the prevalence of first person singular pronouns in this user's posts, it is unsurprising that when comparing usage of *verb + that-clause* of certainty with its child feature *I + verb + that-clause* of certainty, the latter makes up a higher proportion of tokens of the parent feature for User 2 than for any other user. In other words, User 2 overtly inserts their own voice into this construction to a greater extent than the other users, as Figure 42 shows. As a consequence, User 2 is one of the highest-frequency users of the child feature, but not the parent.

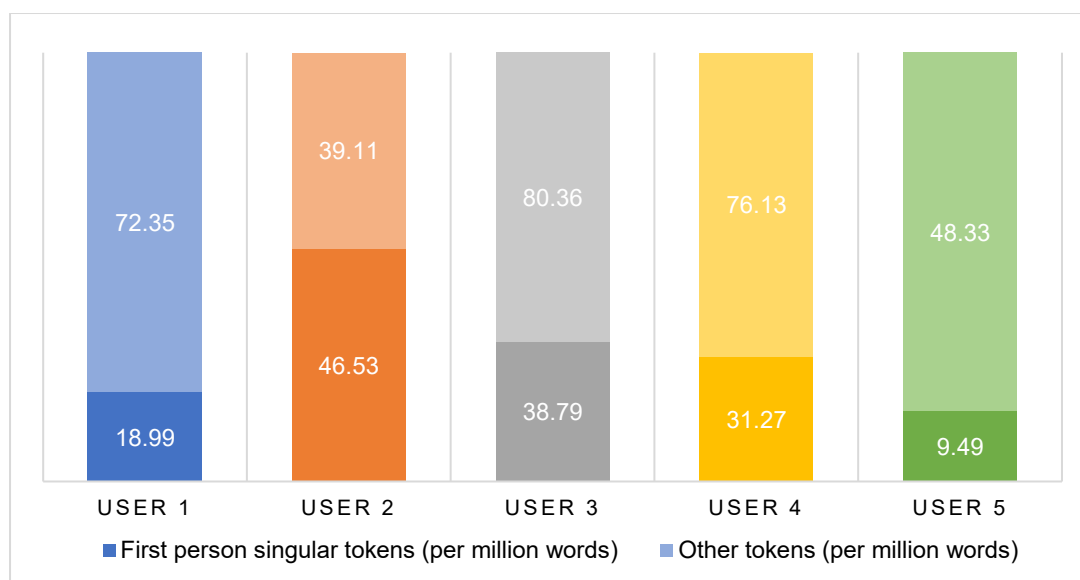


Figure 42: First person singular tokens as a proportion of all verb + that-clause: certainty tokens

Despite the very high rates of usage for some features discussed above, User 2 also shows very low or even no usage for almost half of the twelve features. These are: *noun* + *that-clause* of certainty (n=14) and likelihood (n=10), *adjective* + *that-clause* of likelihood (n=0), *adjective* + *to-clause* of certainty (n=5), and progression verbs (n=21). The user also shows the lowest rate per million words out of all the users for both categories of modal verb, as Figure 22 (§7.4.2.1.2.1) and Figure 24 (§7.4.2.1.2.2) show.

7.4.2.2.3 User 3

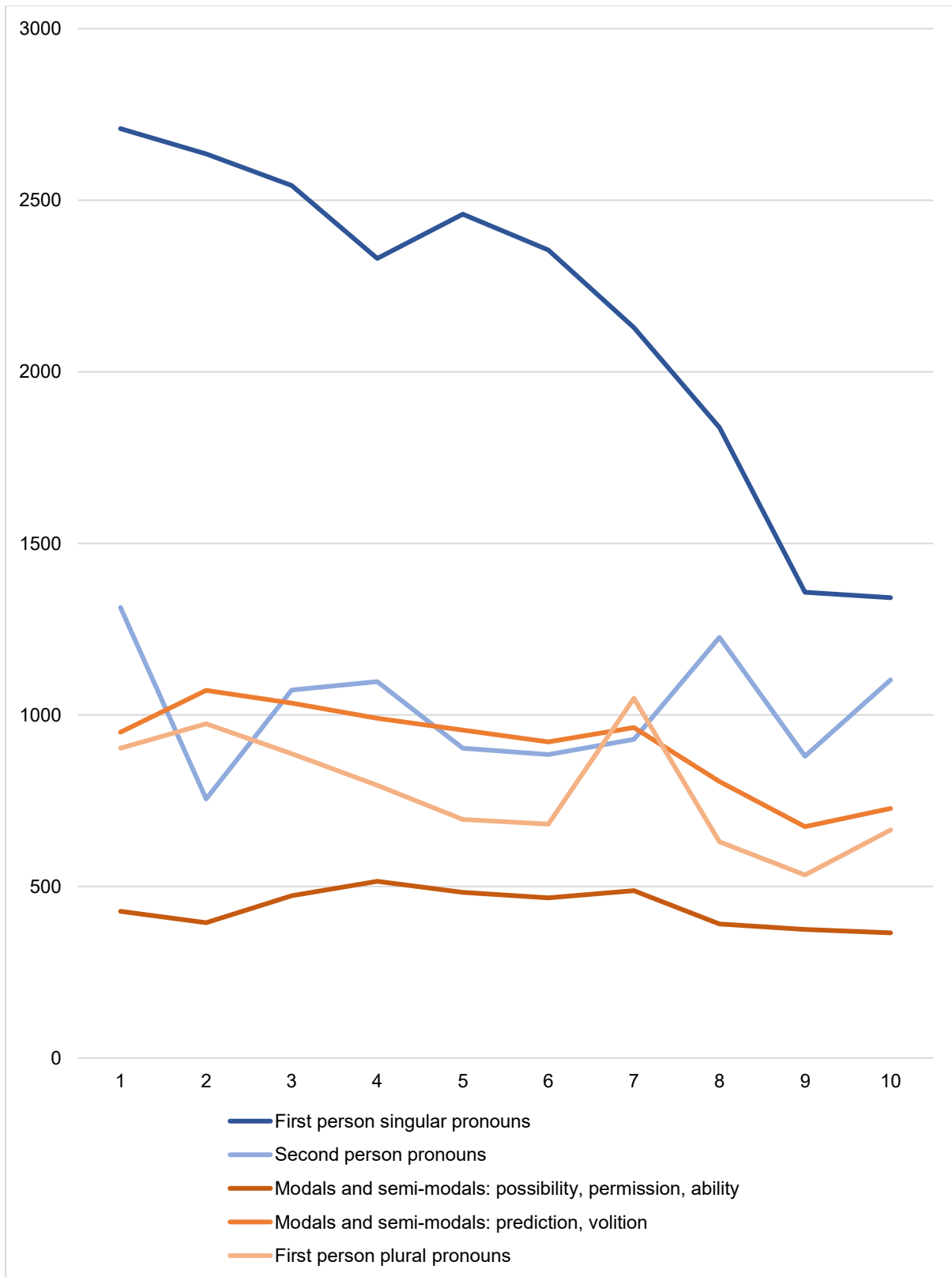


Figure 43: Use of higher-frequency features (pmw) across User 3's career. Mayfly features are represented in blue, and cicada features in orange

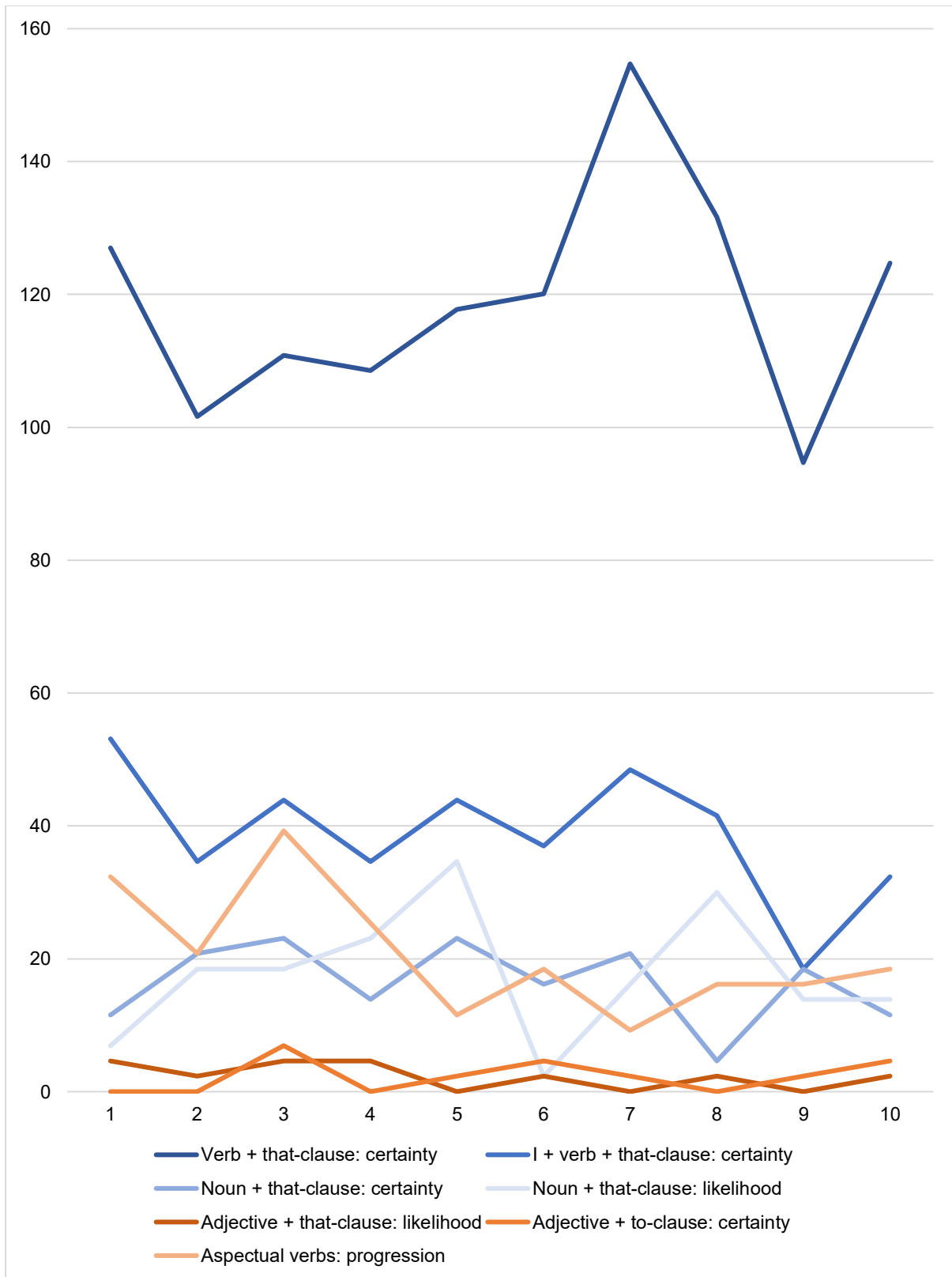


Figure 44: Use of lower-frequency features (pmw) across User 3's career. Mayfly features are represented in blue, and cicada features in orange

User 3 shows a general decrease in usage of many features as their career progresses. This included both mayfly and cicada features. Therefore, User 3's usage both matches and contradicts expectations for an equal number of features. As expected, we find decreasing usage of *I + verb + that-clause: certainty* (but not its parent feature), *noun + that-clause: certainty*, and *first person singular pronouns*, which are all mayfly features. Counter to expectations are the decreasing usage patterns for *modals and semi-modals: prediction and volition*, *first person plural pronouns*, and *aspectual verbs: progression*, all cicada features.

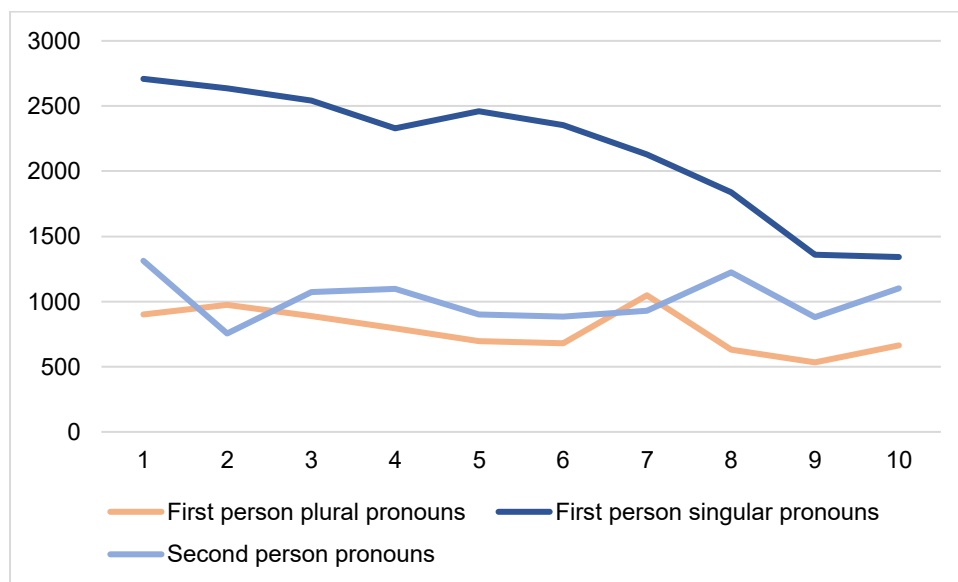


Figure 45: Usage of the three pronominal features (pmw) across User 3's career

There is a very marked, quasi-linear decrease in usage of first person singular pronouns throughout User 3's career, which follows expectations for a mayfly feature. However, unlike User 1 who shows a similar pattern for this feature, this is not matched by the expected corresponding increase in the first person plural; in fact, the latter remains the least-used pronoun type throughout most of the career. Perhaps this indicates that User 3 assimilates towards a collective identity to lesser extent than User 1, or simply that they express such assimilation in different, less overt, ways. The significant decrease in User 3's first person singular pronoun usage means that, while this remains the most frequently used pronoun type throughout, by the end of the career its frequency approaches that of the other two pronoun types, a pattern which is unique among the five users.

Turning to second person pronouns, User 3 is something of an outlier in both frequency and pattern of usage over the career. As Figure 20 (§7.4.2.1.1.6) reflects, this user shows the highest rate of usage of any of the users. While the exact level of usage fluctuates throughout the career, it remains at a similarly high level, with no upward or downward trend. This is in contrast with the other users, who all show some degree of generally decreasing usage as their careers go on.

User 3 shows a relatively low rate of usage for both types of modal verb, especially for modals of possibility, permission and ability; only User 2 shows a lower relative frequency. Similarly, User 3 presents the most marked decrease in usage of modals of prediction and volition across their career, going from one of the most frequent users of the feature to the least frequent. This pattern is contrary to expectations for a cicada feature. Despite these similar lower rates for modal verbs, User 3 presents only two very low frequency features, unlike User 2. Adjectival stance constructions are dispreferred by User 3, with *adjective + that-clause: likelihood* and *adjective + to-clause: certainty* showing only 11 and nine tokens respectively.

7.4.2.2.4 User 4

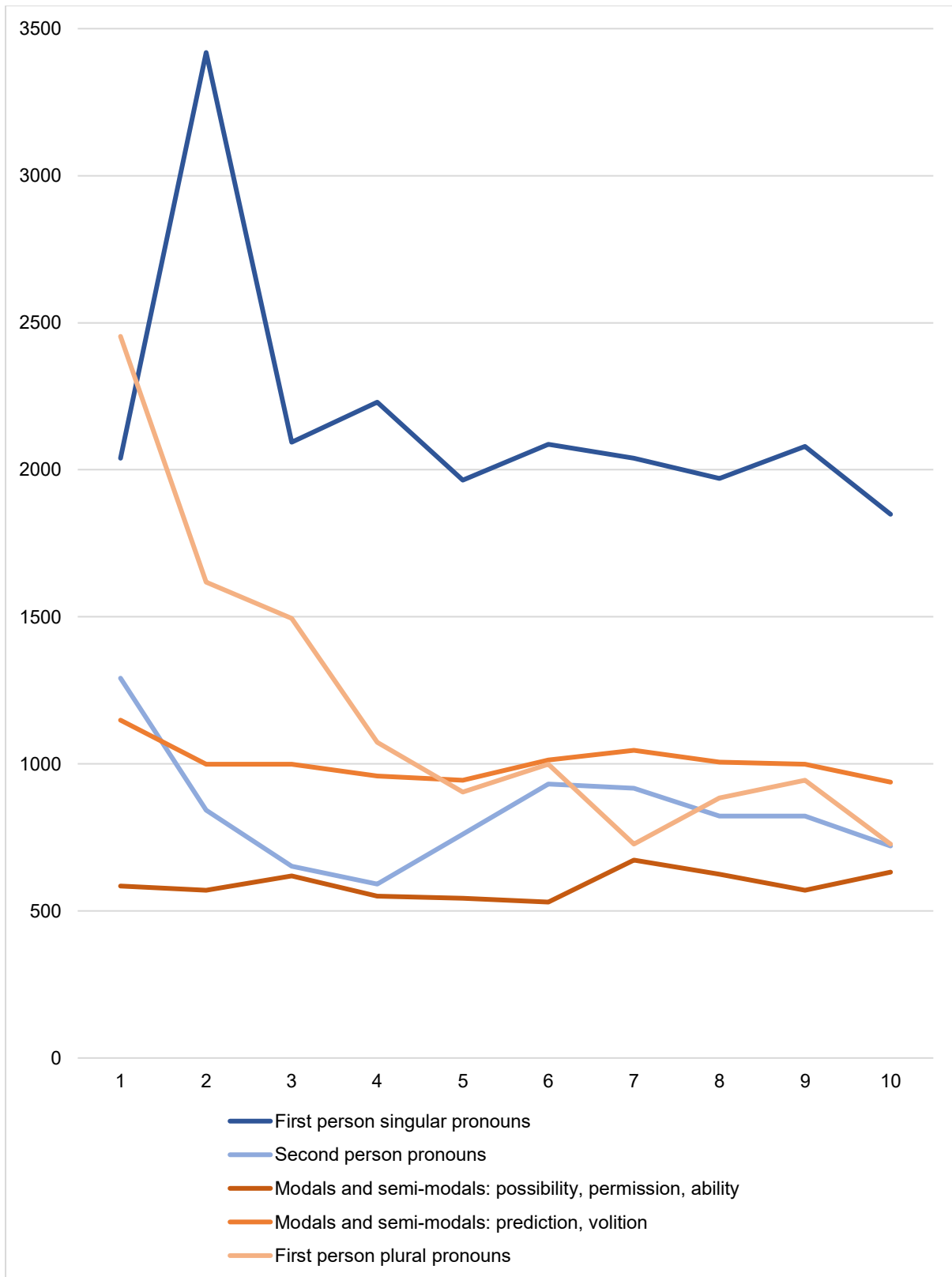


Figure 46: Use of higher-frequency features (pmw) across User 4's career. Mayfly features are represented in blue, and cicada features in orange

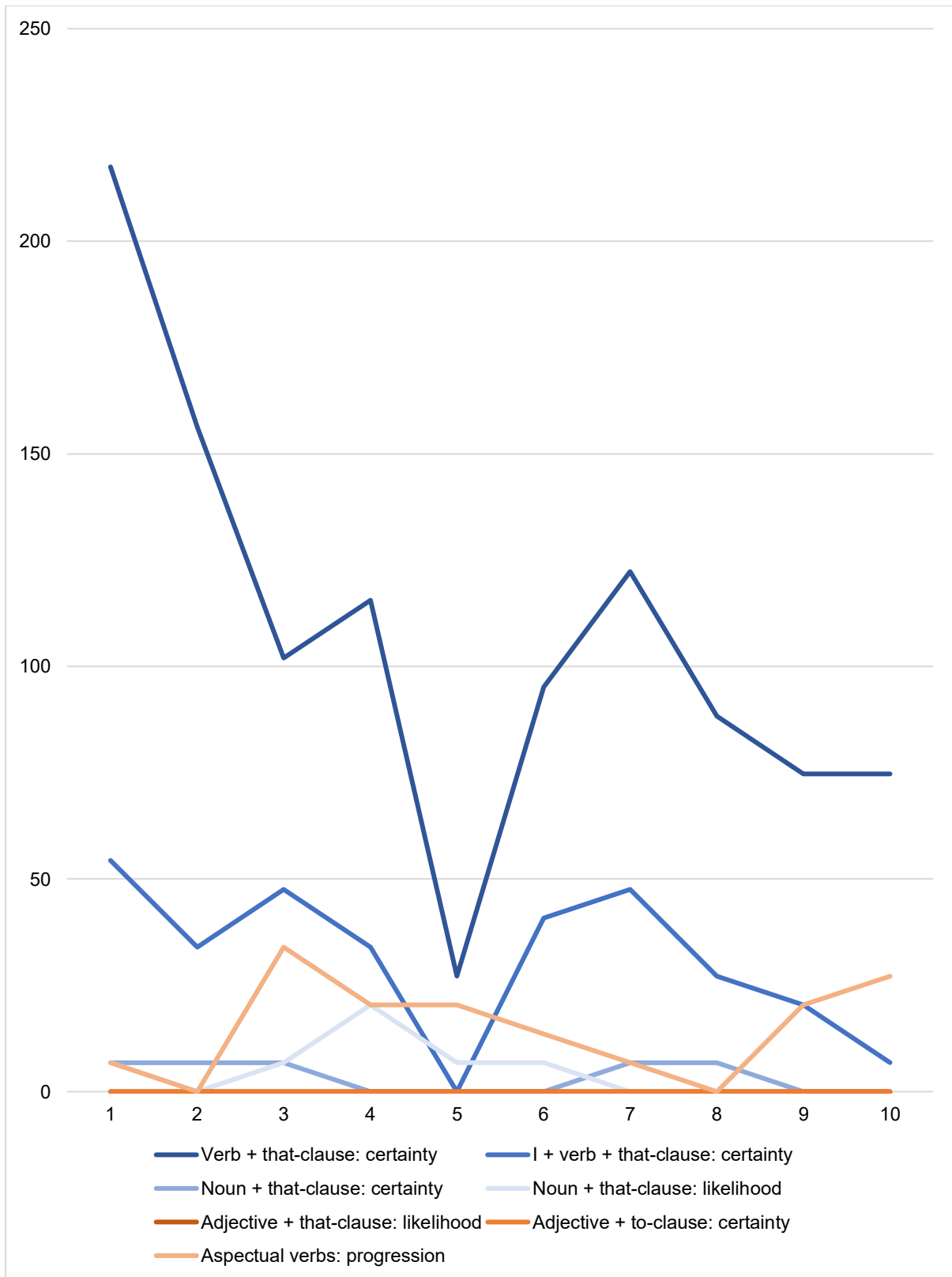


Figure 47: Use of lower-frequency features (pmw) across User 4's career. Mayfly features are represented in blue, and cicada features in orange

User 4 shows similarities with several other users. Arguably, the most salient of these is User 1. Both show a general tendency towards expected patterns. In the case of User 4, six features conform with expectations (*verb + that-clause: certainty*, *I verb + that-clause: certainty*, *noun + that-clause: likelihood*, *first person singular pronouns*, *second person pronouns*, and *modals and semi-modals: possibility, permission, and ability*). Unlike User 1, however, and more similarly to User 3, all but one of these conforming features are mayfly features. Additionally, both of the features for which User 4 exhibits the inverse of the expected pattern – *modals and semi-modals: prediction and volition* and *first person plural pronouns* – are cicada features. Again, this is similar to User 3; User 1, on the other hand, presents none of these inverse patterns.

The expected decreasing usage over time for *verb + that-clause: certainty* is particularly evident for User 4 (see Figure 13, §7.4.2.1.1.1). In the first decile, this user demonstrates by some margin the highest usage of this feature. This drops rapidly, however, with User 4 showing average usage from the third decile onwards. While a general decrease over time is also observed for *I + verb + that-clause: certainty* (Figure 14, §7.4.2.1.1.2), this is much less prominent than for the parent feature; in part due to User 2's preference for this construction, User 4 never shows the highest usage for it throughout their career, and there is no sharp decrease from an early peak.

