IF IT’S NOT SCOTTISH IT’S CRAP!: WHAT TWITTER AND THE SCOTTISH REFERENDUM CAN TELL US

by

Marcella Harrison Morris
A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Conflict Analysis and Resolution

Committee:
_______________________________ Chair of Committee

_______________________________

_______________________________

_______________________________ Graduate Program Director

_______________________________ Dean, School for Conflict Analysis and Resolution

Date: __________________________ Spring Semester 2015
George Mason University
Fairfax, VA
If it’s not Scottish, it’s Crap!: What Twitter and the Scottish Referendum Can Tell Us

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Marcella Harrison Morris, M.S.
Masters of Science
George Mason University, 2015

Director: Thomas E. Flores, Professor
School for Conflict Analysis and Resolution

Spring Semester 2015
George Mason University
Fairfax, VA
This work is licensed under a creative commons attribution-noderivs 3.0 unported license.
DEDICATION

To all the giants who put me on their shoulders
ACKNOWLEDGEMENTS

I would like to thank the family, friends, and others who have made this thesis possible. To my parents, who supported me through this and never said anything but ‘Yes’ to my educational goals, I owe the most gratitude. To my colleagues at START, faculty at UMD, and GMU without their support, direction, time, and speedy emails when most needed, I would not have made it here. Specific thanks go to: Brian Wingenroth, who converted all the data, Brandon Behlendorf, who asked the right questions to me discover the path, and Kate Izsak who allowed so much flexibility in scheduling and direction to allowed me to produce this thesis, to keep my sanity, and keep my job. And finally, I want to thank James Walston who tied it all the disparate pieces I had picked up along the way together for the first time He was the first to help me see where I wanted to go and whose memory continues to inspire my journey.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Scottish Background</td>
<td>1</td>
</tr>
<tr>
<td>Other Independence Referenda</td>
<td>5</td>
</tr>
<tr>
<td>Why Twitter?</td>
<td>8</td>
</tr>
<tr>
<td>Who Uses Twitter?</td>
<td>10</td>
</tr>
<tr>
<td>How is Twitter Different?</td>
<td>11</td>
</tr>
<tr>
<td>Overall Methodology</td>
<td>14</td>
</tr>
<tr>
<td>Filtering tweets from the Firehose</td>
<td>14</td>
</tr>
<tr>
<td>Sampling from the Corpus</td>
<td>15</td>
</tr>
<tr>
<td>Coding by Hand</td>
<td>17</td>
</tr>
<tr>
<td>Location and Time Zones</td>
<td>17</td>
</tr>
<tr>
<td>What Lies Ahead</td>
<td>19</td>
</tr>
<tr>
<td>Referendum Prediction and Results with Twitter</td>
<td>21</td>
</tr>
<tr>
<td>Traditional Election Prediction—To what Twitter-Based Research is Compared?</td>
<td>21</td>
</tr>
<tr>
<td>A Brief History of Polling (in American Politics)</td>
<td>22</td>
</tr>
<tr>
<td>Challenges of Traditional Techniques and Sampling</td>
<td>24</td>
</tr>
<tr>
<td>What Do We Know About Twitter Research and Elections?</td>
<td>27</td>
</tr>
<tr>
<td>We Can Predict Elections on Twitter</td>
<td>27</td>
</tr>
<tr>
<td>What we Know about the Challenges to Prediction on Twitter</td>
<td>30</td>
</tr>
<tr>
<td>What are the Tradeoffs between Polling and Twitter?</td>
<td>35</td>
</tr>
<tr>
<td>Research Design</td>
<td>36</td>
</tr>
<tr>
<td>Coding Rules</td>
<td>37</td>
</tr>
<tr>
<td>Discussion and Results</td>
<td>40</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: Total Count of tweets by region by day</td>
<td>37</td>
</tr>
<tr>
<td>Table 2: Percentages of tweets by topic in each region</td>
<td>39</td>
</tr>
<tr>
<td>Table 3: Percentage breakdown of emotionality by region on September 19</td>
<td>80</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Count of all tweets over time.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Tweets over time broken down by region.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Tweets only from Scotland over time.</td>
<td>43</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Total number of tweets by region over the three day period.</td>
<td>44</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Tweets broken down by day, region, and content of the message.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Count over time of important categories of tweets over time.</td>
<td>47</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Total proportions of Yes/No and undecided tweets over the three-day period.</td>
<td>49</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Total proportion of Yes/No and undecided tweets broken down by day.</td>
<td>49</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Total proportion of Yes/No and undecided tweets broken down by region.</td>
<td>50</td>
</tr>
<tr>
<td>Figure 10</td>
<td>All tweets based on emotionality by day.</td>
<td>76</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Count of emotionality over the full three days by region.</td>
<td>77</td>
</tr>
<tr>
<td>Figure 12</td>
<td>This chart collapsed the overall emotions into four categories progressing up</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>the scale of emotion. 1: neutral tweets, 2: slight emotion, 3: moderate emotion, 4: extreme emotion.</td>
<td></td>
</tr>
<tr>
<td>Figure 13</td>
<td>Emotionality of tweets from September 19 by region.</td>
<td>82</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Tweet Coded as a 0 or neutral message (account).</td>
<td>83</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Tweet coded as a -1 or slight emotion favoring the No campaign (Pilko).</td>
<td>84</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Tweet coded as a 1 or slight emotion favoring the Yes campaign (Podcast).</td>
<td>85</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Tweet coded -2 or moderate emotion favoring the No campaign (Miller).</td>
<td>86</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Tweet Coded 2 or moderate emotion favoring the Yes campaign (King).</td>
<td>