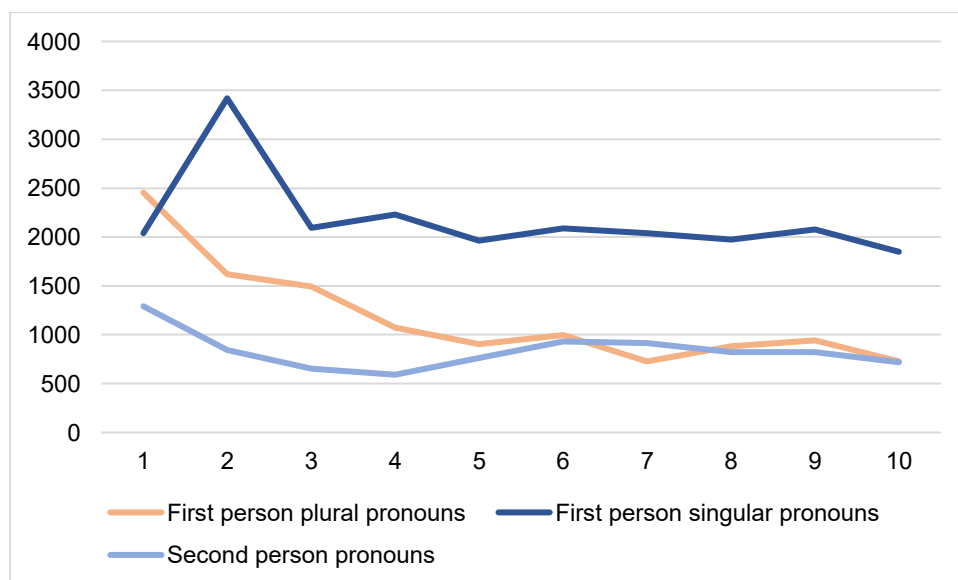


Figure 48: Usage of the three pronominal features (pmw) across User 4's career

Similar to *verb + that-clause: certainty*, User 4's usage of first person plural pronouns shows a very high frequency in the first decile – in fact, they produce over twice as many tokens per million words compared to the next highest frequency (Figure 30, §7.4.2.1.2.5). This contributes to a very marked downward trend in the use of these pronouns across the

career, compared to the four other users. This trend is contrary to the expectation for a cicada feature. Although User 4 is similar to most other users in that their usage of first person singular pronouns is generally the highest of all the pronominal features throughout (most of) the career, their first person plural usage in the first decile is high enough that this is briefly the user's most frequently-used pronoun type.

Much like User 2, almost half of the features are very low frequency or even absent from User 4's discourse. For User 4, these features are *noun + that-clause: certainty* (n=5), *noun + that-clause: likelihood* (n=7), *adjective + that-clause: likelihood* (n=0), *adjective + to-clause: certainty* (n=0), and *aspectual verbs: progression* (n=22). As seen in Table 22 (§7.4.1), both users provide smaller corpora overall, which likely explains this phenomenon.

7.4.2.2.5 User 5

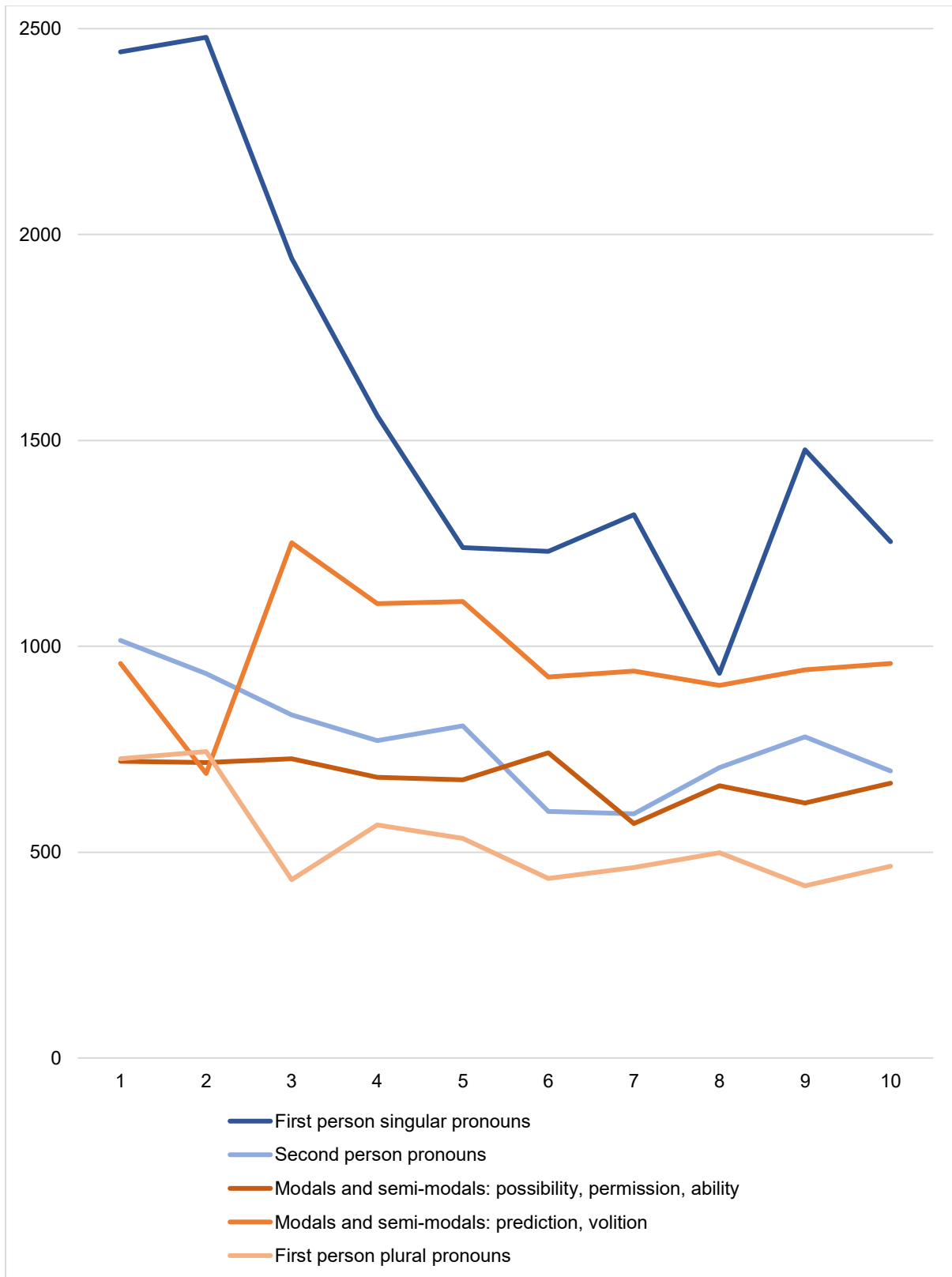


Figure 49: Use of higher-frequency features (pmw) across User 5's career. Mayfly features are represented in blue, and cicada features in orange

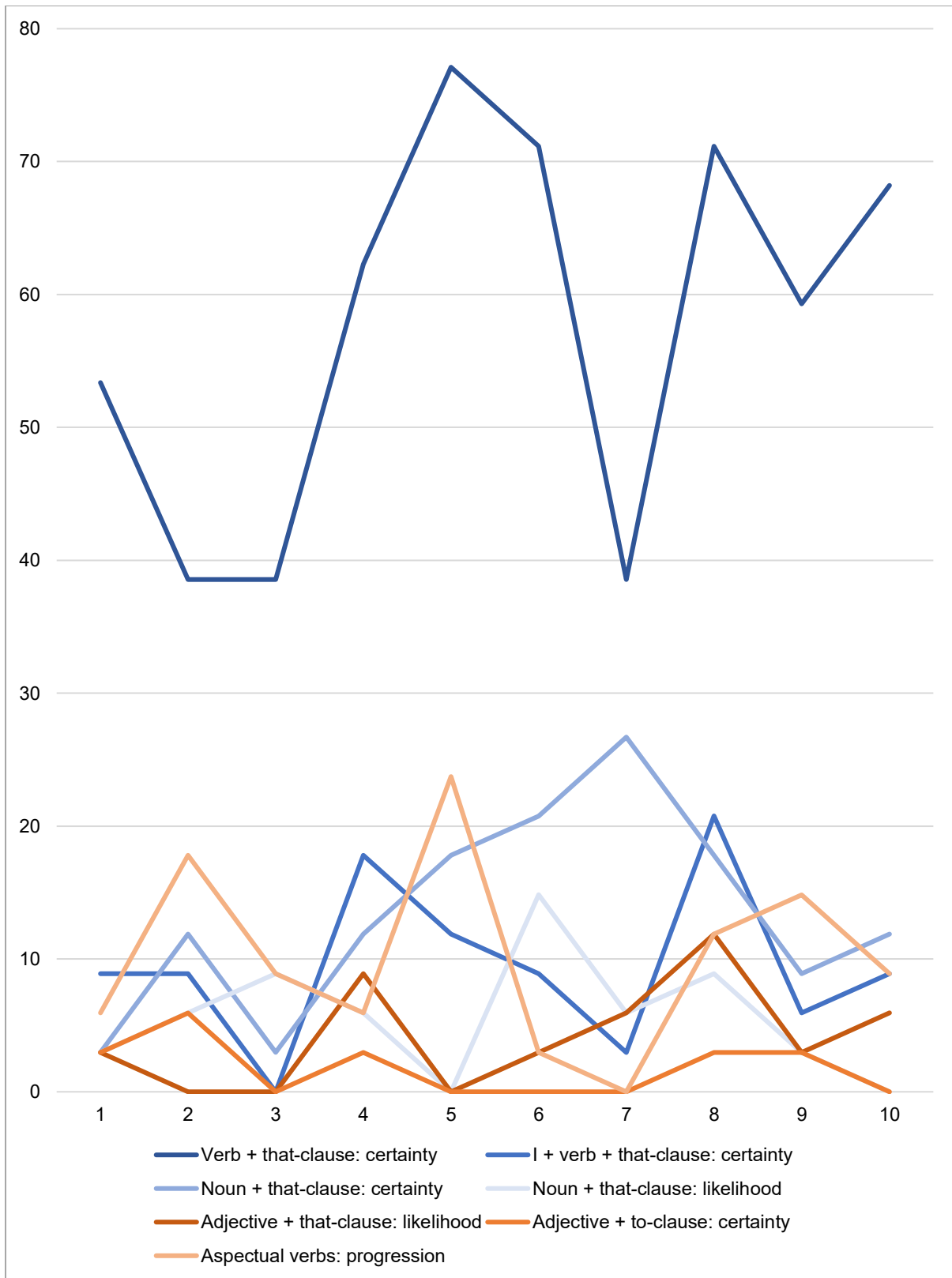


Figure 50: Use of lower-frequency features (pmw) across User 5's career. Mayfly features are represented in blue, and cicada features in orange

The last of the five users under investigation here, User 5, shares similarities with User 2 in that their usage of many features goes contrary to expectations. Indeed, many of these features are the exact same ones exhibited by User 2. These features include a range of both mayfly and cicada features, unlike Users 3 and 4. Features for which User 5 does not conform to expectations are: *verb + that-clause* and *I + verb + that-clause* of certainty; *noun + that-clause* of both certainty and likelihood; *modals and semi-modals* of both possibility, permission and ability, and prediction and volition; and *first person plural pronouns*. However, User 5 does follow expected patterns for some features; these are primarily features for which most users follow expectations, namely *first person singular* and *second person pronouns*. However, User 5 is unique in their (albeit loose) conformity on the cicada feature *adjective + that-clause: likelihood*.

While User 5 shows some notable similarities with User 2, there are nonetheless differences between their usage of some features. This is especially found in the use of the *verb + that-clause* and *I + verb + that-clause* of certainty constructions. Whereas Figure 40 (§7.4.2.2.2) shows that a very high proportion of User 2's *verb + that-clause: certainty* tokens are expressed using the first person singular, User 5 on the contrary shows the lowest proportion. In part, this reflects User 5's relatively low overall usage of first person singular pronouns (Figure 23, §7.4.2.1.1.5), which represents another key difference between this user and User 2. However, while User 5's first person singular usage as a whole decreases throughout the career, tokens of *I + verb + that-clause: certainty* do not follow this decrease; in fact, they generally increase as the career goes on.

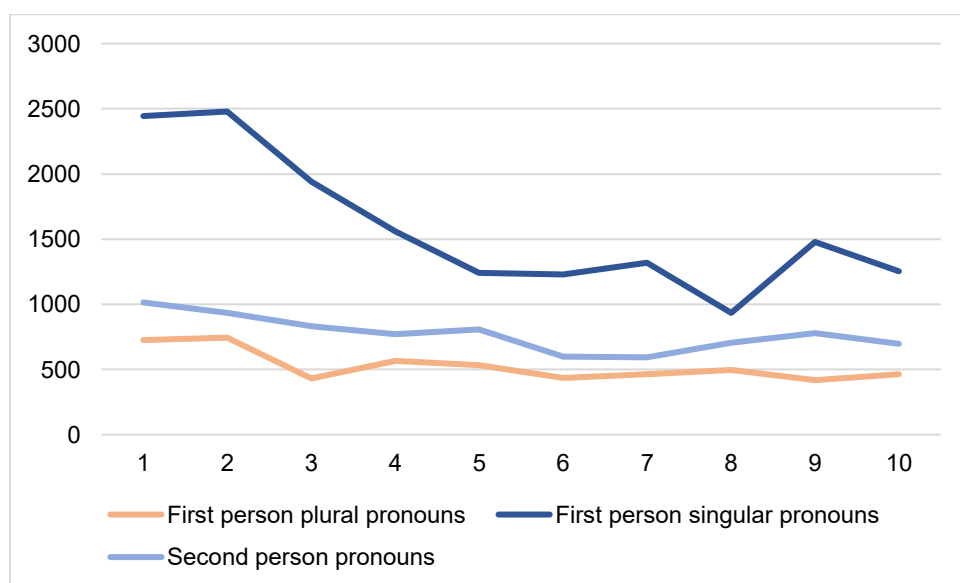


Figure 51: Usage of the three pronominal features (pmw) across User 5's career

The decreased use of first person singular pronouns over time produces a pattern comparable to that shown by User 3. While the first person singular remains User 5's most

frequently used pronominal feature throughout the career, the difference in its usage compared to other pronoun types narrows significantly over time. As is typical across most users, the first person plural and second person pronouns exhibit a relatively similar rate of usage across the career.

Of the five users, User 5 shows the highest rate of usage for modals and semi-modals of possibility, permission and ability in the first half of the career. This persists until the seventh decile, where Users 1 and 4 begin to exhibit the expected increase in this cicada feature's use in the later career. User 5, contrary to expectations, does not show a similar increase, instead showing a downward trend in usage over time.

Finally, User 5 exhibits very low frequency for several features, including *noun + that-clause: likelihood* ($n=19$), *adjective + that-clause: likelihood* ($n=14$), *adjective + to-clause: certainty* ($n=6$). On the contrary, their usage of *adjective + that-clause: likelihood* is unexpectedly high compared to other users. This exemplifies the fact that, while all twelve features were identified as potentially instructive in understanding how users change their linguistic behaviour throughout their careers, not all features prove to be equally useful in this regard, and certainly not for all users. This is illustrated further in the next chapter, where a smaller selection of features are considered for qualitative analysis.

7.4.3 Discussion

To a certain extent, the findings presented in the latter half of the chapter meet the expectations identified in the former. In particular, the use of mayfly features follows the expected patterns; on the individual level, more users conform to expectations with these features than the cicada features, and as a result the average usage over careers for mayfly features is as expected. On the other hand, only one cicada feature meets expectations of usage (*aspectual verbs: progression*), and even here the pattern is somewhat unclear, with fluctuations rather than the hypothesised clear upward trend. Meanwhile, four of the six mayfly features clearly follow the expected patterns (*verb + that-clause: certainty*, *I + verb + that-clause: certainty*, *first person singular pronouns*, and *second person pronouns*).

In interpreting these findings, we must not lose sight of the fact that the results presented in the latter half of this chapter and discussed here are based on only a small sample of cicadas, five out of the total 4,442 identified in Chapter 6 (Table 12, §6.4). These users are the most prolific posters within that group and, while this makes them the most valuable for (quantitative) linguistic analysis due to the amount of data they provide, they may therefore already be considered outliers among the cicadas. It is possible that analysis of another set of five cicadas may have produced different results, either more or less in line with the hypothesised findings.

In addition, some results have been classified as not following expectations simply because they do not show any clear pattern, expected or otherwise. Many of the less clear results stem from small numbers of tokens for certain features. While some features have an average relative frequency (across users and deciles) in the hundreds or even thousands, seven of the features – categorised as ‘lower-frequency’ features above – have an average relative frequency of under 100, and six of under 30. Unsurprisingly, the lower-frequency features show less clear patterns overall. The inclusion of more users, and therefore more tokens, may have revealed clearer patterns.

It is notable that most of the features which follow the expected patterns are mayfly features. This suggests that early-career cicadas do, at least in some respects, behave in the same way as mayflies. This aligns with literature which suggests that some linguistic features are strongly associated with newcomer behaviour, with no distinction identified between those who will go on to have longer careers and those who will not (e.g. Golder & Donath, 2004).

The results from the cicada features complicate this picture, however. Although these features are more strongly associated with longer careers overall, they do not follow the hypothesised pattern of greater usage in the later stages of the career. In other words, while the mayfly features are apparently associated with both short careers *and* the early stages of long careers, cicada features appear to be associated with long careers, but not necessarily with later stages within those long careers. This has two implications.

The first implication follows from the fact that, while these features are positively correlated with career length, and therefore not associated with short careers, the members investigated here use these features most often while their careers are still of short standing. In other words, these findings suggest that these features are primarily used by early-career users who will go on to have long careers on the forum, and that this could distinguish such users from other newcomers. This information could therefore theoretically be used in the early identification of users who will go on to have long careers and possibly to undergo further assimilation into a white nationalist collective identity through their forum membership. This a tentative conclusion, which would require testing with a much larger set of users.

The second implication from this finding is that, while cicadas may follow more predictable patterns early in their careers, different users may adopt distinct linguistic behaviours as their careers progress. We have seen that most of the features I have investigated here are more strongly associated with the early career. This suggests that few, if any, specific linguistic features are specifically associated with later stages of long-term careers, and that there may be significant individual variation between users. This indicates that, as hypothesised,

users develop linguistic capital across their career and are able to draw on a wider pool of linguistic resources by the end of the career, while early-career users remain constrained in their linguistic performance. It also suggests that through the career users become more comfortable in the assertion of their own identity in the group context, and feel less pressure to conform to group norms than earlier-career users. This variety in identity performance may give the appearance that users have not assimilated to a shared collective identity. However, as Melucci (1996) argues, assimilation to the collective identity cannot result in static performances of that collective identity. Because collective identity is a process, members' interactions with the group form part of the negotiations that shift the collective identity over time. This means that collective identity not only leaves room for, but actively requires, individuality in its members' identity performances.

Supporting this hypothesis of greater individuality over time, the five users investigated here all present differences in their linguistic behaviour – there is little evidence here of a typical career across the board. As the graphs presented in §7.4.2.1 show, there are relatively few features for which all, or even most, users follow the same broad pattern of usage. There are no features for which all five users clearly follow the same pattern across their respective careers, and only a small number of features (*modals and semi-modals: prediction and volition* and the three pronominal features) show a majority of users following the same pattern. First person plural pronouns in particular shows two clear but opposing patterns of usage, with User 1 increasing their usage over time in line with expectations, and all four other users showing a more or less marked trend in the opposite direction (Figure 33, §7.4.2.1.2.5).

Overall, I have shown in this chapter that, through a quantitative lens, there is no factor which unites all five users and defines a typical career. In the next chapter, I investigate whether a qualitative approach is able to identify any similarities which unite the five users, or whether the finding of significant individual variation is reproduced.

Chapter 8 The behavioural career II: A qualitative linguistic analysis

8.1 Introduction

In Chapter 7, I carried out a quantitative analysis of linguistic behaviour by forum members across and within career types. In doing so, I was able to identify linguistic features most closely associated with the short-term *mayfly* users and long-term *cicada* users that I had previously described in Chapter 6. I then investigated the use of these features across the careers of five cicada users, examining how individuals' linguistic behaviour changes over time. This high-level quantitative analysis showed that, while some similarities exist between users, the overall picture is that of individual variation with no one typical linguistic career path identified. In this chapter, I use these linguistic features to examine these five users' behaviour across their career from a qualitative perspective. Findings are reported user by user, followed by a discussion which draws together all five users and identifies major patterns across the data.

As outlined in more detail in §3.4, a number of researchers have used qualitative linguistic methods to investigate performances of newcomer and experienced identities in online spaces. There is general agreement that newcomers frequently make use of negative politeness strategies (Brown & Levinson, 2006) to emphasise their low status (Benwell & Stokoe, 2006; Golder & Donath, 2004) and avoid threatening the face of more established community members. Newcomers perform their lower status by seeking, rather than providing, information (Galegher et al., 1998; Golder & Donath, 2004; Newon, 2016). They also explicitly perform their newcomer identity by providing personal narratives which relate their motivations for joining the community (Arguello et al., 2006; Chiang, 2018; Graham, 2016). A smaller proportion of newcomers may perform a 'competent newbie' identity (Chiang, 2018), highlighting their pre-existing expertise in order to outline their value to the community. On the other hand, experienced or veteran members may display the opposite behaviours to newcomers, offering information and advice (Newon, 2011; Rudolf von Rohr et al., 2019; Singh, 2012) and demonstrating their topic expertise (Newon, 2011; Nguyen & Rosé, 2011; Peuronen, 2011). These members may interact with newcomers by providing guidance or support to them (Angouri & Sanderson, 2016; Lampe & Johnston, 2005), and by offering both positive and negative evaluations of others' contributions to the forum (Newon, 2016; Rudolf von Rohr et al., 2019). These evaluations may function in part to gatekeep the community, upholding its boundaries and community norms (Golder & Donath, 2004; Graham, 2016; Honeycutt, 2005). Veteran members may also engage in friendly and off-

topic talk, which may serve to strength bonds within the community (Ang & Zaphiris, 2010; Angouri & Sanderson, 2016).

8.2 Method

The same five users were selected here as for the previous quantitative study. These are the most prolific five users of the cicada group, who have posted to the forum over a period of seven or more years. The same linguistic features were also investigated – those which I have identified (§7.3) as being most strongly correlated with career length. These features are shown in Table 35 below. As noted in §4.4.3, only data from the first and final deciles of each user’s career was investigated. This was due to time constraints, and the assumption that these two deciles would provide the clearest view into how linguistic performance changes across the career. I chose to analyse a maximum of 100 tokens per decile, per feature, per user, as illustrated for User 1 in Table 35. Where a user provided more than 100 tokens of a feature, a random sample of 100 concordance lines was selected using Sketch Engine’s random sample tool. Where a user provided fewer than 100 tokens, every token they did provide was included.

User 1 Feature	Number of concordance lines	
	Decile 1	Decile 10
Verb + <i>that</i> -clause: certainty	41	34
I + verb + <i>that</i> -clause: certainty	12	8
Noun + <i>that</i> -clause: certainty	6	4
Noun + <i>that</i> -clause: likelihood	5	8
First person singular pronouns	100	100
Second person pronouns	100	100
Modals and semi-modals: possibility, permission, ability	100	100
Modals and semi-modals: prediction and volition	100	100
Adjective + <i>that</i> -clause: likelihood	0	0
Adjective + <i>to</i> -clause: certainty	0	3
First person plural pronouns	100	100
Aspectual verbs: progression	15	16

Table 35: Inclusion of feature tokens across the career, demonstrated for User 1

The selected concordance lines were downloaded as an .xlsx file for easier annotation outside of Sketch Engine. As detailed in §4.4.3, I annotated each concordance line, commenting on the content and function of each token. These annotations were then compiled into lists of topical and functional themes for each user, feature, and decile. Although all concordance lines for every feature were annotated, the most uncommon features did not prove useful in identifying diachronic patterns across the career, as it was unclear whether any findings truly reflected patterns. This was particularly the case for features with tokens in only one decile, such as *adjective + to-clause: certainty* in Table 35, as no direct comparison was possible at all. Therefore, in practice, only the more frequent

features were used to understand how users' behaviour changed across the career. Less frequent features were used to supplement these findings, fitting into patterns that had already been identified.

8.3 Observations by user

The following sections explore each user's linguistic behaviour one by one, offering a more detailed qualitative account than was afforded by the quantitative approach in Chapter 7. These sections will characterise each user as a whole, taking into account themes which appear in both deciles of users' careers, and investigate the user's trajectory across the career, considering themes which change from decile to decile. These individual accounts of the five users are followed by a discussion drawing out elements which unite all five users and the nature of the changes that can be observed throughout their careers.

Despite the more detailed qualitative approach taken, the findings reported here should not be understood to be completely comprehensive. This limitation follows from the research design. By using large datasets, only taking the first and last deciles into account, and selecting only more frequent themes in order to interpret users' behaviour over time, some elements of this behaviour will inevitably be lost.

Where quotes from the data are given, these are marked with a (1) or a (10) respectively, to indicate which of the two deciles they are drawn from. Quotes are given verbatim, including any non-standard orthography, except where potentially identifying information has been anonymised (see §4.5).

8.3.1 User 1

User 1 is, in some ways, the prototypical example of what we might expect from a white nationalist forum user. He is critical of both racialised and political outgroups, and aims to rally other members of the movement to action as well as to attract fellow white people to join the movement. However, as the user's career goes on, there is evidence that he has become jaded by the movement. He is more critical of other members, and of other white people, who he increasingly sees as contributing to their own demise through their apathy.

Unsurprisingly, given the nature of the forum, User 1 is critical of racialised outgroups throughout his career (examples 1-4), and accordingly positions white people as victims of these groups (examples 4-7).

1: I think White people are beginning to see that asians like to rip people off (1)

2: *If you put a ***** on an oil rig you will have nothing but trouble. I can just see them trying to steal oil. (1)*

3: *there might be 500,000 muslims in this country right now that are willing to go on some kind of crime spree or commit terrorist acts. (1)*

4: *The Jews have succeeded in brainwashing our people (10)*

5: *They will start to harm White people from the minute they get there. (1)*

6: *I have noticed that negros have no qualms about attacking elderly White people. (1)*

7: *these terrorist training camps are training Muslims to kill as many Whites as they can. (10)*

Although this occurs in both deciles, there are many more examples in this user's early career. By the late career, on the other hand, the user has instead become much more critical of other white people (examples 8-10). This includes other white nationalists (examples 11-13) and even white nationalist leaders (examples 14-15).

8: *I am beginning to think there is some kind of fundamental flaw in White People that keeps them from reaching their full potential (10)*

9: *too bad most Whites will not give us a chance in explaining what the truth really is (10)*

10: *more than can be said for lazy and stupid Whites in America who are doing nothing towards protecting and defending their race. (10)*

11: *WN in the U.S. can help the overall picture simply by being the best we can be which few of us are. (10)*

12: *I still don't see what I have been looking for since joining [forum] in 2005 - serious minded Whites who are willing to come together (10)*

13: *If WN was strong and unified other Whites would begin to follow but they aren't because we don't have our own house in order. (10)*

14: *I was at one time a member of the National Alliance but like most of their followers I decided not to renew my membership due to poor leadership. (10)*

15: I have a hard time understanding how Will Williams¹² can do so many things that are in bad taste and judgement (10)

As example 12 neatly summarises, User 1 appears to have grown restless and frustrated with his membership in a movement which he perceives not to be making any progress in its goals. Both white people in general, and white nationalists, are blamed for this perceived failure. The former are increasingly seen to be complicit in what this user perceives as their own demise. The user appears to distance himself from the latter group, implicitly including himself among the “few of us” (example 11) who behave in a way that benefits the movement. In this way, he goes some way to defining the proper conduct of a white nationalist, and positions himself in proximity to this behaviour. The criticism of white nationalist leaders (particularly Will Williams, leader of the National Alliance) in the late decile is particularly notable. Perhaps the user, now more established among the forum community, feels empowered to make criticisms of powerful figures which he did not previously feel comfortable making. This may also be a symptom of the user’s apparent increasing frustration with the movement at a fundamental or structural level as his career goes on.

This interpretation is supported by the user’s overt enthusiasm for the movement and organisations within it, including the National Alliance, earlier in the career.

16: The question is what is going to be the one organization that will unite the White race? Consider joining the National Alliance. (1)

17: The White race needs to take heed of my words - put your foot down and join the National Alliance today! (1)

18: The best way for you to be compleately supportive of your fellow Whites is to 1. Join a politcal organization such as the National Alliance, the Natioanl Vanguard, Aryan Nations, the Nationalist Socialists White Peoples Party or a similar White unity group. (1)

Of course, as examples 14 and 15 attest, the National Alliance has fallen out of this user’s favour by the late career. While the user continues to promote some organisations, he does so with less enthusiasm than previously.

¹² Leader of the National Alliance neo-Nazi organisation.

19: I still think the Council of Conservative Citizens is the best group we have unless someone can come up with something else I haven't thought about. (10)

20: Now I support NARRG (National Alliance Reform and Restoration Group) after it became apparent that the NA leadership had seriously deteriorated starting in the year 2002. (10)

This may be a result of the user's previous experience with the National Alliance. The user's greater knowledge of the movement and experience gained over time appear to have contributed to his more jaded and less enthusiastic attitude.

Related to the criticism of other white nationalists, this user engages in gatekeeping of white nationalism. This occurs throughout the career, but there are differences in how this is achieved across the two deciles. In both deciles, the user outlines his own understanding of what it is to be a good white nationalist – how white nationalists should act and think.

21: I do not know why anyone would be bothered by the label White Supremecist. (1)

22: What can I say except don't subscribe to a local main stream newspaper if you want to help the White race. (1)

23: If you are a Republican then you are not a White Nationalist even though you may think that you are. (1)

24: other than the fact that the true White Nationalist is a moral person that would not break the law. (10)

25: There are some simple basic ideas that all WN should understand and make an effort to put into effect. (10)

26: Your money should go to the Nationalist movement anyway upon your death. (10)

However, in the later decile the user's gatekeeping also expands in two different ways. First, the user begins to gatekeep not just the movement but this particular forum, including providing explicit explanations of forum rules and norms to other users (example 29). This serves to position the user as an authority on the forum, showcasing his experience as a long-standing forum member.

27: If you approve of race mixing then you are on the wrong forum. (10)

28: You can't have it both ways [username], this is the place of what most people would call "extreme right wing views". (10)

29: I appreciate your post however it is generally known that we speak and write in English on this forum. (10)

The user also appears to interact with new users more frequently than any of the other users investigated here. These interactions are both positive and negative, as examples 30-33 demonstrate.

30: You are on the right team [...] You will feel at home as part of the Nationalist movement. (10)

31: The more you get into it the more you will like it so don't worry about the big bad liberal [...] just get into the game and stay with it (10)

32: Are you a sustaining member? I can see that you aren't although you haven't been here very long either. (10)

33: I see that you just joined [forum] and immediately you are seeking help, the question is why were you not being a supporter when things were going well? (10)

This serves to emphasise the user's authority; he clearly feels empowered to evaluate other members, and positions himself as a guide or mentor for newcomers. The user is thus able to juxtapose his own long-standing career and sustaining (fee-paying) member status with that of the newcomers, emphasising his own status.

Gatekeeping in all of the forms displayed here allows User 1 to implicitly and explicitly define his own understanding of collective identity, outlining how white nationalists and forum members should behave and what beliefs they should have. This includes explicit, targeted praise and criticism of others' behaviour. In doing so, the user also raises his own position within the forum. In part, this stems from the fact that the user (presumably) adheres to his own standards of behaviour and opinion. The user therefore implicitly positions himself as an ideal member of the community, an example for others to follow. This also derives from the user positioning himself as an authority within the forum, someone with the power to include or exclude, and to evaluate others' contributions. This is particularly the case in the later career, where the user positions himself as a guardian or custodian of the forum, enforcing its norms and explaining them to others. Despite this authority, the user also positions himself as a somewhat benevolent figure within the community, welcoming and reassuring some newcomers.

While User 1 therefore appears to increase his social power in the late career, he does not always wield this power in a typically authoritative, direct manner. In line with this, the user often expresses his views more assertively early in the career. This is achieved using a variety of means, including the implicit or explicit assumption that his view is correct (examples 34-35), claiming to speak on behalf of all white nationalists (36) or even white people (37), and offering (sometimes dubious) evidence for his arguments (38) more frequently.

34: some of them have come to the realization that liberalism is stupid and wrong. (1)

35: We are on the side of the facts and conversley the facts are on our side. (1)

36: We are White Nationalists and state senators do not represent us. (1)

37: White people know from our everyday dealings with negros, arabs and hispanics that they do not measure up in everyway to White people (1)

38: A study of history shows that England was as responsible for WW 1 as much as any country (1)

While not aligned with typical newcomer behaviour, which involves the self-conscious performance of 'newness', some researchers (Chiang, 2018; Graham, 2016) show that newcomers do attempt to position themselves as topic enthusiasts or 'competent newbies' in order to 'prove themselves' to the community. Although not entirely absent from the later decile for this user, the techniques exemplified in examples 34-38 are much less frequent there. Instead, there is evidence that rather than attempting to 'prove himself', the user instead allows his long-standing membership, and status as a sustaining member of the forum, to speak for themselves in his later career. The user does allude to his long career, both by narrating the political journey he took into white nationalism (examples 39-40), and by directly alluding to his veteran status (41-42).

39: As the years continued to come and go I noticed that things were beginning to get bad, really bad as a matter of fact. (10)

40: I started doing some research on which Nationalist organization would be best to join and decided the National Alliance was the largest as well as being the best so I signed up with them. (10)

41: As you can see I have been posting on [forum] for 13 years, have joined several Nationalist groups and given away all kinds of money (10)

42: By the way I have been putting out a similar message since I cam here in 2005, not to take any credit for anything. (10)

Taken together, the user appears to express more general authority and knowledge earlier in his career, including some gatekeeping of the white nationalist movement. However, by the later decile the user appears to rely more heavily on community-specific knowledge and resources, such as gatekeeping the forum itself by enforcing norms, welcoming new users, and alluding to his long career as a white nationalist and as a forum member.

Despite seeming more weary of the movement later on, User 1 nonetheless continues to rally white nationalists, and white people more generally, to various kinds of action. Uniquely, this occurs throughout the career – while several other users perform similar calls to action early in their careers, no others continue to do so into the later career with so much frequency as User 1. This includes both calls for white people to come around to the movement (examples 43-45), and for white people and white nationalists to take various kinds of action (examples 46-47).

43: I would like to see all White conservatives comming together in some kind of way and additionally getting more White neutualists to get off the fence and realize that we are not the devils that liberals make us out to be. (1)

44: It would be nice if we could get our fellow Whites to understand they could support better conditions for workers of all types without being followers of what is essentially Communism. (10)

45: it takes a lot, too much as a matter of fact to reach our White brothers and sisters (10)

46: A White Homeland will not be achived unless we stop writing “our Congressman” and start taking matters into our own hands. (1)

47: The White race must march forward all at one time and totally united, that is how we will win. (10).

As example 45 hints, in the later career this is increasingly tied in with a general sense of despair at other white people and their inaction in the face of what this user sees as threats to their existence (see also examples 8-10). Unsurprisingly, given the user’s increasing

criticism of not just white people but the white nationalist movement, the later decile also features an increased focus on the need for unity within the movement, and criticism of the current disunity that the user perceives.

48: I think the greater problem is not differences but rather a lack of interest in uniting, people like having their clubs and would rather remain hobbyists even though such a course of action is not going to get us maximum results. (10)

49: I still don't see what I have been looking for since joining [forum] in 2005 - serious minded Whites who are willing to come together (10)

50: The question is are we finally ready to unify and take action or are we going to go back to our old habits of posting here on [forum] and then doing nothing? (10)

These criticisms in the late career may demonstrate the user's increased awareness of the current state of the movement, reflecting more specialised knowledge of the movement gained over his long career. Additionally, they show that the user is more willing to criticise the movement in his later career, perhaps evidencing more confidence in his own position within the movement and more willingness to challenge the status quo. The user also alludes to his long career (examples 49-50) and expresses frustration at how little has been achieved, showcasing a veteran identity within the movement (Newsome & Grant, in press). Although, as examples 43-47 show, the user is eager to encourage action from all white people throughout his career, these later complaints about the movement's inaction and disunity suggest a more inward-looking tendency developing later in the career, with more criticism directed to social actors closest to the white nationalist centre. This is also reflected in User 1's use of second person pronouns across the two deciles. In the early career, the user addresses these pronouns 'outwards' to a greater extent than in the late career, apparently addressing anyone who is reading, or explicitly addressing guests to the forum (i.e. non-member 'lurkers'); examples 51-53. Later in the career, he appears to directly address other individual users more frequently (examples 54-55).

51: All white guests listen up - it is time to join [forum]. You have been thinking about it and now is the time to do it. (1)

52: To those who have been sitting on the fence wondering if you should join [forum] or the National Alliance this would be a good time to do so (better late than never). (1)

53: I would like to propose that the organization change it's name to the American Federation of White people. Your thoughts? (1)

54: Hang on my friend, you will someday be getting in on all of the action. (10)

55: To answer your question, of course not. (10)

Altogether, the user's continued calls to action throughout the career, combined with his increased criticism of both white people and the movement, gives the impression that despite his frustration, he retains some hope for the prospects of the movement. Supporting this interpretation is evidence that as the career progresses, User 1 becomes more forward-looking in his perspective. The user expresses pessimism for the future in both deciles, as examples 56-59 show. However, there are more tokens of this to be found in the later decile.

56: We are going to start seeing more of this - invaders trying to tell the good White people of this country what to do and if we don't do as they say they will take action. (1)

57: Politicians will pass more anti-White laws. (1)

58: Hopefully more Whites will draw the same conclusion although I doubt it (10)

59: We are still basically in big trouble with the U.S. still slated to become a non-White nation in the next few decades (10)

This continued pessimism throughout the career surely links to the user's increasingly jaded and frustrated perspective in the later career, as he perceives that the situation has not improved throughout his time in the movement. However, in the late career the user also looks to the future in a more optimistic way.

60: 2017 WILL BE A GREAT YEAR FOR US SO LET'S GET GOING! (10)

61: I am sure the ADL¹³ is thrilled at what's happening at our end but a come back is coming. (10)

62: I believe that in coming conflicts those with better genetics will have the upper hand and come out on top. (10)

¹³ The Anti-Defamation League, an advocacy group that researches and works to combat hatred, especially antisemitism.

I therefore contend that, while the user takes a dimmer view of the white nationalist movement in the later career and this is in part reflected in a pessimistic view of the future, this in fact reflects two distinct elements of the user's development across his career. The first is that the user has become more jaded with the movement and with white people in general by the late career. The second is that the user looks to and makes predictions about the future of the movement and the so-called 'white race', both positive and negative, more frequently in his later career.

8.3.2 User 2

Compared with some other users, User 2 shows a relatively high degree of similarity across his entire career. User 2 takes a primarily social approach to forum membership throughout both deciles studied here, often taking part in off-topic discussion and making friendly remarks to other users. However, the user does make more racial and political comments in the later career, perhaps suggesting that this has become a greater concern for him through his membership.

A significant proportion of User 2's posts are 'off-topic', concerning lifestyle and popular culture matters such as music, films, and celebrities.

63: I try 2 stay away from burger joints period (1)

*64: First of all, The English Patient will make you go off into slumberland.
(1)*

65: I don't care for Michael Moore or his liberal politics, I'll wait and see this when it comes out on dvd. (10)

*66: I urge every member who is proud of their whiteness to boycott
DJANGO UNCHAINED and every Quentin Tarantino movie, old or new.
(10)*

As these examples demonstrate, while this user discusses 'lighter' matters throughout his career, he takes a decidedly more critical view later in his career, increasingly viewing these matters through a white nationalist – that is, racist and right-wing – lens. Such perspectives are not absent in the early career, nor are the more benign opinions such as those expressed in examples 63 and 64 absent from the late decile, but the examples presented here represent the more typical views expressed at the two points of this user's career. This may suggest that the user's views have become somewhat more extreme over time, or rather that his existing white nationalist views have become more salient to him, through the career.

This finding is supported by others. While this user is critical of racialised outgroups throughout the career (examples 67-70), contrary to User 1, there are more examples of such sentiment later in the career, suggesting that the user holds stronger views, or more inclination to express these views, later in the career.

67: Maybe we had a white community, maybe no minorities would work in fast food joints. Most of the minorities work there are rude! (1)

68: Yeah, I forgot about that uppity blabbermouth. She doesn't like white people at all. (1)

69: You can say that again. Mexico is a sewer! (10)

70: The Jews want to destroy our white race. (10)

Despite this general trend, the findings for the feature *modals and semi-modals: prediction and volition* show an apparently conflicting pattern, expressing more extreme and violent dislike for racialised groups in the early career. In particular, these tokens celebrate the use of capital punishment for Black prisoners, implying that such punishment is deserved.

71: I don't know if they'll get death, but I hope. (1)

72: He's gonna get the needle! I was waiting to hear about this. (1)

Similarly to User 1, User 2 also appears to be more critical of other white people in the later stage of his career. However, unlike the case of User 1, we have seen that this does not correspond with a decrease in criticism of racialised groups. Additionally, while User 1 expresses despair with white people in general, User 2's focus is more specifically on political rivals, often criticising white liberals rather than white people *per se*.

73: I can't understand liberals and their attitudes supporting a subversive lifestyle. (10)

74: Damn liberal retards can shove it where the sun don't shine! (10)

On the whole, then, this suggests that the user's views may have become more extreme over the course of his career, or, more likely, that he is more preoccupied with the extreme views he already had, making tokens of criticism for racialised and political outgroups more frequent, if not more extreme.

Although, as we have seen, User 2 becomes somewhat more concerned with 'topical' – that is, racial and political – issues later in the career, it is nonetheless clear that his posts to the forum remain largely sociable in nature across both deciles. Throughout the career, the user

engages in a variety of friendly behaviours with other users, including sharing personal details of his offline life (examples 75-76) and praising or agreeing with other members (77-78).

75: She sorta looks like my cousin-inlaw [name]. (1)

76: I live in [state] (10)

77: Good post. I couldn't of thought of a better one. (1)

78: I wish I had \$\$\$,I'd come out to see you. (10).

In line with the user's increased focus on racialised outgroups in the late career, he also shares racist jokes in the final decile of his career.

79: What do you call a black man in a uniform? (10)

80: What do you call a black man in a three piece suit? (10)

The user positions himself as knowledgeable in both deciles of the career. There is no evidence that the user is making increasing claims to knowledge over time, nor decreasing (explicit) claims as seen for User 1. However, these claims are made in different ways across the career. In the early career, the user often positions himself as knowledgeable by offering advice or suggestions to other users on a variety of mostly lifestyle-related topics.

81: You mean AOL as a service will not let you in to [forum]? Have you done a google search or so to let you to [forum]? (1)

82: Hey,better get that done asap. I had one for few months. It is pain! (1)

83: She ran into you cause the snow caused you to slide through a stop sign. She ran into you means she's at fault. (1)

This advice-giving may perhaps be linked back in to the sociable and friendly nature of this user's engagement with the community. However, unusually, the user shows very few examples of advice-giving in the later career, instead using other techniques to showcase his knowledge and expertise. This includes positioning himself as 'pointing out' or bringing new information to others (examples 84-85). Of particular note is the user's increased display of knowledge relating to white nationalism or to the forum, including recommending white nationalist resources.

84: Believe me,the media will do anything, I mean anything to make sure Palin doesn't get media coverage. (10)

85: You do know Rock Hudson was a homo? (10)

86: Usually unapproved group messages are from those who just started posting on [forum] and have less than 50 posts. I hate it when I get those too so you're not alone. (10)

87: I wish something can transpire and move the NationalAlliance into a new direction. Gliebe¹⁴ has got to go. (10)

88: The best newspaper out there is The Nationalist Times American Nationalist Union Don Wassall is the editor. You can get news online for free. (10)

This suggests both a greater awareness of movement and forum issues developed through the career, but also perhaps a greater concern with showcasing this knowledge to perform an experienced identity. However, the user does not seem to be concerned with positioning himself as a gatekeeper of the movement or forum. While he displays movement-specific knowledge, this is usually employed to help or inform others, rather than to gatekeep.

8.3.3 User 3

Like User 2, User 3 shows a degree of stability across his career, in that both deciles are dominated by hostility towards and criticism of other white nationalists. However, this behaviour becomes even more pronounced as the career goes on, with the user focusing primarily on debate with and hostility towards other forum members, rather than the wider white nationalist community.

In the early career, the user addresses a number of discussion topics which are typical of the forum and which we have already observed for other members here, including: the expression of negative sentiments towards racialised and political outgroups; criticism of other, especially liberal, white people; the advocacy of violence; and calls to action for the white nationalist and wider white community.

89: If you take note of what is happening in Paris right now, you'll notice that the savages who are burning down everything and have set a white woman on fire are all non-whites. (1)

¹⁴ Erich Gliebe, former leader of the National Alliance neo-Nazi organisation.

90: As for the Republican Party, if anyone is under the illusion that they will ever do anything any of us would like to see them do [...] it will never happen. (1)

91: you cannot let the apologist, excuse-making liberal whites have the last word (1)

92: But in the end, and I say this because of many reasons; there will only be one way to solve our problems as a race, [...] and that way is a very uncomfortable way. (1)

93: Show up, stand in high numbers in one area and let the community and the klan know that they have your full support. (1)

With the exception of the criticism of racialised groups, which occurs throughout the career, all of these topics receive at most infrequent mention in the later career. Particularly in the case of the latter two themes, exemplified by examples 92 and 93, which advocate for real-world violence or other demonstrations against racialised groups and political opponents, we might consider that the user has in fact become less extreme and more restrained in his discourse by the later career.

However, while the range of targets for User 3's criticism is wider in the early career, it is not the case that he becomes less vitriolic later on. Over the career, the user's focus becomes increasingly dominated by debate and hostility towards other forum members. This is present in the early career (examples 94-95), but to a much lesser extent. In the later career, this is by far the most prominent element of this user's discourse.

94: I see you need glasses too, as one of my parents is from [country] (1)

95: I understand that you're a bit emotional and you feel the need to take shots because you're a self-hater, thats ok. (1)

96: Who do you think you're kidding here (10)

97: People who post here aren't stupid or naive, I know they can see your hypocrisy. (10)

98: your position doesn't hold water, as you're showing bias and alterior motvies. (10)

99: You're a pathetic, cowardly little man who is all talk. (10)

Complementary to this, the early decile reveals frequent tokens of praise for and agreement with other members, which are almost absent later on.

100: I hope alot of people read your posts, its a very important subject. (1)

101: I'm glad I came across your post. (1)

The user also expresses hostile attitudes towards others through gatekeeping of his identities – namely of whiteness, white nationalism, and a national identity which I have anonymised here. While gatekeeping of all three identities occurs in both deciles, it is arguably stronger and more explicit in the later career. In the first decile, the user describes ideal behaviours for these groups – what the ideal member should think or do. This has the implicit function of positioning those who do not meet these standards as lesser members of the respective communities, as well as positioning the user himself as an ideal member.

102: A racist can just see that people like that surely are not white. (1)

103: A true WN doesn't use terms like "wog" or "mick" or "kraut" or frog, we reserve our slander for serpent Jews (1)

104: the great and proud people of [country] who are of white racial stock are part of our beautiful and glorious white race (1)

By the tenth decile, however, the user explicitly excludes people, often other forum members, from these categories.

105: They may have become citizens and call themselves a Spaniard, an Italian, a Frenchman or a German; yet they are not by blood, only by citizenship. (10)

106: There is no way in hell that you have any [nationality] blood. There is no one who is [nationality] or half [nationality] who would say what you have. (10)

107: Oh they're for real alright, they just are not White Nationalists. (10)

He also begins to act as a gatekeeper for the forum itself, suggesting that some users do not belong there. This also takes the form of enforcing forum rules, which state that non-white nationalists should confine themselves to posting in the *Opposing Views* (OV) subforum. This of course has the double function of gatekeeping white nationalism, by suggesting that other members are not true white nationalists. While excluding others, the user also thus

positions himself as an ideal white nationalist and forum member, with the authority and knowledge to police the forum.

108: You obviously are posting on the wrong website and would be better served posting on a site which is more in line to your views. (10)

109: Get over to OV¹⁵ where you belong, you antagonistic, phony WN. (10)

This seems to be part of a wider trend across this user's career whereby he increasingly represents the white nationalist community as divided. Examining the tokens of first person plural pronouns across the two deciles, we find that these pronouns are used to refer to the white nationalist community much more frequently in the first decile.

110: I think what it comes down to is if we can attract people to our cause or not. (1)

111: He is a committed racist and he is on our side. (1)

112: It sure shows the extreme vileness, and true "hate" of our opposition. (1)

This suggests that the user perceives himself to be part of a unified group or movement to a greater extent in the early career. Among the few tokens of first person plural pronouns in the late career which refer to the white nationalist community, several imply division or difference within the community, in particular creating a distinction between 'true' white nationalists and those who do not meet this standard.

113: Beware the ones among us that claim we're "white supremacists" (10)

114: nationalist figures and what they mean to certain but not all WN's as we have many different nationalities on this board. (10)

115: How dare we not like what Hitler did to our people and loved ones and be WN's. (10)

A similarity may be seen here between User 3 and User 1, who also alludes to division within the movement. However, while User 1 explicitly calls for unity between all members of the movement, User 3 instead uses gatekeeping and hostility in an attempt to cast out members who do not meet his standards. As we have seen, User 3's gatekeeping becomes stricter and more focused on internal boundaries (examples 108-109), outlining smaller and

¹⁵ Opposing Views, a subforum where guests, including those who are not white nationalists, are invited to post.

smaller pockets of acceptable behaviour and identities, and becoming more critical of his fellow white nationalists and forum members. This has two implications for collective identity. First, the user's conception of what the collective identity is – or at least what it should be – becomes stronger as the career goes on. On the other hand, the identity that the user promotes appears to become less collective over time, as more and more people are excluded from it.

User 3 appears to be more explicitly assertive of his point of view, and the strength of those views, in the early part of his career. This is demonstrated in a number of ways, including the explicit assumption that his own view is correct (example 116), describing correct predictions he has made (example 117), and citing his personal experience as evidence for his knowledge (example 118). In addition, he emphasises the strength of his feelings by highlighting what he sees as the need for action, even violent action, against racialised and political groups, as I have discussed above (examples 92-93).

*116: The point is, when you know you're correct in what you're saying, (us)
(1)*

117: BTW, I quietly predicted that the media would start the "child abuse" comments when P[russian] Blue¹⁶ was on last week. I was one week too early. (1)

118: The majority of real [nationality], and I know plenty of them, look nothing like Europeans but more like Arabs (but not the dark Arab) (1)

As I have already established, in the later career this user is strongly focused on debating and hostility towards other members. This behaviour of course necessitates the user strongly asserting his own point of view. However, in the late career he claims of knowledge seem to be more concerned with directly disproving others' points in the debating context, often directly addressing their interlocutor with a second person pronoun (examples 119-220). These claims are directly, apparently intentionally, threatening to the face of his interlocutor. In the early career, however, these claims are more general, simply positioning the user as knowledgeable to anyone who might read his posts, rather than as the superior debate partner. As already discussed for User 1 above, this may reflect the user's desire to 'prove himself' as a source of knowledge and authority early in the career.

¹⁶ A neo-Nazi musical duo whose members were young twin sisters.

119: I'm here to tell you, there is absolutely no way he would have picked Foster¹⁷ if he had been leading the pack or had a realistic shot of winning that year. (10)

120: I see the big picture, you don't, even though I explained it. (10)

This sense of direct competition with other users also emerges through User 3's frequent referencing of his interlocutors' previous posts, using these as justifications of his criticisms and thus using others' words against them.

121: I challenge anyone from posters and mods to look at your posts going back the last few months and see if they don't come to the same conclusion. (10)

122: You admitted in post #15 that your knowledge is "secondhand" (10)

The user also refers to his own previous posts in order to support his current argument. This showcases the consistency of his opinion – of his rightness, as he sees it – over time. In particular, example 123 here may function as a claim to a veteran power (Newsome & Grant, in press), as the user refers back to earlier stages of his career, establishing continuity of his views and knowledge over time.

123: I have mentioned this last year or the year before. (10)

124: which is why I wrote what I did (10)

From this discussion, we can see that the major behavioural change which occurs across User 3's career is that he becomes more and more focused on his fellow users as a primary target of criticism and hostility. In light of this, it is surprising that another difference between the early and late stages of this user's career is that the user begins to engage in more off-topic talk later in the career.

125: I played full court basketball with a group of friends in a gym once or twice a week until I was thirty (10)

126: I think he knew way before he got to the NBA¹⁸ that he would have to develop the left as well as the right hand. (10)

¹⁷ Ezola Foster, a conservative African American politician selected as the running mate for far-right politician Pat Buchanan in his 2000 U.S. presidential campaign.

¹⁸ National Basketball Association

127: This is one of my favorites by Stevie Ray¹⁹... (10)

In part, this may align with what could be seen as a reduction of extreme views across time for this user, evidenced by the relative absence of advocacy of violence and other white nationalist demonstrations in the late career. It may also, paradoxically, align with the increased debate and direct hostility towards other members expressed over time. Under this view, User 3 may have begun to see the forum as more of a social space over the course of his career. Of course, the kind of debate evidenced here is not social in the more typical, friendly sense of 'sociable'. Nonetheless, it certainly constitutes an interaction with other members, allowing the user to engage more directly with other users than they did in the early career. The same late-career goal of direct social interaction may be achieved by the more casual, off-topic chat in examples 125-127.

8.3.4 User 4

User 4 is another user displaying relative stability across his career. The user makes frequent reference to elements of his individual identity, namely his old age and Christianity, throughout his career, which appear to influence his white nationalist identity. Nonetheless, the user does exhibit some changes across his career, with his criticism of outgroups becoming more wide-ranging. In addition, like other users, User 4 shows a general decrease in assertiveness in his discourse in the later career.

Like Users 1 and 3, User 4 is generally more assertive in expressing his views earlier in his career, and the views that he expresses are arguably more extreme. Multiple factors contribute to this. First, perhaps the most striking difference between the user's behaviour in the first and tenth deciles is the high prevalence of calls to action for white nationalists and other white people in the early decile. While not absent from the later decile, such encouragement is much less frequent. Additionally, the tone of these posts is markedly different across the two deciles. In the early career, many tokens advocate for more violent or otherwise high commitment actions (examples 128-130), even mocking lower-effort actions such as voting. Later-career tokens, as well as being fewer in number, are also less extreme in nature, generally focusing on the need for white people to 'wake up' to white nationalism, and for white nationalists already involved in the movement to show greater unity in their efforts.

128: I'm afraid that only blood will make these invaders go home and stay there. (1)

¹⁹ Stevie Ray Vaughan, a blues musician

129: We all need to get off our ass and go to D.C. and raise holy hell, or replace the bastards. (1)

130: It would seem that goverment, media, schools, and the churches are all against us. the only way to stop it is to fight, and if that means blood in the streets so be it. (1)

131: Off course if we become that lazy or chicken we could always vote for a 3rd. party that doesn't take much courage. (1)

132: If we could just get the parents and other whites to stand together.... (10)

133: I wish the white folks there would wake up and understand who is doing to them and the whole country. (10)

In addition to the more violent tone expressed in the early decile, the user further strengthens his calls for others to act by making appeals to both history and readers' emotions.

134: I've said it before and I'll keep saying it, if the founders were here they would have already marched on Wasington with a gun in one hand and a rope in the other. (1)

135: our childern and grand childern will ask why we did nothing when we could have, and they will hate you. (1)

The user positions himself as knowledgeable throughout the career. However, like other users, User 4 appears to be more concerned with this in the early decile, producing more tokens and drawing on a wider range of resources to perform this knowledgeable identity. These include using his personal experience and observations as evidence for his claims (examples 136-137) and positioning himself as understanding things that others do not (example 138).

136: I have noticed that for the last 45 years or so, that there has been open warfare against the American people by it's owm goverment. (1)

137: I carry different guns depending where I'm going and so far have never had a problem with the cops (1)

138: don't your people understand where many of the illnesses that humans now suffer from came to us by people doing these kinds of things? (1)

Additionally, contrary to typical expectations of the behaviour of newcomers and longer-standing members, the user in fact seeks information and advice on several occasions in the later career, as well as expressing uncertainty about his statements. While similar features are also found (at lower rates) in the first decile, these are offset by the greater assertiveness expressed there.

139: On the news they said it was nearly a 1,000 yard shot with a AK47, I'm thinking about this, can that really happen? (10)

140: Maybe someone out there smarter than me can put more light on this. (10)

141: I heard this years ago, don't know if it's true or not (10)

As for other users, then, User 4 appears to be less and less preoccupied with appearing knowledgeable as his career goes on. On the other hand, he may feel pressure to 'prove himself' in his early career by expressing greater knowledge, and zeal for the cause and for action.

Unlike User 1, however, User 4 does not straightforwardly begin to rely on his 'veteran power' (Newsome & Grant, in press) later in the career. Forms of what might be called veteran power are drawn on throughout the career. However, it is in the early career where we find 'true' veteran power as described by Newsome and Grant (in press), where the user refers directly to the time he has spent in the movement to position himself as a credible authority within it.

142: In 1965 I went to everyone I knew to rise up and go to Washington and take our country back, but no-one would go with me. (1)

143: Am I ready, you bet. I've been waiting for years. (1)

144: I remember 25 years ago these ivaders were saying that they would overwhelm us with their numbers and then vote us out. (1)

These overt references to his long career may also reflect a desire to 'prove himself' by explicitly identifying himself as a long-standing member of the white nationalist community, despite his short-standing career in this specific forum. Later on, the user positions himself less explicitly as a veteran, by referring instead to his old age.

145: I know I grew up a long time ago, maybe it was another world. (10)

146: But I'm a old dog and can't learn new tricks. (10)

147: In my day he would have got something else. (10)

This user also displays a more pessimistic attitude in his later than his earlier career, including reflections on how the perceived situation for white people and white nationalism has got worse over time. This signals the user's long history of white nationalist thinking in an even more indirect way.

148: This counrty has gone from, 'I regret that I have but only one life to give for my country', to 'What is the meaning of is', or 'We have to pass this bill to know what is in it'. (10)

149: But just when did we let the darkies and queers take over? (10)

In line with the general trend of this user declaring his knowledge and enthusiasm for the movement more clearly in the early career, it seems that this user may feel that his long forum career, combined with less overt references to his long-term investment in the movement, speaks for itself in signalling his authority and prestige within the forum and movement.

As for most users, criticism of both racialised and political outgroups is present at both ends of the career for User 4. However, as for User 2, this is more frequent in the later decile.

150: We have millions of people who have crossed into our lands without permission (1)

151: The Joos working behind the scene to flood the USA with the invaders (1)

152: Have you ever noticed that no matter which one of the two major parties that get into power, things only get worse for America (1)

153: To all queers that are looking in to see what we the normal people think of you. You are not liked and we all wish you would go away (10)

154: this is the problem with blacks, they want to be like us but can't. (10)

155: I'll just bet that some joo was in charge. (10)

156: Government could ban smokes because they cause sickness and death, but they won't, because there is too much money in it. THEY DON'T CARE ABOUT YOU! (10)

157: the FBI can't find it's own butt with both hands. (10)

158: These folks²⁰ will be comming home one day. but don't worry they won't take your guns, rape your daughter, or shoot you, or will they. (10)

As these examples go some way to demonstrate, the user focuses on many more targets for his dislike and criticism in the late decile than he did in the early. Relatedly, like User 1, User 4 also begins to engage in some criticism of other white people in the later decile.

159: Maybe that is the trouble with white folk, they don't speak up when we see something that is wrong. (10)

160: Why do our own people mess with us, it's no way to stick together (10)

However, this tendency is not as marked as for User 1, nor does this behaviour largely replace criticism of outgroups, as it does for User 1, instead representing yet another new target of the user's disdain in the later decile.

The finding that User 4 seems to have negative feelings about more social actors in the later career may seem inconsistent with the idea suggested above that the user is more assertive and – perhaps – extreme early in the career. However, it is consistent with Berger's (2018) hypothesis that as an individual becomes more involved in an extremist group, they typically develop negative feelings about more and more actors as their own in-group becomes more closely defined. This is reflected in User 4's increasingly negative views of white outgroups, such as U.S. government agencies, homosexuals, and non-extremist white people, as his career goes on. However, while the user may appear more typically extreme in this way, I argue that while these negative feelings have become broader in their focus, encompassing more targets than in the early career, there is no evidence that these feelings have become deeper or more extreme, or at least that the user expresses them in a more extreme way. In fact, the opposite appears to be true, as the prevalence of advocations of violence in the early decile attests to. Nevertheless, this finding shows another way in which the user has changed across his forum career, perhaps influenced by other members and discourses on the forum to distrust a wider range of social actors.

²⁰ U.S. soldiers.

Christianity is clearly an important facet of this user's identity, and he makes frequent reference to it throughout the career. In the early career, this largely takes the straightforward form of quotes from and reference to the Bible and to the Christian God. In the later career, however, the user increasingly turns to gatekeeping of Christianity and criticism of other Christians, in turn positioning himself as a good Christian more explicitly and more frequently than in the early decile.

161: I believe in God and His laws, learn this, ex22:2 it's the law. (1)

162: Paul has shown us that we must know both God's law and the law of the lands we live in (1)

163: My prayers are with you lads. (1)

164: Mods forgive me, but I'm tired of these misinformed Christians. (10)

165: No one will ever fly away, Ez. 13:20, I wish someone would read the book. (10)

166: But he is saying things that you will never hear in most Christian churchs that should be heard. (10)

This increased interest with positioning himself as a good Christian stands in contrast to the user's apparent lack of concern with explicitly performing an established or veteran white nationalist identity in the late career, as discussed above. There are several possible explanations for this apparent discrepancy. The first is that this increased criticism of other Christians and Churches is linked to the general increase in criticism of institutions and other white people in the later decile, with these actors simply representing another new target of disdain. Another possible interpretation is that the user has become more defensive of his own version of Christianity over time, and therefore more likely to criticise others to protect and bolster his own position. While rejecting the label, User 4 expresses views that are similar to, and sympathetic towards, a subsection of the white nationalist movement called Christian Identity. Christian Identity is a racist interpretation of Christianity which claims that only white people are descendants of Adam, and that Jewish people and others perceived as non-white are instead descendants of Cain, who is understood to be the result of a union between Eve and the Serpent (ADL, 2017).

167: most say that I am CI [Christian Identity], just because I know who my fathers were. (10)

168: In the beginning there were no jews. I wish people could read the book and know what it says. (10)

169: Looks like someone could use a long chat with a CI pastor. (10)

Christian Identity is a controversial ideology even among white nationalists; indeed, User 3 uses this as a means by which to criticise other users, and even suggests that Christian Identity adherents are not true white nationalists:

170: You can't shine Buchanan's²¹ shoes you ci flunky. (User 3, 10)

171: Now why don't you go defend the clownish ci movement in the ci forum with your buddies? (User 3, 10)

172: Intelligent and thoughtful WN's aren't interested in ci kookery. (User 3, 10)

Having learned by the late career that the forum is not an altogether sympathetic space to his particular beliefs, User 4 may therefore feel more inclined to defend his own version of Christianity more strongly. By criticising other Christians, he positions his own version of Christianity, and by extension himself, as superior to others.

As is the case for other users, most notably User 2, there is evidence that the forum has a social as well as an ideological function for this user. Most notably, he engages in off-topic discussion of lifestyle and travel throughout his career.

173: I love and respect the Irish people and their culture and that is one of the reasons I would travel there, it was pure magic. (1)

174: If you ever go to NH, go to North Salem route 93 and you will find a place called America's Stonehenge (1)

175: There are some things that help you live longer, but in the end it's going to kill you. If you smoke stop. (10)

176: Look at the label on these supplements, they say with a proper diet and exercise you will lose fat and weight. (10)

As these examples showcase, this off-topic talk often has an advice-giving function. This is part of a general pattern across all five users whereby advice-giving tends not to be directly relevant to white nationalist topics (although is often tangential, e.g. advice about firearms).

²¹ Pat Buchanan, a far-right politician who ran as a nominee in the 2000 U.S. presidential election

For User 4, advice-giving tokens appear somewhat more frequently in the early decile, suggesting that despite the off-topic and friendly nature of this advice, it may still form part of an effort by the user to position himself as knowledgeable in the early career. The user also engages in other forms of friendly engagement throughout the career, including well-wishing (particularly in the early decile) and a joking or mocking tone (particularly in the later career).

177: I wish you the best, and they will be in my prayers. (1)

178: if we never meet know that we are brothers in more ways than you think. (1)

179: I know black paper with white ink, no, that would be racist. (10)

180: Are you sure we can still use the word boy anymore? I mean no one wants to be a racist, well almost no one. (10)

As we will see, this tendency to take a more ironic tone later in the career is also shared by User 5. Perhaps this is symbolic of a greater sense of comfort and ease within the community, as well as an increasingly sociable approach to forum membership. For User 4, this also links back to the focus on a wider range of outgroups as the career progresses, as a range of topics, including racialised outgroups and so-called 'PC culture', are frequently the subject of these mocking comments.

8.3.5 User 5

Among the five users investigated here, User 5 is perhaps the one who changes the most across his career. Across the two deciles, the focus of his discourse shifts drastically as he apparently re-evaluates who the enemy is and who is victimised by that enemy. In the early career, he performs a neo-Confederate identity which strongly opposes the U.S. federal government. Later, he displays a more typical white nationalist identity which recognises a wider range of racialised and political outgroups as a threat.

Early in the career, by far this user's most defining trait is his neo-Confederate identity, which manifests in multiple ways. In particular, the user showcases his knowledge of Confederate figures and history, declares pride in the Confederate heritage of his state and in his own Confederate identity, and shows support for modern-day neo-Confederate ideals such as new attempts at state secession.

181: State Secession or outright rebellion is the only remedy for your "freedom". (1)

182: As was stated from the beginnings of the Confederacy. We only wish to be left alone. (1)

183: Most of the movements to contact that are used in today's infantry can be traced to Gen. Lee, Gen. Jackson and Gen. Forrest.²² (1)

184: Technically, we still have a Confederate Government. It's seats are vacant, but the Southern National Congress is active at this very moment. (1)

185: There's not a prouder stock of people that I'd cast my lot with, than those who tread on the road to Independence. (1)

As a result of this identity, the user is clear in the early career that the U.S. federal government is the primary threat to his in-group, which he understands to comprise his fellow white southerners. In line with this, the user positions white southerners and their culture (which is apparently equated with the Confederate heritage of their states) as victims of oppression by the federal government, whose rule over the former Confederate states the user perceives as illegitimate.

186: May we one day reclaim our customs and culture that has been taken from us. (1)

187: And now Southern people are being overtaken by transplant northerners that will accept the "dark ones" among them. (1)

188: I believe that the world would be a more peaceful one without the empire we call the u. S government. (1)

189: If people would read the un-ratified 14th amendment²³ in it's entirety. They will see that we are mere subjects of the u. S. government. (1)

The user, however, remains defiant against this perceived enemy and foresees a future where the southern states can secede and – as the user sees it – reclaim their sovereignty.

²² Robert E. Lee, Thomas Jonathan 'Stonewall' Jackson, and Nathan Bedford Forrest, generals of the Confederate Army during the American Civil War

²³ The 14th Amendment to the U.S. Constitution grants U.S. citizenship to any person born in the country. It was intended to bestow citizenship rights on formerly enslaved people following the Civil War. The former Confederate states were forced to ratify the amendment in order to regain their right to political representation.

190: We will have much to gain for each other with the impending collapse of the empire. (1)

191: Granted, it's not Sumter²⁴, but we have a chance to reclaim our Liberty lost 146 years ago. (1)

Despite User 5's strong identification with a white, southern identity, he mirrors the other users investigated here by engaging in criticism of other white people, although as we will see this is limited to the early career in the case of User 5. For this user, this largely takes the form of criticising other southern white people who support the federal government, who the user sees as complicit in their own continued oppression.

192: I ask them why do you allow the u. S. flag fly in front of the churches, knowing full well that the display of that wretched rag implies federal ownership? (1)

193: Most that receive "benefits" from their masters in DC. Care not if you or I suffer the ransom. (1)

In the later career, while the theme of the user's mistrust for federal authorities persists, this diversifies hugely, with a much wider range of threats to the user's ingroup identified. This is similar to the phenomenon observed for User 4. Most of the new targets of disdain mentioned by User 5 in the late career are adjacent to the U.S. federal government, but new targets are also identified which received only scant mention in the early career, such as the media and bankers.

194: I'd bet a dollar to a doughnut that not one of the neo-cons and warmongers could answer these simple questions. (10)

195: Civil Rights law(s) will be foisted upon the majority by the judiciary. (10)

196: The police department will hand over the "investigation" to it's PR lackeys and lawyers. (10)

197: I've noticed on my local zio-con radio station the jew Suze Orman is telling little old ladies to put their money back into the bank instead of their shoeboxes. (10)

²⁴ The Battle of Fort Sumter triggered the American Civil War.

198: I don't carry the water for any group of lobbyist that can whisper sweet nothings in politician's ears. (10)

Uniquely among the five users, User 5 engages in very little criticism of racialised outgroups in his early career – example 187 is a rare example of such sentiment, and arguably places more emphasis on non-racist white people than on the racialised group – with the U.S. federal government practically the only perceived outgroup to be recognised as a threat to (southern) white people. This changes in the later decile:

199: only a filthy greedy jew would demand payment for protection (10)

200: So it seems to me, if she²⁵ is that "thin skinned" in majority White America, she should seek out other thin skinned tribes..... IN AFRICA!!!!!!! (10)

201: And this picture of the Japs²⁶, shows that Patton's²⁷ description of their sty-like living conditions haven't changed 65 years or so. (10)

The racism expressed in the later decile has a number of facets, suggesting that the user is much more preoccupied with questions of race than he was early in his career. For instance, while the user expresses pro-gun sentiment in his early career, by the late career this position appears to have developed to being pro-gun for white people only.

202: My permit is the 2nd Amendment.²⁸ (1)

203: Open carry²⁹ would only work "peacefully" in an ALL WHITE society. (10)

Relatedly, the final decile sees the user mention conspiracy theories, especially antisemitic conspiracy theories, much more frequently than he did earlier in his career. These theories reflect a belief in a (Jewish) 'elite' who work to ensure regular (white) people remain in an inferior social position.

204: If this ACT OF WAR Bill is not approved in the senate by the end of the year. We will look for scandals and plane crashes to occur in short order. It's the Mossad's way. (10)

²⁵ Mexican-Kenyan actress Lupita Nyong'o, who had spoken out about the challenges she had faced due to her dark skin.

²⁶ Jewish American Princesses

²⁷ George S. Patton, a U.S. general who oversaw post-WWII denazification efforts in Germany despite his strongly antisemitic views and professed respect for the Nazi Party.

²⁸ The 2nd Amendment to the U.S. Constitution grants citizens the right to bear arms.

²⁹ The act of visibly carrying a firearm in public places, permitted in most U.S. states.

205: That Americans would forget about all the treachery committed in our name for his masters in IsRaEl. (10)

206: Once again we witness the slimy tentacles of the international jew prohibiting any competition to their Ponzi rackets, known as the federal reserve, IMF, World Bank, BOEngland and Bank of International Settlements. (10)

Hand-in-hand with this shift in perceived enemies, the user also appears to change his perception of who is victimised by these enemies. While in the early career white southerners are represented as being under threat from the U.S. federal government, in the later career regular (non-elite) white people are seen as being victimised by this range of new perceived enemies.

207: It seems to me that there is an elite group of controllers that operate outside of societies norms. In other words. A super duper god like entity that robs, kidnaps, assaults and murders at will. (10)

208: You're a smuck, backwoods, no tooth hick to be fleeced each and every payday. (10)

209: And as we seen in Cyprus. It's not the big wheels who lose their money. It's the little old lady who worked her a\$\$ off at a company for 30 years. (10)

Interestingly, this change in the user's understanding of victimised parties coincides with a lack of criticism of white people in the later career, presenting the opposite pattern to User 1. In the early career, by contrast, white southerners were criticised for their alleged complicity with the U.S. government. This may derive from the user's new outlook which places white people in an altogether more helpless position, victimised from all sides by political enemies and racialised groups, which are seen to work covertly against white people's interests. On the other hand, as we have seen (examples 190-191), in the early career the user appears to believe that the fall of the U.S. government and the resurgence of the Confederacy are inevitable, and therefore that white southerners will be freed from their alleged oppression. The user may therefore see uncooperative white southerners as an impediment to this. In the late career, however, white people are increasingly seen as unable to escape their fate, and therefore unworthy of criticism.

In the later career, these newer targets of criticism – political or other authorities, and racialised outgroups – are often the subject of mocking remarks or jokes, targeting so-called

'PC culture' and language, and trivialising violence faced by racialised groups (examples 210-212). This is in stark contrast to the early career, where User 5 spoke of the target of his criticism – the U.S. government – in much more solemn and serious tones (examples 213-214).

210: And we all know that the very worst thing that could happen to a White man is being called a racist or bigot. (10)

211: Maybe we can start calling bank robbers the withdrawal community. (10)

212: Remember that little incident between Zimmerman and Trayvon "skittles" Martin?³⁰ (10)

213: The media will pound the masses with the evils of Secession. For this strikes at the very heart of the empire's existence. You will see more of this active separation shortly. The legs of the beast are shaking from the top heavy strain of bureaucrats and usurpers. (1)

214: True, the cannons are silent, but the shadow of oppression still lingers in every nook and cranny of our Southern homeland. (1)

As for User 4, this may signal a more ironic, comfortable, and sociable approach to ideological talk in the later decile. Given the wider range of targets of criticism in the later decile, it may also signal that, like User 4, negative sentiments towards outgroups have become *broader*, but not necessarily *deeper* and more extreme, through the user's membership in the forum.

The wider range of outgroups referenced in the later decile, particularly the new focus on racialised groups, brings User 5 much closer in line with the ideologies expressed on the forum as a whole. Neo-Confederacy is a relatively niche ideology on the site, and the user mentions that he has received criticism for it from other users.

215: Another thread on this site asked "What would you do to change the u. S. constitution?" I said. Replace it with the Confederate one. I was jeered at. (1)

³⁰ Trayvon Martin was an unarmed African American teenager who was shot dead while walking home by George Zimmerman, who claimed that Martin looked 'suspicious'. The Skittles sweets that Martin had purchased just before his death became a symbol of Martin's childhood innocence.

The user has therefore become more mainstream – by the standards of the forum – over time.

As the user moves closer to a typical white nationalist ideology from a more specific neo-Confederate focus, we also begin to see less of the user's individual identity reflected in his posts. In large part, this comes from no longer mentioning their neo-Confederate identity. This was very frequent in the early career, where the use of first person plural pronouns (e.g. examples 182, 184, 186, and 191) made it clear that the user identified strongly with fellow Confederates and residents of former Confederate states, both past and present. Reference to neo-Confederate identity, or to the Confederacy at all, are almost absent in the later decile. However, decreased reference to a neo-Confederate identity is just one part of a wider pattern. The user provides much more personal information, and narratives about his own life, in the early decile. In particular, personal information is almost absent from the late decile, although there are many examples from the early career.

216: My sister passed away from [condition]. (1)

217: Visualise that my driveway is over 1/4 mile long. And 3/4 of the way down is a gate with posted signs and house phone number. (1)

218: I had an office outside [city]. (1)

Similarly, while some life narratives appear in the later decile, these are more frequent in the earlier career. Functionally, these narratives play similar roles across the two deciles, providing evidence for claimed knowledge, sharing conspiracy theories, or positioning himself as a target of attacks or surveillance.

219: And told them I would tell everyone I know about Rite Aid supporting a racist/hate group. (1)

220: I've made a flyer, and have been passing them out to people that have DOTS on their mailboxes. (1)

221: Last week a deputy sheriff got right on my bumper. Then dropped way back. I think my Secession sticker spooked him (1)

222: When he seen me he started to babble about not knowing about the off limits upstairs. It took all of my energy from shooting this filthy vermin on sight. (10)

223: I TOLD HIM!!!! I would shoot him in the face if he came to my door to take any of my personal property (10)

224: I reached for my small “pouch”³¹ and rested my hand. His eyes became big as cup saucers and adjusted his gate and direction immediately. (10)

In the early decile these narratives tend to highlight actions and activism the user has carried out to advance the white nationalist cause. In the later decile, these narratives take on a more physical or violent nature, recounting occasions where the user has threatened others with violence. This trend may suggest that the user is becoming more violent and aggressive as the career goes on, in contrast to the pattern observed for Users 1, 3, and 4.

As for most other users, there is no particular evidence that User 5 is concerned with presenting himself as more knowledgeable as the career goes on. There are a similar number of tokens across the two deciles where this user positions himself as knowledgeable.

225: I was my companies armorer in the Army. The M60MG was being phased out because of the introduction of the M249 SAW. (1)

226: Only an order of Dissolution of the federal government by Constitutional Convention (34 States) or independent Secession resolutions of the States, can we truly reclaim sovereignty. (1)

227: You can tell a bite from a sting by the shape of the incision. (10)

228: There’s only so much you can do with felled cows/bulls. The human consumption of such animals are regulated to jerky and hot sausages. (10)

If tokens of advice-giving are also taken into account, the user may be considered to position himself as more knowledgeable in the early career, as these are more frequent here.

229: But please do not choose hopelessness. I would not pursue full time employment, until you feel like you can function as comfortably as possible. (1)

230: But if it's stabbing and slashing you want. I recommend the Rex Applegate/WE Fairbairn (German Boker-sousen) double edge, SOG Daggert1³² (1)

As I have discussed, however, such tokens may also serve a sociable or friendly, not solely authoritative, purpose, as example 229 exemplifies. Perhaps also refuting the hypothesis of

³¹ Gun holster

³² A variety of combat knife.

greater authority in the early career, the user also seeks advice or information, or positions himself as uncertain, more frequently here.

On the whole, we can conclude that like the other four users, User 5 does not seem to be more concerned with appearing knowledgeable as the career goes on. However, while other users apparently make a greater effort to showcase their knowledge in the early career, perhaps in an effort to 'prove themselves', this does not appear to be a particular concern for User 5 in either decile.

8.4 Discussion

The findings presented in this chapter show that no two users are completely similar, with each bringing some individual elements to the construction of their identity across the career. There are no discursive themes which appear for all users at both career deciles investigated here. Unsurprisingly for a white nationalist forum which explicitly concerns itself with racial issues, the most widespread theme is that of negative sentiment expressed towards racialised outgroups, which is frequent in all possible user/decile contexts except for User 5/Decile 1. Other very common themes include negative sentiment towards political outgroups, which is frequent in eight of ten possible user/decile contexts. This is also unsurprising in the context of a white nationalist forum. As I have shown elsewhere (Booth, 2023), despite the claimed racial focus of the forum, political or ideological outgroups often receive as much, if not more, negative attention than racialised groups. Other common themes are the provision of advice and suggestions and positioning of the self as knowledgeable, which each appear in eight of ten possible contexts. Analysis of each user shows themes which are unique to them, showcasing the individual identities which the users bring to the community and which survive assimilation to a collective identity. We do not, then, observe the complete *depersonalization* (Hogg, 2012; see §2.5.2) of individuals which may occur as an individual becomes more involved with an extremist group.

Nonetheless, we can identify an important high-level similarity between all five cicada users investigated here; namely, that all five show evidence of change in how they engage with the community across their career. Further, some of these changes show trends in the same direction across users. This offers support for the validity of the career as a valid model through which to understand online engagement with extremist movements over time, implying that members may pass through different stages, marked by different (linguistic) behaviours, and that different users may converge over time, assimilating to what we might understand as a collective identity. Of course, further research considering a greater number of users would be required to demonstrate this claim more strongly.

Themes that show a general increase in frequency across the career include negative sentiment towards the media and/or Hollywood, appearing for two users in the early career and four in the late; negative sentiment towards other white people, appearing for one user in the early career and three in the late; a mocking or joking tone, which is not frequently used by any early-career users but is adopted by two later on; and off-topic talk, including discourse on lifestyle and popular culture, which also appears for one user in the early career and three later. Themes which generally decrease in frequency over time are the giving of advice and suggestions, which appears for five users in the first decile and three in the tenth; negative sentiment towards political outgroups, which also appears for five users early on and three later; and calls to action, which represents the most significant difference across the career, appearing frequently for four users in the early career, and only one in the late career.

On the whole, each of the users explored here become less overt, extreme and assertive in their discourse as their career progresses. This general statement has two major strands. The first derives from the fact that the most significant difference across all five users' careers is that these users tend to rally their fellow white people and white nationalists to action much less frequently in the late career. Hand-in-hand with this, as we have seen above, several users show a general decrease in advocacy of violence in the later career. Together, these two trends suggest that the five users become less extreme across their careers, in the sense of being less inclined to violent or other extremist acts (McCauley & Moskalenko, 2017). The decrease in calls to action also has other implications, especially when considering the general increase over time of criticism of other white people. Taken together, the trends suggest that these users have become jaded with the movement and its chances of success in its goals, and are therefore less proactive about attempting to achieve these goals. Because of their jadedness, they may no longer see the point of encouraging action, seeing it as fruitless because other white people, who are not awake to their own alleged oppression, will not get involved.

The second strand relates to the decreased advice-giving over time, which suggests that these users have become generally less interested in positioning themselves as knowledgeable and asserting their own authority. This is also supported by the changing expression of knowledge over time. Although this remains a major theme in most user/decile contexts, and does not show a marked decrease over time, the finer analysis presented above shows that there are generally fewer or similar numbers of tokens in the later career, and that in the early career the users employ more methods to appear knowledgeable. This suggests that these users become less interested overall in positioning themselves as knowledgeable, or at least that appearing knowledgeable is not a growing concern across

the career for these users as the literature would lead us to expect (e.g. Ang & Zaphiris, 2010; Danescu-Niculescu-Mizil et al., 2013; Newon, 2011). These users may feel that their long-standing career speaks for itself without them needing to further demonstrate their credibility, and no longer feel pressure to 'prove themselves' later in the career as they may have done earlier on.

Another general trend is that the five users' engagement with the forum becomes more social in nature over time. This is evidenced by an increase in joking and mocking behaviour across the career. While much of this content contains ideological elements, including racist jokes, the tone of this engagement is nonetheless more ironic than is typical in the early career, and suggests that the members are increasingly interested in having fun as a part of the community. The increase in off-topic talk also speaks to an increasingly social engagement with the forum, as this discourse is often entirely non-ideological, or with only tangential connections to white nationalist ideologies. Elsewhere, Angouri and Sanderson (2016) claim that off-topic talk serves to build rapport between members.

An increase in sociality may also take non-traditional forms. While User 3 becomes increasingly unfriendly as their career goes on, the targets of his hostility are often his fellow forum members. This reflects a greater interest in one-on-one interaction with other members, albeit of an unusual kind. This trend suggests that, while users may be brought to the forum by a desire to find ideologically similar individuals and to discuss these ideologies, later in the career they may seek a more social environment, less informed by – although not void of – ideological talk.

A final general trend observed across all five of the careers is that the users are typically more outward-looking earlier in the career, and more inward-looking later on. This manifests in several different ways. The first is in the targets of the criticism and negative sentiment that these users express across their careers. While such attitudes are generally expressed towards racialised outgroups throughout the career, posters' criticism of their fellow white people tends to become more prominent as the career goes on. Clearly, these users are casting a more critical eye closer to home than they typically did in the early career. This tendency is particularly striking for User 1, for whom criticism of racialised outgroups is largely replaced by that of fellow white people and even white nationalists. Secondly, we might understand the decreasing number of examples of calls to action over time to reflect a more inward-looking focus. While early in the career the five users frequently encourage taking their ideology offline by engaging in real-world action, this behaviour is generally much diminished over time. By the late career, most of the users appear to be content simply to post to the forum, rather than trying to rouse further action. Finally, and related to

the previous point, we find the increased sociality that tends to occur across the career. These users appear to have more interest in the social function of the forum as the career goes on, engaging more frequently with other users and engaging with the forum in a way that does not involve the ideological discussion of outgroups.

The findings presented here are of course limited. I have investigated only five cicada users in detail, and there is no guarantee that similar patterns would be observed for other users – either other cicadas, or members following other career paths. However, it is promising that I have been able to identify some general trends across these five users, which may also be applicable to others. In the next chapter, I draw on these findings, along with those of Chapters 5, 6, and 7, and further reflect on these in the light of my research questions (§3.8.3) and the broader context of this research.

Chapter 9 Discussion

9.1 Thesis summary

This thesis has addressed the following research questions, which were formulated in Chapter 3 with reference to existing literature and to my inductive research design process.

What does the career look like in the context of this online white nationalist forum?

RQ1: *How, if at all, can the structure of the career be conceptualised in a stable and reproducible way?*

RQ2: *How far are the individualised vs. structural approaches to defining the career supported in this forum context?*

a: *From a temporal perspective, to what extent, if any, can a typical career path be said to exist?*

b: *To what extent, if any, are longer careers associated with the accrual of linguistic capital in the form of a wider repertoire of linguistic resources? What resources are acquired, if any?*

c: *To what extent, if any, do members appear to radicalise across the career? For example, through more extreme boundary drawing and dehumanisation of out-groups, and/or greater acceptance of violent means to achieve ideological ends.*

RQ3: *How does the career observed in the forum intersect with the development of collective identity, in terms of its three elements (boundaries, consciousness, and negotiation)?*

I addressed RQ1 in Chapter 5. In the sociological literature, the career has primarily been discussed in an abstract way, which offers little guidance to researchers aiming to apply the career concept in new contexts. In Chapter 5, I took a novel approach to systematising the literature which describes career roles and stages, identifying general principles for understanding the structure of careers. I found that behavioural factors and temporal factors (such as the duration of the career) are the most important elements to consider when studying the career, together offering a more comprehensive understanding of careers in a given context. These elements can be useful for the study of careers both offline and online, although there are finer differences in how best to achieve this in different contexts.

The sequence analysis presented in Chapter 6 investigated careers from a temporal perspective (RQ2a). I showed that, while many different career paths are attested in the data, there is also a high degree of homogeneity among users, with most forum members following one of a small number of different career paths. These results support a view of careers that falls somewhere between the individual (Fillieule, 2010) and structural (Becker, 1973; Goffman, 1961) approaches, recognising both homogeneity and heterogeneity in users' careers (RQ2). I therefore identified not one, but four broad career paths based on duration. The shortest career path – the *mayfly* career comprising only a single calendar year of posting – is by far the most common, accounting for 76% of user accounts. A further 10% of members each follow the two year *dragonfly* career path, and the *mantis* career path of between three and six years. Finally, the *cicada* career path of seven or more years (up to eighteen years) accounts for only 4% of users.

The large-scale quantitative linguistic study presented in Chapter 7 found that general trends exist across career types in the usage of a number of linguistic features (RQ2b). Shorter mayfly careers are more strongly associated with the use of second and first person singular pronouns. They also favour epistemic stance complement clauses controlled by verbs and nouns. Contrary to expectations based on the literature on language use in the early career, these stance resources typically convey *certainty*, or heightened epistemic commitment. On the other hand, longer cicada careers are more closely associated with first person plural pronouns, progressive aspectual verb constructions such as *continue + -ing* or *carry on + -ing*, modal verbs, and epistemic stance complement clauses controlled by adjectives.

Later in Chapter 7, I carried out a quantitative analysis of usage of these key linguistic features across the careers of five individual cicada users. While I hypothesized that the 'cicada features' – those associated with the long-term cicada career type as a whole – would be more frequent in the later career, this was not the case, and most features showed decreasing usage across the career. In other words, no features were identified which are strongly associated with the late stages of long careers. This suggests that users are drawing on a larger pool of resources in their late career, meaning that no one feature shows particularly high rates of usage (RQ2b).

In Chapter 8, I conducted a qualitative analysis of the linguistic behaviour of the five cicada users. This revealed a degree of individual variation, with no features unifying all five users at all times. However, all five users were shown to change their discursive behaviour across their career, and a number of general patterns were identified (RQ2). In line with the greater certainty expressed in mayfly careers, cicadas' discourse became less assertive and aggressive as their careers went on, with fewer overt displays of authority and expertise.

This also manifested in less violent discourse, suggesting a less extreme viewpoint (RQ2c). Discourse also became somewhat less ideological, with more 'off-topic' and ironic interactions observed in the late career. This suggests that long-term users are not solely motivated by ideology, but by the social relations they have established with the community. Where ideological discourse did appear in the later career, its focus had typically shifted from the early career. Early-career discourse was characterised by criticism of racialised out-groups. Later in the career, groups closer to the deictic centre of the extremist group were criticised, including non-racialised political outsiders, other white people, and even other white nationalists. This appears to represent an increased concern among members with gatekeeping their identities (RQ2b) and demonstrates a new conception of the boundaries of the group (RQ3).

9.2 Reflections

9.2.1 Collective identity and symbolic capital

The career is characterised by a move towards a collective identity. This involves the development of symbolic capital, affording members communicative competence in the context of the group.

In Chapter 8 I identify that, while retaining a degree of individuality, the five users investigated do begin to show greater similarity across their careers. We can therefore conclude that these users do move towards a shared collective identity, comprising shared meanings and shared understandings of, and orientations towards, Self and Other (RQ3).

In §3.6, I identified Taylor and Whittier's (1992) work describing elements of collective identity as a useful framework to understand how a user's discourse across the career might reflect their assimilation towards a collective identity. These three elements include *boundaries*, how members distinguish between their in- and out-group(s); *consciousness*, members' understandings of the social relations between in- and out-group(s) and their social position in relation to one another; and *negotiation*, the "politicization of everyday life" (Taylor & Whittier, 1992, p. 122), which concerns how everyday practices and symbols (such as linguistic resources) are used to reflect members' identification with the group. I will begin this section by characterizing the collective identity of this forum in terms of this framework. There is significant similarity between *boundaries* and *consciousness* in this context; in Goffmanian terms, both concepts concern how members' *footings* towards different actors change across the career (Goffman, 1981). There will therefore be overlap in these discussions.

Contrary to our expectations of collective identity developed through a career, there is evidence that most users may share a similar consciousness from the beginning of their careers, demonstrating a similar understanding of the relations between social actors. Unsurprisingly for a racist white nationalist forum, a dichotomy is constructed between the white in-group and the racialised out-group. The out-group is positioned as a significant threat, even an existential threat, to the in-group. Both footings described by Zickmund (1997), of racialised out-groups as conspirators and as social contaminants, are present in the discourse. White people are seen as the victim of violence and even eventual replacement by these groups. The ubiquity of these footings from the beginning of users' careers may lead us believe that members do not assimilate to a collective identity throughout the career.

User 5 serves as the exception that proves this rule. In his early career, he displays a consciousness which is unusual even by the standards of the community. Contrary to the norm, discussion and criticism of racialised out-groups is almost absent from User 5's early-career discourse. Instead, his neo-Confederate consciousness positions the U.S. federal government as the primary threat to his claimed in-group of white people in the former Confederate states. By the late career, however, he exhibits a consciousness much more closely aligned with that of the community as a whole, in particular partaking in conspiracy theory discourse that blames racialised groups for white people's perceived misfortune.

This therefore suggests that the consciousness outlined above, of white in-groups under threat from racialised out-groups, is characteristic of the group's collective identity, and that User 5 assimilates to this identity throughout his career, evidenced through his change in footing towards various actors over time. In this light, the apparent lack of consciousness shifting among the other members can be understood as them already sharing this view of in- and out-groups by the time of their arrival. This may reflect the online context of this group and collective identity. We have seen that offline, individuals may become involved in groups because of their social networks, rather than because of a pre-existing commitment to a group's ideology or consciousness (Ahmad, 2016; Blee, 2002; McAdam, 1986). In the more impersonalised and anonymous online space, it may be more likely that newcomers sought out the site because of views they already held, making it less likely that the views they encounter on the site will challenge and shift their existing consciousness. Rather than this unchanging consciousness being seen as antithetical to a collective identity, then, it can be understood as a feature of that identity.

Despite this level of stability, there is evidence of new boundaries and consciousness developing across users' careers. While the early career primarily focuses on racialised out-

groups, late-career users take a more inward-looking approach, offering greater criticism of their fellow white people, white nationalists, and forum users. This is in line with Berger's (2018; see §2.5.3) claim that extremist groups may begin to "treat eligibility as obligation" (p. 63), treating members of their extended in-group as a "functional out-group" (p. 63) due to their failure to embrace extremist ideology. This suggests a shift of consciousness concerning who is perceived as the cause of harm to users' in-group(s), with greater blame ascribed to these groups for their alleged complicity in their own downfall. This also indicates the development of new boundaries. In addition to that of race, a distinction is now made between 'good' and 'bad' members of the in-group. In line with this, the enforcement of boundaries appears to become more of a preoccupation for members in their late careers, perhaps reflecting a greater consciousness of the perceived damage others that are nominally members of the in-group may do to the movement.

Overall, then, we see evidence of collective identity in the form of both pre-existing and new shared understandings of *boundaries* and *consciousness*. These are evidenced through consistent and shifting footings taken towards actors both beyond and within the white and white nationalist in-groups. There is also some evidence of *negotiation*, which concerns the redefinition of everyday practices and linguistic resources with movement-related meaning. This is demonstrated through the apparent re-evaluation of the forum as a more social, rather than simply ideological, space. Over the career, the users investigated here typically begin to engage in more 'off-topic' talk, discussing lifestyle and popular culture. They also take a more ironic or mocking tone in their discourse. This indicates a greater integration of the movement with members' everyday lives, as they begin to turn to the forum not only for ideological matters but also for everyday social talk. In this way, late-career members bring more of their identity to the forum, allowing non-ideological elements to become part of their overall identity performance in the forum. This imbues these non-ideological elements of their identity with movement-related meaning.

However, despite what Taylor and Whittier's (1992) description of *negotiation* leads us to expect, and the emergence of discursive similarities outlined above, the five users investigated in depth here do not appear to converge on specific linguistic forms which can be said to index a shared identity. On the contrary, my findings suggest that there may be no linguistic features which are particularly associated with the late stages of long careers, and therefore with a collective identity. Instead, late-career members appear to draw on a wider pool of resources and showcase greater individuality. I contend that users develop more symbolic linguistic capital as their career goes on, which gives them greater freedom to choose from a larger repertoire of linguistic resources. The features identified in Chapter 7 are those most closely associated with the more limited repertoires of the early career,

including both mayfly careers and the early stages of cicada careers. This explains why certain features, although associated with longer cicada careers, do not show greater usage over time within these careers. In the late career, identity performances are less constrained and draw on a wider range of resources.

The capital developed across the career allows late-career members to draw on atypical or non-normative resources for their identity performances without threatening their own face. Early in the career, when members have little to no capital, they are more likely to feel the need to impress others and prove themselves as a legitimate member of the community. Newcomers who behave in non-normative ways may be perceived as misunderstanding or disrespecting the norms and values of the community, and therefore be viewed unfavourably. Late-career members, however, can act in non-normative ways without experiencing a loss of face. Their capital – which is both symbolic, reflected in their discourse, and institutionalised, reflected in the *Join Date* field displayed next to their posts – affords them “the power to ignore the rules” (Graham, 2016, p. 313), and therefore a greater degree of agency and choice in their identity performance.

This explains why we see a decrease in traditional and normative ways of expressing authority and legitimacy within the movement as users’ careers progress. The early career is typically characterised by more ideological talk and advocacy of violence, used by members to emphasise the depth of their commitment to the movement and its ideology. In the late career, however, talk of violence is minimal, and ideological talk is accompanied by ‘off-topic’ social talk, indicating that late-career users feel they can still be seen as legitimate members of the group without ‘proving themselves’ with more violently ideological talk. Likewise, the drop in assertive claims and overt displays of knowledge indicates that late-career members do not need to use these resources to have their authority recognised. In Goffmanian terms, the user’s footing towards the community appears to change across their career as they develop more capital. While early-career members are keen to position themselves within the community and assert the legitimacy of their place there, late-career members’ performance indicates that they feel they can take their legitimacy for granted, with no need to overtly stake their claim to membership in this way.

If the late career is characterised by greater agency and therefore greater individuality in identity, does this present a contradiction with the idea of a collective identity developed over time? Perhaps – but this paradox is inherent to the concept of collective identity as described by Melucci (1996) and others. One key principle of collective identity is that individuals assimilate towards this identity through their involvement in the community – that is, individuals’ identity performances converge to reflect the collective identity. However,

another key principle according to Melucci's influential view is that collective identity is a moving target. Because collective identity a *process*, not a *product*, it is renegotiated incrementally over time. Once an individual assimilates to the collective identity they become part of these renegotiations, having the "privilege to push and extend boundaries of acceptable conversation", knowledge, and beliefs (Golder & Donath, 2004, p. 22), or in Bourdieusian terms, to set the standard of legitimate language and behaviour (Bourdieu, 1991), without suffering a loss of face. Naturally, the pushing of boundaries implies stepping away from collective norms, instead exhibiting individuality.

We therefore expect to see individuality in the early career, among members who have not yet assimilated to a collective identity. We *also* expect to see individuality in the late career, among members who are now implicated in the process of collective identity renegotiation and change. This reveals a fundamental difficulty with the concept of collective identity: that of finding empirical evidence for what Melucci himself describes as "a concept, an analytical tool and not a datum or an essence, a 'thing' with a 'real' existence" (Melucci, 1996, p. 77). To resolve this, we must turn to what Melucci calls *reifications*, those artefacts – practices, items, linguistic resources – which may have shared meaning for a group and therefore serve as "empirical indicators of a possible collective identity" (Melucci, 1996, p. 75). Melucci denies that these reifications are synonymous with the collective identity. In the case of individual identities, however, the social constructionist view holds that discursive performance *is* synonymous with identity, rather than simply reflecting some underlying identity (Antaki & Widdicombe, 1998; Benwell & Stokoe, 2006). Therefore, the changing identify performances seen across the five individuals can be seen not only as changes in performance, but as changes in identity. Where these individual identities change and converge in the same direction, this can be understood as assimilation to the same, collective, identity, and therefore as evidence for collective identity in this environment. While individuality is retained across the five users, we have seen that collective identity does not preclude individuality; in fact, individuality is inherent in the collective.

9.2.2 Applying the career concept

The career is a useful theoretical concept to understand identity change over time in online contexts. A flexible understanding and application of the concept is valuable, allowing for a wider range of insights on the career.

In the present thesis, I have demonstrated that the concept of the career is a valid model for the investigation of identity change over time. Specifically, I have demonstrated that this is the case in online communities, where the concept has not been applied previously. More generally, the existing literature which explicitly makes use of the career has been highly

theoretical in nature, often developing the concept in a single, offline context (Becker, 1973; Fillieule, 2001; Goffman, 1961). It is unclear from this how the concept might be operationalised by researchers working in other contexts, that is, how the career can practically be studied or measured. This is particularly the case for online contexts, as these older foundational texts typically take decidedly face-to-face, ethnographic research approaches, which have little in common with the more restricted data which can be accessed concerning members' engagement with online communities.

In Chapter 5, however, I systemised literature which does not work specifically within the career framework, but which nonetheless functions to describe careers in the sense of describing how community members can be categorised according to their role or stage of progression through the community. In doing so, I have been able to show how the career can be operationalised in a way which is theoretically applicable to any research context, including online (RQ1). These general principles are a focus on both *temporal* elements and *behavioural* elements of the career. Within this general framework, the career can and should be researched in a way that remains sensitive to the specific characteristics of the research context. The findings suggest that the most salient piece of context is the online/offline distinction, with certain methods better suited to each. However, because multiple correspondence analysis creates dimensions, with different typologies falling on different positions along these axes, the framework naturally offers flexibility, allowing for greater or lesser focus on the temporal or behavioural element depending on the research context and researcher's goals. The framework is therefore intended as a guide, not an instruction manual; context is central to designing a study of careers.

The flexibility afforded by this framework reflects the literature on careers, which often shows disagreement in key areas. In particular, researchers have vastly different understandings of the proper unit of analysis for the study of careers. Foundational texts on the career (Becker, 1973; Goffman, 1961) take the approach of making generalisations across all individuals in the sample, identifying a single general career path. While recognising that individuals may vary around this general path, "unique outcomes are neglected in favor of such changes over time as are basic and common to the members of a social category" (Goffman, 1961, p. 127). In this view, the proper unit of analysis is the entire 'social category' under analysis, who follow more or less the same career path.

However, this one-size-fits-all approach is not always observed in the empirical literature both within and beyond the explicit career framework. In Chapter 5, I outlined a wealth of literature from across disciplines which explores career types. Some of these follow Becker and Goffman in assuming that members typically follow the same set of stages along the

same career path (e.g. Dennen, 2008; Lave & Wenger, 1991; Singh, 2012; Smith Risser & Bottoms, 2014). However, other researchers take a less universalist approach, providing typologies which imply that multiple different career paths are possible in a given community context (e.g. Altier et al., 2022; Chiang et al., 2021; Gill & Horgan, 2013; Marcoccia, 2004; Waters & Gasson, 2005). Within the social movement context, Fillieule (2001) goes even further, advocating for the individual as the unit of analysis. He cites the myriad different ways in which different individuals engage with the collective, conditioned by both the nature of the collective but also by the individual's own lifestyle and previous experiences. Although not working directly within the career framework, Pilkington (2023a, 2023b) takes a similar approach, expressing scepticism about the value of career typologies which, she says, inevitably mask the reality of individualism and leave us with an incomplete understanding (Pilkington, 2023a). Pilkington (2023b) demonstrates her granular ethnographic approach to (de)radicalisation careers. Because individuals' career choices are "shaped [...] by a range of affective and situational and dispositional factors," Pilkington (2023b, p. 4) argues, no two individuals can follow exactly the same career path.

In some ways, this work supports the views of Fillieule and Pilkington. While the five users I have explored here do show some similarities to one another, they nonetheless resist grouping into 'types' in any meaningful way based on their linguistic behaviour. Of course, this may be related to the small sample size – perhaps a larger set would have shown clearer types of users. However, following the resource-constraint model of identity performance I have drawn on here (§2.4), while some of the resources individuals draw on in group contexts will be those derived from and therefore shared with the group, inevitably individuals' performances will also be coloured by resources derived from other contexts. No two individuals' identity performances, even in the late career, will be identical. Therefore, taking the individual as the unit of analysis for a study of career will always provide the richest and most accurate results.

Nevertheless, I believe there is value in analysis of the career at higher levels, such as the identification of broad career types in a given research context. As I have shown in Chapters 6 and 7, typologies can help us to make sense of larger datasets such as the one I have presented here (§4.3), which contains data for over 120,000 users. Providing the 'full story' of careers in such a large dataset, especially at the level of detail advocated by Pilkington (2023a, 2023b), would be unfeasible. While Pilkington's approach might provide the 'full story' for a small number of users, the remainder of the careers in the dataset would remain unexplored and unknown, sacrificing the 'full story' of the dataset as a whole. Indeed, as Pilkington (2023a) herself recognises, the practical demands of publication necessitate a loss of detail when considering only a very small sample of individual careers. In Chapter 6, I

have identified a set of four broad user types based on career duration – *mayflies*, *dragonflies*, *mantises*, and *cicadas*. This generalisation has been a useful tool for this work, helping to guide subsequent studies by serving as a ‘stepping stone’ or ‘way in’ to the identification of suitable individuals for further study. However, the career types are also valuable in their own right. As Chapter 7 shows, the four types show correlations – including some very strong correlations – with the use of certain linguistic features. This shows that, despite representing a significant abstraction away from the individual career, these types are able to show us real patterns across the wider dataset.

Far from coming down on any side of the debate, this research has drawn on all three notions of the proper unit of analysis for the career (RQ2). In Chapters 7 and 8, I have taken the individual as the unit of analysis, characterising each of the five users’ linguistic performance over time in both quantitative and qualitative terms. I have also grouped members into four broad types in Chapters 6 and 7. These findings, as I have argued, have value both as a stepping stone to further study, and in their own right. Finally, I have taken steps in Chapter 8 and in the present chapter towards a generalised approach, using the findings from the five users to suggest general patterns in identity change over users’ careers (albeit while recognising the limitations of extrapolating from such a small sample). This demonstrates that the study of the career has multiple legitimate units of analysis which may be appropriate to different contexts and research questions.

9.2.3 The interdisciplinary approach

Combining concepts and methods from sociology and linguistics is a promising way of characterising membership in online groups, allowing us to understand the nature of individuals’ attachment to the group and the identities that individuals assume as part of the group.

In this thesis, I have demonstrated that sociological and linguistic theory can be brought together fruitfully to examine the nature of participation in online communities in a more complete way than is afforded by either discipline by itself. As I discuss in this section, each of the disciplines has its own shortcomings in their application to my research aims, which can be addressed by the inclusion of the other (Bryman, 2006).

Primary among these benefits is the recognition of the role of language in identity construction. As I have outlined in §3.8.2, language is frequently cited in the sociological literature on collective identity as a key element in the (re)negotiation of identity. However, the role of language is never explored further than these brief mentions. On the contrary, within linguistics, the centrality of language for individual and group identity construction has long since been recognised, and has, accordingly, been investigated in a very wide range of

contexts, of which only a small selection are highlighted in Chapter 2. From this perspective, the absence of a thorough consideration of language in the sociological literature is extremely striking, and a linguistically-informed approach clearly has much to add in this unexplored area.

Likewise, the construction of identity *online* has been largely overlooked in the sociological study of social movements. Meanwhile, language has been widely identified as a particularly important resource for identity construction in online contexts due to the disembodied and often largely text-based nature of participation in such environments. This is reflected in a large linguistic literature on identity construction online (see §2.5.1). While a small number of studies exist on social movements and particularly collective identity in online environments, consideration of language is almost absent from this literature; only Gerbaudo and Treré (2015) highlight the role of features like lexicons and naming practices in the development of collective identities online. Bringing together the sociological and linguistic perspectives has allowed me to explore in detail, apparently for the first time, not only how language is used to construct collective identities, but also how this manifests in online social movements.

Despite the absence of consideration of language, the sociological literature nonetheless provides useful frameworks (namely, collective identity and the career) for the understanding of group identity, and identity change over time in a group context, in the specific context of ideological and social movement-related groups. While the term *collective identity* has been adopted in some linguistic research (e.g. Koller, 2009, 2012), it is used in a generic way synonymous with 'group identity' rather than in the specialised way I have outlined in this work (§3.6). The notion of the *career* is apparently entirely absent from the linguistic literature, although as shown in Chapter 5 some research employing linguistic methods has identified roles and stages in a community in a way that is very similar to work on careers (e.g. Chiang et al., 2021; Dennen, 2008; Marcoccia, 2004; Radin, 2006). Meanwhile, only scant attention has been paid to language in the careers literature. Only Becker (1973) remarks that "[t]he vocabularies in which deviant motivations are phrased reveal that their users acquire them in interaction with other deviants" (pp. 30-31); this idea has not been pursued further until now.

One framework which has been widely employed in linguistic studies that shares similarities with both collective identity and the career is the *communities of practice* (CoP) framework (Lave & Wenger, 1991; Wenger, 1998; Wenger et al., 2002), which describes the process by which individuals become progressively more involved in a group as they assimilate to the 'moving target' of that group's norms. While the CoP framework is arguably applicable to the present research context, those of collective identity and the career provide a better fit to the

specifics of the community. While CoPs are fundamentally organised around a practice or activity of some kind, collective identity and the career provide concepts which specifically describe (or have been adapted to describe) the nature of engagement in social movement contexts, where ideology rather than activity is the central binding factor. This allows for more context-appropriate analysis, such as my characterisation of the forum and its members' identity drawing on Taylor and Whittier's (1992) three key elements of collective identity, *boundaries*, *consciousness*, and *negotiation* (see §3.6 and §9.2.1).

As I have demonstrated in Chapter 5, the career is also a useful lens through which to bring together and analyse a highly disparate and multidisciplinary literature which describes individuals' progression through stages or different roles across different social contexts. By uniting these papers according to this important, but previously unexplored, similarity, I have been able to develop a unified theory which I have employed in this thesis to explore the forum posting career in a systematic way (RQ1). The findings of the multiple correspondence analysis recounted in Chapter 5 further justify the use of both sociological and linguistic approaches. The temporal and behavioural elements which I identified as central to the analysis of a career in any context lend themselves to a combination of methods from both disciplines, namely sequence analysis (Chapter 6) and corpus-assisted discourse analysis (Chapters 7 and 8).

In summary, while sociology has largely neglected useful methods which would allow the discipline to explore the importance of language in (collective) identity construction, linguistics has often explored, but undertheorized, research contexts of sociological interest. Linguistic studies may therefore be enriched by stronger engagement with these sociological concepts and theories. Despite shared roots in the Goffmanian notion of identity as performance (Goffman, 1956), the connection between the two disciplines has, in more recent years, been largely overlooked. In this thesis, I have reunited the two, demonstrating that despite shifts in the development in both fields, the two remain compatible and complementary.

9.3 Limitations and future directions

As I have noted throughout, this thesis was exploratory in nature, seeking to investigate the potential of combining sociological concepts such as collective identity and the career with linguistic methods. Naturally, then, scope of this thesis has been limited by a number of theoretical as well as practical considerations. This creates a number of opportunities for future work.

Chapter 7 identified wider quantitative linguistic trends, overall taking into account most users of the forum. However, such large-scale approaches naturally lose nuance (e.g. Egbert et al., 2020). While I did carry out more detailed quantitative and qualitative study of individual careers in the latter half of Chapter 7 and in Chapter 8, time constraints meant that I was only able to do so for five users from the cicada group. A valuable future direction may be to expand this more detailed study to a larger sample of users, including users following different career types, in order to confirm or amend the findings I have reported here.

The linguistic studies presented here are centred around a limited set of linguistic features. As I detailed in Chapter 7, I believe that the selection of features was justified by the exploratory study of the key n-gram lists for each subcorpus. Nevertheless, these features likely do not represent an exhaustive list of those which may distinguish between user types, or between early and late stages of the same career. As I highlighted in §3.4, the literature on newcomers and veterans shows that the former perform in powerless or non-authoritative ways compared with the latter. Contrary to expectations, I found that users appear to become less overtly assertive and authoritative as their careers go on. In any case, power and authority seem to be central elements of linguistic identity performance across the career. A further exploration of a wider range of lexico-grammatical features used to index power may therefore be worthwhile. The work in development by Busso et al. (2023), following Newsome and Grant (in press), may be a useful starting point for this.

One interesting finding from this thesis concerns the apparent similarities and differences between the linguistic behaviour of mayfly users on the one hand, and early-career cicada users on the other. As Chapter 7 shows, certain linguistic features are more strongly associated with long-term cicada careers, but appear to be used more frequently in the early stages of these careers. This suggests that there may be some features which could be used to differentiate between the posts of users who will have only short careers, and those who will go on to have longer careers. This could have applications in the early detection of more dedicated forum users. This hypothesis would require further testing using a much larger set of users, which was beyond the scope of this thesis.

More generally, in this thesis I have developed ideas which may theoretically be applied to other communities, particularly online communities, to investigate the nature of participation and identity in these spaces. These include the combination of sociological and linguistic principles to study the nature of identity change over time in a theoretically sound way, and, more practically, a method for operationalising the abstract concept of the career (Chapter 5). In particular, it would be interesting to compare this white nationalist forum with other

extremist communities, to investigate whether patterns hold across different communities or whether different communities behave idiosyncratically.

9.4 Summary of contributions

The existing sociological literature on careers is largely abstract, taking the structure of the career for granted, and providing little guidance for other would-be researchers of the career. My primary contribution to the literature has been to develop a framework for the analysis of the career across contexts, applying this to the study of identity over time in a white nationalist forum. The specific contributions of each empirical chapter are summarised below, followed by a reflection on the potential practical or policy-related implications of this thesis.

9.4.1 Chapter 5 – How to study the career: A multiple correspondence analysis

- It is possible to systematise the literature which describes career roles and stages and to identify key elements for the study of careers, bringing more order and direction to this literature and providing guidance for the operationalisation of the career concept.
- The two major elements for the study of careers are the *temporal* element – the amount of time the individual has dedicated to the community, in terms of frequency but especially duration of engagement – and the *behavioural* element – the nature of the engagement the individual has with the community.
- Within this framework, there is room for a flexible approach depending on the research context, with the online vs. offline context being particularly relevant.

9.4.2 Chapter 6 – The temporal career: A sequence analysis

- Neither the structural approach (Becker, 1973; Goffman, 1961) nor the individualised approach (Fillieule, 2010) to understanding careers is fully supported. Instead, a number of typical career types were identified (*mayflies*, one year duration, 76% of users; *dragonflies*, two years' duration, 10% of users; *mantises*, three to six years' duration, 10% of users; and *cicadas*, seven or more years' duration, 4% of users).
- Most people follow a small number of career paths; beyond these paths, however, there is significant heterogeneity.
- The most common careers I have identified, which are short and characterised by low rates of engagement, are very different to what the literature on careers (§3.3.2) would lead us to believe, which would be a longer career with multiple steps.

9.4.3 Chapter 7 – The behavioural career I: A quantitative linguistic analysis

- There are linguistic features which correlate with career length (i.e. that are more common in shorter careers and less common in longer careers, and vice versa). These are primarily pronominal features but also some stance features (complement clauses headed by nouns and verbs, and modal verbs), and aspectual verbs of progression.
- Investigating usage of these features across the careers of individual users shows that overall, most of the features are more prevalent in the early career. Therefore, no features were identified which are associated specifically with the later stages of long career. This suggests that users draw from a wider repertoire of possible resources in their later career.

9.4.4 Chapter 8 – The behavioural career II: A qualitative linguistic analysis

- Users who have long careers do not *necessarily* become more extreme in their beliefs across the career. Other researchers (e.g. Abay Gaspar et al., 2020; McCauley & Moskalenko, 2017; Pilkington, 2023b) have also shown that radicalisation to the most extreme possible end point is not inevitable. However, I have demonstrated this in a new way, showing how this manifests discursively across the career.
- Despite the wider range of resources available to them in the later career, users' identity performances become more similar to one another over time in several important ways, suggesting that they develop a collective identity through their career. These key similarities include: decreased usage of overt expressions of authority and violent ideation; increased engagement in off-topic and other rapport-building talk; and gatekeeping of their white and white nationalist identities, through increased criticism of members of these in-groups.
- Users gain symbolic capital over time, meaning that they accrue new resources across their career. In the late career, users therefore need not draw on typical or overt resources to index their authority in the forum, and are able to do so without threatening their own, or others', face. Early career users are more constrained in their identity performances.

9.4.5 Practical implications

This thesis does not seek to make direct recommendations for law enforcement, policymakers, or other stakeholders. Nevertheless, this work is of potential value to such parties. As I highlighted in §1.2, social media environments such as the forum context I have explored here are a high priority for security services, as such platforms are increasingly

implicated in the development of extremist groups and individuals who may pose a violent, or even deadly, threat (Europol, 2023; Mutton et al., 2023). It is therefore necessary to collect as much information as possible about such groups and platforms in order to better tackle this threat (Bentley, 2015), a goal to which this thesis contributes.

From a practical perspective, one major contribution of this thesis is how it may aid in the prioritisation of particular targets for further investigation and intervention by law enforcement. The four user groups I have identified are distinguished on the basis of their quantitatively-measured commitment to the forum over time. Therefore, the cicada group, who have the longest forum careers, can be understood as the most dedicated group of users, at least by this measure. As the cicada group represent only a small proportion of the total user base, these users may be conveniently seen as an initial priority for further investigation.

Further to this, the thesis sheds light on the nature of the connection that these most dedicated members develop with the community through their long-term membership. As I have shown, this connection may not be simply ideological in nature, as the cicada users investigated here also engage more socially with the community in their late career. This has implications for parties with an interest in disrupting the connection between individuals and the white nationalist milieu, as turning individuals away from white nationalist ideologies is likely to be insufficient to fully sever this connection.

In addition, the quantitative analysis of the use of particular linguistic features between and across career types has shown that some features could be associated with the early stages of long careers (§7.4.3). Although significant further investigation into this possibility is required, this finding may have an application in the early identification of users who will go on to show greater commitment to the forum over time, again aiding in the prioritisation of targets for further investigation and intervention.

Finally, while this thesis has focused on the career paths of (collections of) individuals, the work reported here may have applications at the site level. Through sequence analysis, identifying the most common career lengths, I have demonstrated the level of 'stickiness' of this site, in terms of the proportion of users making repeated use of the site over different lengths of time. While I have only considered a single site, the same method could be applied to other fora in order to estimate their relative impact within the wider online white nationalist or far-right movement. This may therefore allow investigators to prioritise particular sites as targets for investigation or infiltration, prior to identifying key users within those sites.

9.5 Concluding remarks

This thesis has introduced a novel interdisciplinary approach to the study of identity change over time in an online forum context. It has fruitfully combined the sociological notions of career and collective identity with the social constructionist view of identity as interactionally (re)negotiated through language, reconciling two fields which share a common theoretical grounding in the works of Goffman. To this end, I have demonstrated the value of a multimethodological approach to analysis of the career, comprising sequence analysis to capture temporal elements of the career and both quantitative and qualitative corpus-assisted discourse analysis to investigate behavioural elements. I am hopeful that this work will contribute to the body of knowledge on online extremist communities, with an ultimate view to disrupting the collective identity processes taking place there, and that the approach I have taken here may be of use to other scholars of the career.

References

- Abay Gaspar, H., Daase, C., Deitelhoff, N., Junk, J., & Sold, M. (2020). Radicalization and political violence: Challenges of conceptualizing and researching origins, processes and politics of illiberal beliefs. *International Journal of Conflict and Violence*, 14(2), 1-18.
<https://doi.org/10.4119/ijcv-3802>
- Adams, J., & Roscigno, V. J. (2005). White supremacists, oppositional culture and the World Wide Web. *Social Forces*, 84(2), 759-778.
- Adkins, L. (2008). Social capital put to the test. *Sociology Compass*, 2(4), 1209-1227.
<https://doi.org/10.1111/j.1751-9020.2008.00123.x>
- ADL. (2017). Christian Identity. Retrieved 24th May 2023 from
<https://www.adl.org/resources/backgrounders/christian-identity>
- Ahmad, A. (2016). The ties that bind and blind: Embeddedness and radicalisation of youth in one Islamist organisation in Pakistan. *The Journal of Development Studies*, 52(1), 5-21.
<https://doi.org/10.1080/00220388.2015.1075976>
- Aiston, J. (2022). *Argumentation strategies in an online male separatist community* [PhD thesis, Lancaster University].
- Altier, M. B., Leonard Boyle, E., & Horgan, J. G. (2022). Terrorist transformations: The link between terrorist roles and terrorist disengagement. *Studies in Conflict & Terrorism*, 45(9), 753-777. <https://doi.org/10.1080/1057610x.2019.1700038>
- Anahita, S. (2006). Blogging the borders: Virtual skinheads, hypermasculinity, and heteronormativity. *Journal of Political and Military Sociology*, 32, 143-164.
- Andrew, P. (2016). Constructing age identity: The case of Mexican EFL learners. In S. Preece (Ed.), *The Routledge handbook of language and identity* (pp. 337-350). Routledge.
- Andrews, T. (2012). What is social constructionism? *The Grounded Theory Review*, 11(1), 39-46.
- Ang, C. S., & Zaphiris, P. (2010). Social roles of players in MMORPG guilds. *Information, Communication & Society*, 13(4), 592-614. <https://doi.org/10.1080/13691180903266952>
- Angouri, J., & Sanderson, T. (2016). 'You'll find lots of help here': Unpacking the function of an online rheumatoid arthritis (RA) forum. *Language & Communication*, 46, 1-13.
<https://doi.org/10.1016/j.langcom.2015.10.001>

Antaki, C., & Widdicombe, S. (1998). Identity as an achievement and as a tool. In C. Antaki & S. Widdicombe (Eds.), *Identities in talk* (pp. 1-14). SAGE.

Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C. P., & Wang, X. (2006). *Talk to me: Foundations for successful individual-group interactions in online communities*. Conference on Human Factors in Computing Systems, Montréal, Canada.

BAAL. (2021). *Recommendations on good practice in applied linguistics*.
<https://www.baal.org.uk/wp-content/uploads/2021/03/BAAL-Good-Practice-Guidelines-2021.pdf>

Bakardjieva, M. (2015). Do clouds have politics? Collective actors in social media land. *Information, Communication & Society*, 18(8), 983-990.
<https://doi.org/10.1080/1369118X.2015.1043320>

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Baker, P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies*, 9(3), 247-256.
<https://doi.org/10.1080/17405904.2012.688297>

Baker, P., Vessey, R., & McEnery, T. (2021). *The language of violent jihad*. Cambridge University Press.

Bakhtin, M. M. (1981). *The dialogic imagination* (M. Holquist & C. Emerson, Trans.). University of Texas Press.

Baron, N. S. (2008). *Always on: Language in an online and mobile world*. Oxford University Press.

Barton, D., & Lee, C. (2013). *Language online: Investigating digital texts and practices*. Routledge.

Bassett, E. H., & O'Riordan, K. (2002). Ethics of internet research: Contesting the human subjects research model. *Ethics and Information Technology*, 4, 233-247.

Baumgarten, N. (2017). Othering practice in a right-wing extremist online forum. *Language@Internet*, 14(1).

Becker, H. S. (1973). *Outsiders: Studies in the sociology of deviance* (2nd ed.). Free Press.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2, 8-14. <https://doi.org/10.1016/j.npls.2016.01.001>

- Benne, K. D., & Sheats, P. (1948). Functional roles of group members. *Journal of Social Issues*, 4(2), 41-49. <https://doi.org/10.1111/j.1540-4560.1948.tb01783.x>
- Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739-768. <https://doi.org/10.1080/1369118X.2012.670661>
- Bentley, M. (2015). Recognition making response: Preventing far-right extremism and radicalisation. In C. Baker-Beall, C. Heath-Kelly, & L. Jarvis (Eds.), *Counter-radicalisation: Critical perspectives* (pp. 106-123). Routledge.
- Benwell, B., & Stokoe, E. (2006). *Discourse and identity*. Edinburgh University Press.
- Berger, J. M. (2018). *Extremism*. MIT Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.
- Binnick, R. I. (1991). *Time and the verb: A guide to tense and aspect*. Oxford University Press.
- Blee, K. M. (2002). *Inside organized racism: Women in the hate movement*. University of California Press.
- Block, D. (2016). Class in language and identity research. In S. Preece (Ed.), *The Routledge handbook of language and identity* (pp. 241-254). Routledge.
- Blommaert, J. (2018). *Durkheim and the internet: Sociolinguistics and the sociological imagination*. Bloomsbury.
- Blommaert, J., & Backus, A. (2013). Superdiverse repertoires and the individual. In I. de Saint-Georges & J.-J. Weber (Eds.), *Multilingualism and multimodality: Current challenges for educational studies* (pp. 11-32). Sense Publishers.
- Blommaert, J., & Varis, P. (2013). Enough is enough: The heuristics of authenticity in superdiversity. In J. Duarte & I. Gogolin (Eds.), *Linguistic superdiversity in urban areas: Research approaches* (pp. 143-159). John Benjamins.
- Boostrom, R. E. (2008). The social construction of virtual reality and the stigmatized identity of the newbie. *Journal For Virtual Worlds Research*, 1(2). <https://doi.org/10.4101/jvwr.v1i2.302>

A. E. Booth, PhD Thesis, Aston University 2023

Booth, A. (2023). Fractured in-group identity negotiation in an online white nationalist forum. *Applied Corpus Linguistics*, 3(3), 1-8. <https://doi.org/10.1016/j.acorp.2023.100062>

Borum, R. (2003). Understanding the terrorist mind-set. *FBI Law Enforcement Bulletin*, 72(7), 7-10.

Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. *Journal of Strategic Security*, 4(4), 7-36. <https://doi.org/10.5038/1944-0472.4.4.1>

Bostdorff, D. M. (2004). The internet rhetoric of the Ku Klux Klan: A case study in web site community building run amok. *Communication Studies*, 55(2), 340-361.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). Routledge.

Bourdieu, P. (1991). The production and reproduction of legitimate language (G. Raymond & M. Adamson, Trans.). In J. B. Thompson (Ed.), *Language and symbolic power* (pp. 43-65). Polity. (Reprinted from *Ce que parler veut dire: l'économie des échanges linguistiques*, pp. 23-58, by P. Bourdieu, 1982, Librairie Anthème Fayard.)

Bourdieu, P. (2004). The forms of capital. In S. J. Ball (Ed.), *The RoutledgeFalmer reader in sociology of education* (pp. 15-29). RoutledgeFalmer. (Reprinted from *Handbook of theory and research for the sociology of education*, pp. 241-258, by J. C. Richardson, Ed., 1986, Greenwood Publishing Group)

Bowden, Z. A. (2008). Poriadok and bardak (Order and Chaos): The neo-fascist project of articulating a Russian "People". *Journal of Language and Politics*, 7(2), 321-347. <https://doi.org/10.1075/jlp.7.2.07bow>

boyd, d. m. (2008). *Taken out of context: American teen sociality in networked publics* [PhD thesis, University of California, Berkeley].

boyd, d. m. (2011). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. A. Papacharissi (Ed.), *A networked self: Identity, community, and culture on social network sites* (pp. 39-58). Routledge.

Brindle, A. (2016). *The language of hate: A corpus linguistic analysis of white supremacist language*. Routledge.

Brookes, G., & McEnery, T. (2020). Correlation, collocation and cohesion: A corpus-based critical analysis of violent jihadist discourse. *Discourse & Society*, 31(4), 351-373. <https://doi.org/10.1177/0957926520903528>

A. E. Booth, PhD Thesis, Aston University 2023

Brown, P., & Levinson, S. C. (2006). Politeness: Some universals in language use. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed., pp. 311-323). Routledge. (Reprinted from *Politeness: Some universals in language usage*, by P. Brown and S. C. Levinson, 1987, Cambridge University Press.)

Brush, A. J. B., Wang, X., Turner, T. C., & Smith, M. A. (2005). *Assessing differential usage of Usenet social accounting meta-data*. SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, USA.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113. <https://doi.org/10.1177/1468794106058877>

Bucholtz, M. (2007). Shop talk: Branding, consumption, and gender in American middle-class youth interaction. In B. S. McElhinny (Ed.), *Words, worlds, and material girls: Language, gender, globalization* (pp. 371-402). Mouton de Gruyter.

Bucholtz, M., & Hall, K. (2004). Language and identity. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 369-394). Blackwell.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585-614. <https://doi.org/10.1177/1461445605054407>

Bühlmann, F., Benz, P., Mach, A., & Rossier, T. (2017). Mapping the power of law professors: The role of scientific and social capital. *Minerva*, 55(4), 509-531. <https://doi.org/10.1007/s11024-017-9333-1>

Bunnage, L. A. (2014). Social movement engagement over the long haul: Understanding activist retention. *Sociology Compass*, 8(4), 433-445. <https://doi.org/10.1111/soc4.12141>

Burr, V., & Dick, P. (2017). Social constructionism. In B. Gough (Ed.), *The Palgrave handbook of critical social psychology* (pp. 59-80). Palgrave Macmillan.

Busch, B. (2012). The linguistic repertoire revisited. *Applied Linguistics*, 33(5), 503-523. <https://doi.org/10.1093/applin/ams056>

Busso, L., Htait, A., & Grant, T. (2023, July 5th). *Hierarchies of power in digital communities: A corpus linguistics study*. 16th Biennial Conference of the International Association for Forensic and Legal Linguistics, Manila, Philippines.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Cameron, D. (1999). Performing gender identity: Young men's talk and the construction of heterosexual masculinity. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (pp. 442-458). Routledge.

A. E. Booth, PhD Thesis, Aston University 2023

Chiang, E. (2018). *Rhetorical moves and identity performance in online child sexual abuse interactions* [PhD thesis, Aston University].

Chiang, E., Nguyen, D., Towler, A., Haas, M., & Grieve, J. (2021). Linguistic analysis of suspected child sexual offenders' interactions in a dark web image exchange chatroom. *International Journal of Speech, Language, and the Law*, 27(2), 129-161.
<https://doi.org/10.1558/ijssl.41446>

Chidgey, R. (2021). Postfeminism™: celebrity feminism, branding and the performance of activist capital. *Feminist Media Studies*, 21(7), 1055-1071.
<https://doi.org/10.1080/14680777.2020.1804431>

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). *Bryman's social research methods* (6th ed.). Oxford University Press.

Clarke, I. (2019). Functional linguistic variation in Twitter trolling. *International Journal of Speech, Language, and the Law*, 26(1), 57-84. <https://doi.org/10.1558/ijssl.34803>

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95-S120.

Collins, R. (2004). *Interaction ritual chains*. Princeton University Press.

Conway, M. (2021). Online extremism and terrorism research ethics: Researcher safety, informed consent, and the need for tailored guidelines. *Terrorism and Political Violence*, 33(2), 367-380. <https://doi.org/10.1080/09546553.2021.1880235>

Corrigan-Brown, C. (2012). *Patterns of protest: Trajectories of participation in social movements*. Stanford University Press.

Cortese, D. K. (2015). I'm a "good" activist, you're a "bad" activist, and everything I do is activism: parsing the different types of "activist" identities in LGBTQ organizing. *Interface*, 7(1), 215-246.

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25(4), 431-447. <https://doi.org/10.1093/applin/25.4.431>

Crossley, N. (2014). Social class. In M. J. Grenfell (Ed.), *Pierre Bourdieu: Key concepts* (pp. 85-97). Routledge.

Dalgaard-Nielsen, A. (2010). Violent radicalization in Europe: What we know and what we do not know. *Studies in Conflict & Terrorism*, 33(9), 797-814.

<https://doi.org/10.1080/1057610X.2010.501423>

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). *No country for old members: User lifecycle and linguistic change in online communities*. 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil.

Darics, E., & Gatti, M. C. (2019). Talking a team into being in online workplace collaborations: The discourse of virtual work. *Discourse Studies*, 21(3), 237-257.

<https://doi.org/10.1177/1461445619829240>

De Fina, A., Schriffin, D., & Bamberg, M. (2006). *Discourse and identity*. Cambridge University Press.

De Koster, W., & Houtman, D. (2008). 'Stormfront is like a second home to me': On virtual community formation by right-wing extremists. *Information, Communication & Society*, 11(8), 1155-1176. <https://doi.org/10.1080/13691180802266665>

de Saussure, F. (1995). *Cours de linguistique générale [Course in general linguistics]*. Payot. (Original work published 1916)

Dennen, V. P. (2008). *Intersecting communities of practice: Merging roles across the academic and blogging worlds*. IADIS International Conference, Algarve, Portugal.

Doherty, B., & Hayes, G. (2019). Tactics and strategic action. In D. A. Snow, S. A. Soule, H. Kriesi, & H. J. McCammon (Eds.), *The Wiley Blackwell companion to social movements* (2nd ed., pp. 271-288). John Wiley & Sons Ltd.

Doran, M. (2004). Negotiating between *bourge* and *racaille*: Verlan as youth identity practice in suburban Paris. In A. Pavlenko & A. Blackledge (Eds.), *Negotiation of identities in multilingual contexts* (pp. 93-124). Multilingual Matters Ltd.

Douglas, K. M., McGarty, C., Bliuc, A.-M., & Lala, G. (2005). Understanding cyberhate: Social competition and social creativity in online white supremacist groups. *Social Science Computer Review*, 23(1), 68-76. <https://doi.org/10.1177/0894439304271538>

Dowty, D. R. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Kluwer Academic Publishers.

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139-182). John Benjamins Publishing Company.

Duffy, M. E. (2003). Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry*, 27(3), 291-312.

<https://doi.org/10.1177/0196859903252850>

Eatwell, R. (2004). Introduction: the new extreme right challenge. In R. Eatwell & C. Mudde (Eds.), *Western democracies and the new extreme right challenge*. Routledge.

Eckert, P. (1996). Vowels and nail polish: The emergence of linguistic style in the preadolescent heterosexual marketplace. In N. Warner, J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, & M. Chen (Eds.), *Gender and belief systems: Proceedings of the fourth Berkeley Women and Language Conference* (pp. 183-190). Berkeley Women and Language Group.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender and community-based practice. *Annual Review of Anthropology*, 21(1), 461-490.

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utraiainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*, 4(1), 1-10.

<https://doi.org/10.1177/2158244014522633>

Europol. (2020). *European Union terrorism situation and trend report 2020*.

<https://www.europol.europa.eu/activities-services/main-reports/european-union-terrorism-situation-and-trend-report-te-sat-2020>

Europol. (2023). *European Union terrorism situation and trend report 2023*.

<https://www.europol.europa.eu/publication-events/main-reports/european-union-terrorism-situation-and-trend-report-2023-te-sat>

Evans, J. D. (1996). *Straightforward statistics for the behavioural sciences*. Thomson Brooks/Cole Publishing Co.

Fairclough, N. L. (2013). *Language and power* (2nd ed.). Longman.

Fillieule, O. (2001). Propositions pour une analyse processuelle de l'engagement individuel [Proposals for process analysis of individual commitment]. *Revue française de science politique*, 51(1-2), 199-215. <https://doi.org/10.3917/rfsp.511.0199>

Fillieule, O. (2010). Some elements of an interactionist approach to political disengagement. *Social Movement Studies*, 9(1), 1-15. <https://doi.org/10.1080/14742830903442436>

Fillieule, O. (2020). Carrière militante [Activist career]. In O. Fillieule, L. Mathieu, & C. Péchu (Eds.), *Dictionnaire des mouvements sociaux* (2nd ed., pp. 91-98). Presses de Sciences Po.

A. E. Booth, PhD Thesis, Aston University 2023

Fillieule, O., & Neveu, E. (2019). Activists' trajectories in space and time. In O. Fillieule & E. Neveu (Eds.), *Activists forever? Long-term impacts of political activism* (pp. 1-36). Cambridge University Press.

Finlayson, A. (2021). Neoliberalism, the alt-right and the intellectual dark web. *Theory, Culture & Society*, 38(6), 167-190. <https://doi.org/10.1177/02632764211036731>

Firer-Blaess, S. (2016). The collective identity of Anonymous: Web of meanings in a digitally enabled movement [PhD thesis, Uppsala Universitet].

Flesher Fominaya, C. (2010). Collective identity in social movements: Central concepts and debates. *Sociology Compass*, 4(6), 393-404. <https://doi.org/10.1111/j.1751-9020.2010.00287.x>

Flesher Fominaya, C. (2019). Collective identity in social movements: Assessing the limits of a theoretical framework. In D. A. Snow, S. A. Soule, H. Kriesi, & H. J. McCammon (Eds.), *The Wiley Blackwell companion to social movements* (2nd ed., pp. 429-445). John Wiley & Sons Ltd.

Flesher Fominaya, C., & Gillan, K. (2017). Navigating the technology-media-movements complex. *Social Movement Studies*, 16(4), 383-402. <https://doi.org/10.1080/14742837.2017.1338943>

Ford, J. (2021). *Gender disguise and linguistic identity performance in online writings: production, perception, and forensic applications* [PhD thesis, Aston University].

Futrell, R., & Simi, P. (2004). Free spaces, collective identity, and the persistence of U.S. white power activism. *Social Problems*, 51(1), 16-42.

Gabardinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1-37. <https://doi.org/10.18637/jss.v040.i04>

Gabardinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2010). *Mining sequence data in R with the TraMineR package: A user's guide*. University of Geneva. <http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf>

Galegher, J., Sproull, L., & Kiesler, S. (1998). Legitimacy, authority, and community in electronic support groups. *Written Communication*, 15(4), 493-530. <https://doi.org/10.1177/0741088398015004003>

Gautier Morin, J., & Rossier, T. (2021). The interaction of elite networks in the Pinochet regime's macroeconomic policies. *Global Networks*, 21(2), 339-364.

<https://doi.org/10.1111/glob.12300>

Gerbaudo, P. (2015). Protest avatars as memetic signifiers: Political profile pictures and the construction of collective identity on social media in the 2011 protest wave. *Information, Communication & Society*, 18(8), 916-929. <https://doi.org/10.1080/1369118X.2015.1043316>

Gerbaudo, P., & Treré, E. (2015). In search of the 'we' of social media activism: Introduction to the special issue on social media and protest identities. *Information, Communication & Society*, 18(8), 865-871. <https://doi.org/10.1080/1369118x.2015.1043319>

Giles, D. (2006). Constructing identities in cyberspace: The case of eating disorders. *British Journal of Social Psychology*, 45(3), 463-477. <https://doi.org/10.1348/014466605x53596>

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist use of the internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy*, 16(1), 99-117. <https://doi.org/10.1111/1745-9133.12249>

Gill, P., & Horgan, J. (2013). Who were the volunteers? The shifting sociological and operational profile of 1240 Provisional Irish Republican Army members. *Terrorism and Political Violence*, 25(3), 435-456. <https://doi.org/10.1080/09546553.2012.664587>

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15-31. <https://doi.org/10.1177/1094428112452151>

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. (2009). *A conceptual and operational definition of 'social role' in online community*. 42nd Hawaii International Conference on System Sciences, Waikoloa, HI, USA.

Goffman, E. (1956). *The presentation of self in everyday life*. University of Edinburgh Social Sciences Research Centre.

Goffman, E. (1961). *Asylums: Essays on the social situation of mental patients and other inmates*. Doubleday Anchor Books.

Goffman, E. (1963). *Behavior in public places: Notes on the social organization of gatherings*. The Free Press.

Goffman, E. (1964). The neglected situation. *American Anthropologist*, 66(6), 133-136.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Pantheon.

A. E. Booth, PhD Thesis, Aston University 2023

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Golder, S. A., & Donath, J. (2004). *Social roles in electronic communities*. Internet Research 5.0, Brighton, UK.

Graham, S. L. (2016). Relationality, friendship, and identity in digital communication. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 305-320). Routledge.

Grant, T. (2013). Txt 4N6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy*, 21(2), 467-494.

Grant, T. (2022). *The idea of progress in forensic authorship analysis*. Cambridge University Press.

Grant, T., & MacLeod, N. (2020). *Language and online identities: The undercover policing of internet sexual crime*. Cambridge University Press.

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal*, 331(7524), 1064-1065. <https://doi.org/10.1136/bmj.38636.593461.68>

Gumperz, J. J. (1964). Linguistic and social interaction in two communities. *American Anthropologist*, 66(6), 137-153.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Hodder Arnold.

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioural Scientist*, 57(5), 664-688. <https://doi.org/10.1177/0002764212469365>

Hammersley, M. (1992). *What's wrong with ethnography?* Routledge.

Hasbullah, S. S., Maynard, D., Wan Chik, R. Z., Mohd, F., & Noor, M. (2016). *Automated content analysis: A sentiment analysis on Malaysian government social media*. 10th International Conference on Ubiquitous Information Management and Communication, Da Nang, Vietnam.

Hawley, G. (2017). *Making sense of the alt-right*. Columbia University Press.

- Heritage, F., & Koller, V. (2020). Incels, in-groups, and ideologies: The representation of gendered social actors in a sexuality-based online community. *Journal of Language and Sexuality*, 9(2), 152-178. <https://doi.org/10.1075/jls.19014.her>
- Herring, S. C. (1996). Gender and democracy in computer-mediated communication. In R. Kling (Ed.), *Computerization and controversy: Value conflicts and social choices* (2nd ed., pp. 476-489). Academic Press.
- Herring, S. C. (2001). Computer-mediated discourse. In D. Schrifin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 612-634). Blackwell.
- Herring, S. C. (2003). Gender and power in online communication. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 202-228). Blackwell.
- Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, 4.
- Herring, S. C., & Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. In D. Tannen, H. E. Hamilton, & D. Schrifin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 127-151). John Wiley & Sons Ltd.
- Herring, S. C., & Stoerger, S. (2014). Gender and (a)nonymity in computer-mediated communication. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The handbook of language, gender, and sexuality* (2nd ed., pp. 567-586). John Wiley & Sons Ltd.
- Hirschman, A. O. (1970). *Exit, voice and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.
- Hogg, M. A. (2012). Self-uncertainty, social identity, and the solace of extremism. In M. A. Hogg & D. L. Blaycock (Eds.), *Extremism and the psychology of uncertainty* (pp. 19-35). Wiley-Blackwell.
- Hogg, M. A., & Adelman, J. (2013). Uncertainty-identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues*, 69(3), 436-454. <https://doi.org/10.1111/josi.12023>
- Holt, T. J., Freilich, J. D., & Chermak, S. D. (2022). Examining the online expression of ideology among far-right extremist forum users. *Terrorism and Political Violence*, 34(2), 364-384. <https://doi.org/10.1080/09546553.2019.1701446>
- Honeycutt, C. (2005). Hazing as a process of boundary maintenance in an online community. *Journal of Computer-Mediated Communication*, 10(2).

A. E. Booth, PhD Thesis, Aston University 2023

Horgan, J. (2009). *Walking away from terrorism: Accounts of disengagement from radical and extremist movements*. Routledge.

Horgan, J. (2012). Discussion point: The end of radicalization? *National Consortium for the Study of Terrorism and Responses to Terrorism*.

<https://www.start.umd.edu/news/discussion-point-end-radicalization>

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

Hughes, E. C. (1937). Institutional office and the person. *American Journal of Sociology*, 34(3), 404-413.

Hund, S. A., & Benford, R. A. (2004). Collective identity, solidarity, and commitment. In D. A. Snow, S. A. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 433-457). Blackwell.

Hunter, M. (2022). *Violent ideologies: An investigation of the relationship between linguistic evaluative patterns and psychopathology in three types of violent offender* [PhD thesis, Aston University].

Husu, H.-M. (2013). Bourdieu and social movements: Considering identity movements in terms of field, capital and habitus. *Social Movement Studies*, 12(3), 264-279.

<https://doi.org/10.1080/14742837.2012.704174>

Institute for Economics and Peace. (2023). *Global terrorism index 2023: Measuring the impact of terrorism*. <https://www.visionofhumanity.org/wp-content/uploads/2023/03/GTI-2023-web-170423.pdf>

Introvigne, M. (1999). Defectors, ordinary leave-takers, and apostates: A quantitative study of former members of New Acropolis in France. *Nova Religio*, 3(1), 83-99.

<https://doi.org/10.1525/nr.1999.3.1.83>

Irwin, A. (2011). Social constructionism. In R. Wodak, B. Johnstone, & P. Kerswill (Eds.), *The SAGE handbook of sociolinguistics* (pp. 100-112). SAGE.

Jane, E. A. (2014). 'Back to the kitchen, cunt': Speaking the unspeakable about online misogyny. *Continuum*, 28(4), 558-570. <https://doi.org/10.1080/10304312.2014.924479>

Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, 17(1), 65-87. <https://doi.org/10.1007/s10676-015-9362-0>

Jasper, J. M. (1997). *The art of moral protest: Culture, biography, and creativity in social movements*. University of Chicago Press.

Jasper, J. M. (1998). The emotions of protest: Affective and reactive emotions in and around social movements. *Sociological Forum*, 13(3), 397-424.

Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review*, 27(2), 265-276.

<https://doi.org/10.1016/j.hrmr.2016.08.003>

Jensen, C., King, S., & Kuechler, V. (2011). *Joining free/open source software communities: An analysis of newbies' first interactions on project mailing lists*. 44th Hawaii International Conference on System Sciences, Kauai, HI, USA.

Johnston, H., Laraña, E., & Gusfield, J. R. (1994). Identities, grievances and new social movements. In E. Laraña, H. Johnston, & J. R. Gusfield (Eds.), *New social movements: From ideology to identity*. Temple University Press.

Johnstone, B. (1996). *The linguistic individual: Self-expression in language and linguistics*. Oxford University Press.

Johnstone, B. (2009). Stance, style, and the linguistic individual. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives* (pp. 29-52). Oxford University Press.

Josey, C. S. (2010). Hate speech and identity: An analysis of neo racism and the indexing of identity. *Discourse & Society*, 21(1), 27-39. <https://doi.org/10.1177/0957926509345071>

Julien, C. (2015). Bourdieu, social capital and online interaction. *Sociology*, 29(2), 356-373. <https://doi.org/10.1177/0038038514535862>

Junk, J., Süß, C.-A., Daase, C., & Deitelhoff, N. (2020). What do we know about radicalisation? Overview of the structure and key findings of the focus section. *International Journal of Conflict and Violence*, 14(2), 1-5. <https://doi.org/10.4119/ijcv-3875>

Juris, J. S. (2012). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, 39(2), 259-279. <https://doi.org/10.1111/j.1548-1425.2012.01362.x>

Kavada, A. (2015). Creating the collective: social media, the Occupy Movement and its constitution as a collective actor. *Information, Communication & Society*, 18(8), 872-886. <https://doi.org/10.1080/1369118x.2015.1043318>

Kenyon, J., Binder, J., & Baker-Beall, C. (2021). *Exploring the role of the internet in radicalisation and offending of convicted extremists*. Ministry of Justice Analytical Series.

<https://www.gov.uk/government/publications/exploring-the-role-of-the-internet-in-radicalisation-and-offending-of-convicted-extremists>

Kenyon, J., Binder, J., & Baker-Beall, C. (2022). *The internet and radicalisation pathways: Technological advances, relevance of mental health and role of attackers*. Ministry of Justice Analytical Series. <https://www.gov.uk/government/publications/internet-and-radicalisation-pathways-technological-advances-relevance-of-mental-health-and-role-of-attackers>

Kilgarriff, A. (2009). *Simple maths for keywords*. Corpus Linguistic Conference, Liverpool, UK.

Kim, A. J. (2000). *Community building on the web: Secret strategies for successful online communities*. Peachpit Press.

Klandermans, B. (1994). Transient identities? Membership patterns in the Dutch peace movement. In E. Laraña, H. Johnston, & J. R. Gusfield (Eds.), *New social movements: From ideology to identity* (pp. 168-184). Temple University Press.

Klandermans, B. (1997). *The social psychology of protest*. Blackwell.

Klandermans, B. (2004). The demand and supply of participation: Social-psychological correlates of participation in social movements. In D. A. Snow, S. A. Soule, & H. Kriesi (Eds.), *The Blackwell companion to social movements* (pp. 360-379). Blackwell.

Klandermans, B., & de Weerd, M. (2000). Group identification and political protest. In S. Stryker, T. J. Owens, & R. W. White (Eds.), *Self, identity, and social movements* (pp. 68-90). University of Minnesota Press.

Klandermans, B., & Linden, A. (2004). *Repression and stigmatization of extreme right activists*. American Sociological Association Annual Meeting 2004, San Francisco, CA, USA.

Kleinberg, B., van der Vegt, I., & Gill, P. (2021). The temporal evolution of a far-right forum. *Journal of Computational Social Science*, 4, 1-23. <https://doi.org/10.1007/s42001-020-00064-x>

Koller, V. (2009). Analysing collective identity in discourse: Social actors and contexts. *Semen*, 27.

Koller, V. (2012). How to analyse collective identity in discourse: Textual and contextual parameters. *Critical Approaches to Discourse Analysis across Disciplines*, 5(2), 19-38.

Kredens, K., & Pezik, P. (2021). *White supremacist discussion forum corpus*. Forensic Linguistic Databank. <https://fold.aston.ac.uk/handle/123456789/19>

A. E. Booth, PhD Thesis, Aston University 2023

Krendel, A. (2020). The men and women, guys and girls of the 'manosphere': A corpus-assisted discourse approach. *Discourse & Society*, 31(6), 607-630.

<https://doi.org/10.1177/0957926520939690>

Krendel, A., McGlashan, M., & Koller, V. (2022). The representation of gendered social actors across five manosphere communities on Reddit. *Corpora*, 17(2), 291-321.

<https://doi.org/10.3366/cor.2022.0257>

Lakoff, G. (1966). Stative adjectives and verbs in English. In *Report No. NSF-17 to the National Science Foundation*. Aiken Computation Laboratory, Harvard University.

Lampe, C., & Johnston, E. (2005). Follow the (slash) dot: Effects of feedback on new members in an online community. International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL, USA.

Larsen, A. G., Ellersgaard, C., & Andrade, S. (2016). *soc.ca: Specific correspondence analysis for the social sciences* (version 0.7.3). <https://CRAN.R-project.org/package=soc.ca>

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Le Roux, B., & Rouanet, H. (2010). *Multiple correspondence analysis*. SAGE.

Leppänen, S., Kytölä, S., Jousmäki, H., Peuronen, S., & Westinen, E. (2014). Entextualization and resemiotization as resources for identification in social media. In P. Seargeant & C. Tagg (Eds.), *The language of social media: Identity and community on the internet* (pp. 112-136). Palgrave Macmillan.

Leshed, G. (2005). *Posters, lurkers, and in between: A multidimensional model of online community participation patterns*. HCI International, Las Vegas, NV, USA.

Levin, B. (2002). Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioural Scientist*, 45(6), 958-988.

Lin, A., & Kutoba, R. (2011). Discourse and race. In K. Hyland & B. Paltridge (Eds.), *The Bloomsbury companion to discourse analysis* (pp. 277-290). Bloomsbury.

Lindekilde, L., Malthaner, S., & O'Connor, F. (2019). Peripheral and embedded: relational patterns of lone-actor terrorist radicalization. *Dynamics of Asymmetric Conflict*, 12(1), 20-41.

<https://doi.org/10.1080/17467586.2018.1551557>

Lindesmith, A. R., Strauss, A. L., & Denzin, N. K. (1999). *Social psychology* (8th ed.). SAGE.

Lorenzo-Dus, N., & Nouri, L. (2021). The discourse of the US alt-right online - a case study of the Traditionalist Worker Party blog. *Critical Discourse Studies*, 18(4), 410-428.

<https://doi.org/10.1080/17405904.2019.1708763>

Lüders, M. (2015). Researching social media: Confidentiality, anonymity and reconstructing online practices. In H. Fossheim & H. Ingjerd (Eds.), *Internet Research Ethics* (pp. 77-97). Cappelen Damm Akademisk.

Lyall, C., & Meagher, L. A. (2012). A Masterclass in interdisciplinarity: Research into practice in training the next generation of interdisciplinary researchers. *Futures*, 44(6), 608-617.

<https://doi.org/10.1016/j.futures.2012.03.011>

Lytra, V. (2016). Language and ethnic identity. In S. Preece (Ed.), *The Routledge handbook of language and identity* (pp. 131-145). Routledge.

Mackenzie, J. (2016). Identifying informational norms in Mumsnet talk: A reflexive-linguistic approach to internet research ethics. *Applied Linguistics Review*, 8(2-3), 293-314.

<https://doi.org/10.1515/applirev-2016-1042>

Maloney-Krichmar, D., & Preece, J. (2005). A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction*, 12(2), 201-232. <https://doi.org/10.1145/1067860.1067864>

Mandel, D. R. (2009). Radicalization: What does it mean? In T. M. Pick, A. Speckhard, & B. Jacuch (Eds.), *Home-grown terrorism* (pp. 101-113). IOS Press.

Marcoccia, M. (2004). On-line polylogues: conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics*, 36(1), 115-145.

[https://doi.org/10.1016/S0378-2166\(03\)00038-9](https://doi.org/10.1016/S0378-2166(03)00038-9)

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Massanari, A. L. (2018). Rethinking research ethics, power, and the risk of visibility in the era of the "alt-right" gaze. *Social Media + Society*, 4(2), 1-9.

<https://doi.org/10.1177/2056305118768302>

McAdam, D. (1986). Recruitment to high-risk activism: The case of Freedom Summer. *American Journal of Sociology*, 92(1), 64-90.

McAuley, J. J., & Leskovec, J. (2013). *From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews*. International World Wide Web Conference, Rio de Janeiro, Brazil.

McCarthy, J. D., & Zald, M. N. (1973). *The trend of social movements in America: Professionalization and resource mobilization*. General Learning Press.

McCauley, C., & Moskaleiko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, 30(3), 415-433.

<https://doi.org/10.1080/09546550802073367>

McCauley, C., & Moskaleiko, S. (2017). Understanding political radicalization: The two-pyramids model. *American Psychologist*, 72(3), 205-216.

<https://doi.org/10.1037/amp0000062>

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Melucci, A. (1996). *Challenging codes: Collective action in the information age*. Cambridge University Press.

Mihailidis, P., & Viotty, S. (2017). Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in “post-fact” society. *American Behavioural Scientist*, 61(4), 441-454. <https://doi.org/10.1177/0002764217701217>

Milan, S. (2013). *Social movements and their technologies: Wiring social change*. Palgrave Macmillan.

Milan, S. (2015). From social movements to cloud protesting: the evolution of collective identity. *Information, Communication & Society*, 18(8), 887-900.

<https://doi.org/10.1080/1369118x.2015.1043135>

Milbrath, L. W. (1965). *Political participation: How and why do people get involved in politics?* Rand McNally & Company.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE.

Mo, P. K. H., & Coulson, N. S. (2010). Empowering processes in online support groups among people living with HIV/AIDS: A comparative analysis of ‘lurkers’ and ‘posters’. *Computers in Human Behavior*, 26(5), 1183-1193. <https://doi.org/10.1016/j.chb.2010.03.028>

Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist*, 60(2), 161-169. <https://doi.org/10.1037/0003-066X.60.2.161>

Mondon, A., & Winter, A. (2020). *Reactionary democracy: How racism and the populist far right became mainstream*. Verso.

A. E. Booth, PhD Thesis, Aston University 2023

Moore, R. (2014). Capital. In M. J. Grenfell (Ed.), *Pierre Bourdieu: Key concepts* (pp. 98-113). Routledge.

Morozov, E. (2009). *Brave new world of slacktivism*. NPR. Retrieved 27th July 2023 from <https://www.npr.org/templates/story/story.php?storyId=104302141>

Morrison, K., & Lui, I. (2000). Ideology, linguistic capital and the medium of instruction in Hong Kong. *Journal of Multilingual and Multicultural Development*, 21(6), 471-486.
<https://doi.org/10.1080/01434630008666418>

Mourelatos, A. P. D. (1978). Events, processes, and states. *Linguistics and Philosophy*, 2(3), 415-434.

Mudde, C. (1996). The war of words: Defining the extreme right party family. *West European Politics*, 19(2), 225-248.

Mudde, C. (2019). *The far right today*. Polity.

Mutton, R., Lewis, J., & Marsden, S. (2023). *Online radicalisation: A rapid review of the literature*. Centre for Research and Evidence on Security Threats.

<https://crestresearch.ac.uk/resources/online-radicalisation-a-rapid-review-of-the-literature/>

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, 31(3), 327-362.

Nasser-Eddine, M., Garnham, B., Agostino, K., & Caluya, G. (2011). *Countering violent extremism (CVE) literature review*. Australian Government Department of Defence.

<https://apps.dtic.mil/sti/citations/ADA543686>

Nepstad, S. E. (2004). Persistent resistance: Commitment and community in the Plowshares movement. *Social Problems*, 51(1), 43-60.

Newon, L. (2011). Multimodal creativity and identities of expertise in the digital ecology of a World of Warcraft guild. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. 131-153). Oxford University Press.

Newon, L. (2016). Online multiplayer games. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 289-304). Routledge.

Newsome, H., & Grant, T. (in press). Developing a resource model of power and authority in anonymous online criminal interactions. *Language and Law / Linguagem e Direito*.

A. E. Booth, PhD Thesis, Aston University 2023

Nguyen, D., & Rosé, C. P. (2011). *Language use as a reflection of socialization in online communities*. Workshop on Language in Social Media, Portland, OR, USA.

Nini, A. (2023). *A theory of linguistic individuality for authorship analysis*. Cambridge University Press.

Nissani, M. (1997). Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research. *The Social Science Journal*, 43(2), 201-216.
[https://doi.org/10.1016/S0362-3319\(97\)90051-3](https://doi.org/10.1016/S0362-3319(97)90051-3)

Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), 7-20.
<https://doi.org/10.1007/s10660-006-5985-x>

Nonnecke, B., & Preece, J. (2000). *Persistence and lurkers in discussion lists: A pilot study*. Hawaii International Conference on System Sciences, Maui, HI, USA.

Nonnecke, B., & Preece, J. (2001). *Why lurkers lurk*. Americas Conference on Information Systems, Boston, MA, USA.

Oliver, P. (1984). If you don't do it, nobody else will: Active and token contributors to local collective action. *American Sociological Review*, 49, 601-610.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Harvard University Press.

Page, R. (2014). Hoaxes, hacking and humour: Analysing impersonated identity on social network sites. In P. Seargeant & C. Tagg (Eds.), *The language of social media: Identity and community on the internet* (pp. 46-64). Palgrave Macmillan.

Page, R., Barton, D., Unger, J. W., & Zappavigna, M. (2014). *Researching language and social media: A student guide*. Routledge.

Pasieka, A. (2019). Anthropology of the far right: What if we like the 'unlikeable' others? *Anthropology Today*, 35(1), 3-6.

Pedahzur, A., & Perliger, A. (2006). The changing nature of suicide attacks: A social network perspective. *Social Forces*, 84(4), 1987-2008.

Perliger, A., Koehler-Derrick, G., & Pedahzur, A. (2016). The gap between participation and violence: Why we need to disaggregate terrorist 'profiles'. *International Studies Quarterly*, 60(2), 220-229. <https://doi.org/10.1093/isq/sqv010>

Perry, B., & Olsson, P. (2009). Cyberhate: the globalization of hate. *Information & Communications Technology Law*, 18(2), 185-199.

<https://doi.org/10.1080/13600830902814984>

Peuronen, S. (2011). "Ride hard, live forever": Translocal identities in an online community of extreme sports Christians. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. 154-176). Oxford University Press.

Pichler, P. (2009). *Talking young femininities*. Palgrave.

Pilkington, H. (2023a, March 2nd). *Can less be more? Writing on the 'extreme-right' from an ethnographic position*. Researching the Far Right: Methods and Ethics Seminar Series, Centre for Research on Extremism, Oslo, Norway.

Pilkington, H. (2023b). Radicalization as and in process: Tracing journeys through an "extreme-right" milieu. *Studies in Conflict & Terrorism*, 0(0), 1-27.

<https://doi.org/10.1080/1057610X.2023.2169896>

Pitcavage, M. (2023). *Murder and extremism in the United States in 2022*. Anti-Defamation League. <https://www.adl.org/resources/report/murder-and-extremism-united-states-2022>

Pluempavarn, P., Panteli, N., Joinson, A., Eubanks, D., Watts, L., & Dove, J. (2011). *Social roles in online communities: Relations and trajectories*. 6th Mediterranean Conference on Information Systems, Limassol, Cyprus.

Poell, T., & van Dijck, J. (2018). Social media and new protest movements. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 546-561). SAGE.

Polletta, F. (2006). *It was like a fever: Storytelling in protest and politics*. University of Chicago Press.

Polletta, F., & Jasper, J. M. (2001). Collective identity and social movements. *Annual Review of Sociology*, 27, 283-305.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 201-223.

<https://doi.org/10.1016/j.chb.2003.10.015>

Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1), 13-32.

Preece, S. (2009). 'A group of lads, innit?': Performances of laddish masculinity in British higher education. In P. Pichler & E. Eppler (Eds.), *Gender and spoken interaction* (pp. 115-138). Palgrave Macmillan.

Prentice, S., Taylor, P. J., Hoskins, A., & O'Loughlin, B. (2011). Analyzing the semantic content and persuasive composition of extremist media: A case study of texts produced during the Gaza conflict. *Information Systems Frontiers*, 13, 61-73.

<https://doi.org/10.1007/s10796-010-9272-y>

Putnam, R. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6, 65-78.

Radin, P. (2006). "To me, it's my life": Medical communication, trust, and activism in cyberspace. *Social Science & Medicine*, 62(3), 591-601.

<https://doi.org/10.1016/j.socscimed.2005.06.022>

Resnick, P. J., Janney, A. W., Buis, L. R., & Richardson, C. R. (2010). Adding an online community to an internet-mediated walking program part 2: Strategies for encouraging community participation. *Journal of Medical Internet Research*, 12(4), e72.

<https://doi.org/10.2196/jmir.1339>

Rickford, J., & Price, M. (2013). Girls II women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics*, 17(2), 143-179.

Roth, S. (2019). Linguistic capital and inequality in aid relations. *Sociological Research Online*, 21(1), 38-54. <https://doi.org/10.1177/1360780418803958>

Rüdiger, S., & Dayter, D. (2017). The ethics of researching unlikeable subjects. *Applied Linguistics Review*, 8(2-3), 251-269. <https://doi.org/10.1515/applirev-2016-1038>

Rudolf von Rohr, M.-T., Thurnherr, F., & Locher, M. A. (2019). Linguistic expert creation in online health practices. In P. Bou-Franch & P. Garcés-Conejos Blitvich (Eds.), *Analyzing digital discourse: New insights and future directions* (pp. 219-250). Palgrave Macmillan.

Rydgren, J. (Ed.). (2018). *The Oxford handbook of the radical right*. Oxford University Press.

Sandell, R. (1999). Organizational life aboard the moving bandwagons: A network analysis of dropouts from a Swedish temperance organization. *Acta Sociologica*, 42(1), 3-15.

Saunders, C. (2008). Double-edged swords? Collective identity and solidarity in the environment movement. *The British Journal of Sociology*, 59(2), 227-253.

<https://doi.org/10.1111/j.1468-4446.2008.00191.x>

Saunders, C., Grasso, M., Olcese, C., Rainsford, E., & Rootes, C. (2012). Explaining differential protest participation: Novices, returners, repeaters, and stalwarts. *Mobilization: An International Quarterly*, 17(3), 263-280.

<https://doi.org/10.17813/maiq.17.3.bqm553573058t478>

Sayers, A. (2007). Tips and tricks in performing a systematic review: Chapter 2, Preliminary evidence gathering: snowballing and reverse snowballing. *British Journal of General Practice*, 57(542), 759.

Schmid, A. P. (2013). *Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review*. International Centre for Counter-Terrorism.

<https://www.icct.nl/publication/radicalisation-de-radicalisation-counter-radicalisation-conceptual-discussion-and>

Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 170-183). SAGE.

Schussman, A., & Soule, S. A. (2005). Process and protest: Accounting for individual protest participation. *Social Forces*, 84(2), 1083-1108. <https://doi.org/10.1353/sof.2006.0034>

Scott, J. (Ed.). (2014). *A dictionary of sociology* (4th ed.). Oxford University Press.

Scott, S., & Hardie-Bick, J. (2022). Moral career. In M. H. Jacobsen & G. Smith (Eds.), *The Routledge international handbook of Goffman studies* (pp. 74-84). Routledge.

Scrivens, R., Davies, G., & Frank, R. (2020). Measuring the evolution of radical right-wing posting behaviors online. *Deviant Behavior*, 41(2), 216-232.

<https://doi.org/10.1080/01639625.2018.1556994>

Scrivens, R., Wojciechowski, T. W., & Frank, R. (2022). Examining the developmental pathways of online posting behavior in violent right-wing extremist forums. *Terrorism and Political Violence*, 34(8), 1721-1738. <https://doi.org/10.1080/09546553.2020.1833862>

Seargeant, P., & Tagg, C. (2014). Introduction: The language of social media. In P. Seargeant & C. Tagg (Eds.), *The language of social media: Identity and community on the internet* (pp. 1-20). Palgrave Macmillan.

Shin, H. (2014). Social class, habitus, and language learning: The case of Korean early study-abroad students. *Journal of Language, Identity, and Education*, 13(2), 99-103.

<https://doi.org/10.1080/15348458.2014.901821>

Silver, R. E. (2005). The discourse of linguistic capital: Language and economic policy planning in Singapore. *Language Policy*, 4, 47-66.

Simpson, P. (1993). *Language, ideology and point of view*. Routledge.

Singh, V. (2012). *Newcomer integration and learning in technical support communities for open source software*. 17th ACM International Conference on Supporting Group Work, Sanibel Island, FL, USA.

Smith, C. S. (1991). *The parameter of aspect*. Kluwer Academic Publishers.

Smith Risser, H., & Bottoms, S. (2014). "Newbies" and "Celebrities": Detecting social roles in an online network of teachers via participation patterns. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 433-450. <https://doi.org/10.1007/s11412-014-9197-4>

Smits, J., & Gündüz-Hoşgör, A. (2003). Linguistic capital: Language as a socio-economic resource among Kurdish and Arabic women in Turkey. *Ethnic and Racial Studies*, 26(5), 829-853. <https://doi.org/10.1080/0141987032000109050>

Snow, D. A. (2001). Collective identity and expressive forms. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 2212-2219). Pergamon.

Snow, D. A., Zurcher Jr., L. A., & Eklund-Olson, S. (1980). Social networks and social movements: A microstructural approach to differential recruitment. *American Sociological Review*, 45(5), 787-801.

Somma, N. M. (2009). How strong are strong ties? The conditional effectiveness of strong ties in recruitment attempts. *Sociological Perspectives*, 52(3), 289-308. <https://doi.org/10.1525/sop.2009.52.3.289>

Strauss, A. L. (1959). *Mirrors and masks*. The Free Press.

Sveningsson Elm, M. (2009). How do various notions of privacy influence decisions in qualitative internet research? In A. N. Markham & N. K. Baym (Eds.), *Internet inquiry: Conversations about method* (pp. 69-87). SAGE.

Sveningsson, M. (2003). Ethics in internet ethnography. In E. A. Buchanan (Ed.), *Readings in virtual research ethics: Issues and controversies* (pp. 45-61). Information Science Printing.

Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge.

Tagg, C. (2020). English language and social media. In S. Adolphs & D. Knight (Eds.), *The Routledge handbook of English language and digital humanities* (pp. 568-586). Routledge.

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behaviour. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 367-390). Psychology

Press. (Reprinted from *Psychology of intergroup relations*, pp. 7-24, by S. Worchel and W. G. Austin, Eds., 1986, Nelson-Hall Publishers)

Taylor, V., Kimport, K., Van Dyke, N., & Andersen, E. A. (2009). Culture and mobilization: Tactical repertoires, same-sex weddings, and the impact on gay activism. *American Sociological Review*, 74(6), 865-890.

Taylor, V., & Whittier, N. E. (1992). Collective identity in social movement communities: Lesbian feminist mobilization. In A. D. Morris & C. M. Mueller (Eds.), *Frontiers in social movement theory* (pp. 104-129). Yale University Press.

Thompson, J. B. (1991). Editor's introduction. In J. B. Thompson (Ed.), *Language and symbolic power* (pp. 1-31). Polity.

Thurlow, C., & Mroczek, K. (2011). Introduction: Fresh perspectives on new media sociolinguistics. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. xix-xliv). Oxford University Press.

Tilly, C. (1993). Contentious repertoires in Great Britain, 1758-1834. *Social Science History*, 17(2), 253-280.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10(4). <https://doi.org/10.1111/j.1083-6101.2005.tb00270.x>

Vaast, E. (2007). Playing with masks: Fragmentation and continuity in the presentation of self in an occupational online forum. *Information Technology & People*, 20(4), 334-351. <https://doi.org/10.1108/09593840710839789>

Valocchi, S. (2012). Activism as a career, calling, and way of life. *Journal of Contemporary Ethnography*, 42(2), 169-200. <https://doi.org/10.1177/0891241612467202>

van Dijk, T. A. (1998). Opinions and ideologies in the press. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse* (pp. 21-63). Blackwell.

Van Dyke, N., & Dixon, M. (2013). Activist human capital: Skills acquisition and the development of commitment to social movement activism. *Mobilization*, 18(2), 197-212.

van Uden-Kraan, C. F., Drossaert, C. H. C., Taal, E., Seydel, E. R., & van de Laar, M. A. F. J. (2008). Self-reported differences in empowerment between lurkers and posters in online patient support groups. *Journal of Medical Internet Research*, 10(2), e18. <https://doi.org/10.2196/jmir.992>

van Wyk, I. (2013). Beyond ethical imperatives in South African anthropology: Morally repugnant and unlikeable subjects. *Anthropology Southern Africa*, 36(1-2), 68-79.

<https://doi.org/10.1080/02580144.2013.10887025>

Verba, S., Scholzman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.

Viégas, F. B., & Smith, M. (2004). *Newsgroup Crowds and AuthorLines: Visualizing the activity of individuals in conversational cyberspaces*. 37th Hawaii International Conference on System Sciences, Big Island, HI, USA.

Walton, S., & Jaffe, A. (2011). "Stuff white people like": Stance, class, race, and internet commentary. In C. Thurlow & K. Mroczek (Eds.), *Digital discourse: Language in the new media* (pp. 199-219). Oxford University Press.

Wang, Y.-C., Kraut, R., & Levine, J. M. (2012). *To stay or leave?: The relationship of emotional and informational support to commitment in online health support groups*. ACM 2012 Conference on Computer Supported Cooperative Work, Seattle, WA, USA.

Waters, J., & Gasson, S. (2005). *Strategies employed by participants in virtual learning communities*. 38th Hawaii International Conference on System Sciences, Big Island, HI, USA.

Wellman, B., Quan-Haase, A., Boase, J., Chen, W., Hampton, K. N., Díaz, I., & Miyata, K. (2003). The social affordances of the internet for networked individualism. *Journal of Computer-Mediated Communication*, 8(3). <https://doi.org/10.1111/j.1083-6101.2003.tb00216.x>

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). *Finding social roles in Wikipedia*. iConference, Seattle, WA, USA.

Welser, H. T., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8(2), 1-32.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business School Press.

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends*, 55(1), 22-45.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A grammar of data manipulation* (version 1.0.5). <https://CRAN.R-project.org/package=dplyr>

Wiktorowicz, Q. (2005). *Radical Islam rising: Muslim extremism in the West*. Rowman & Littlefield.

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J. (2020). Online extremism: Research trends in internet activism, radicalization, and counter-strategies. *International Journal of Conflict and Violence*, 14(2), 1-20.

<https://doi.org/10.4119/ijcv-3809>

Wodak, R., de Cillia, R., Reisigl, M., & Liebhart, K. (2009). *The discursive construction of national identity* (2nd ed.). Edinburgh University Press.

Wodak, R., & Meyer, M. (2015). *Methods of critical discourse studies*. SAGE.

Wohlin, C. (2014). *Guidelines for snowballing in systematic literature studies and a replication in software engineering*. 18th International Conference on Evaluation and Assessment in Software Engineering, London, UK.

Wong, M. A., Frank, R., & Allsup, R. (2015). The supremacy of online white supremacists – An analysis of online discussions by white supremacists. *Information & Communications Technology Law*, 24(1), 41-73. <https://doi.org/10.1080/13600834.2015.1011845>

Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: Developing a textsensitive methodology for authorship research. *International Journal of Speech, Language, and the Law*, 20(1), 45-75. s

Wright, D. (2017). Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics*, 22(2), 212-241. <https://doi.org/10.1075/ijcl.22.2.03wri>

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, 13(5), 788-806. <https://doi.org/10.1177/1461444810385097>

Zickmund, S. (1997). Approaching the radical other: The discursive culture of cyberhate. In S. G. Jones (Ed.), *Virtual culture: Identity and communication in cybersociety* (pp. 165-182). SAGE.

Zihnioglu, Ö. (2023). Strategizing post-protest activism in abeyance: retaining activist capital under political constraint. *Social Movement Studies*, 22(1), 122-137. <https://doi.org/10.1080/14742837.2021.2003193>

List of appendices

Appendix Ai	Ethics form submitted to Aston Institute for Forensic Linguistics Research Ethics Committee
Appendix Aii	Follow-up questions and responses submitted to Aston Institute for Forensic Linguistics Research Ethics Committee
Appendix Aiii	Outcome letter from Aston Institute for Forensic Linguistics Research Ethics Committee
Appendix B	Full dataset used for multiple correspondence analysis
Appendix C	R script used for multiple correspondence analysis
Appendix D	Full set of metadata as received
Appendix E	R script used for data restructuring and sequence analysis
Appendix F	Frequency data for all 4,225 career sequences
Appendix Gi	Mayflies subcorpus
Appendix Gii	Dragonflies subcorpus
Appendix Giii	Mantises subcorpus
Appendix Giv	Cicadas subcorpus
Appendix H	Frequency and keyness information for key n-grams
Appendix li	Corpus for User 1
Appendix lii	Corpus for User 2
Appendix liii	Corpus for User 3
Appendix liv	Corpus for User 4
Appendix lv	Corpus for User 5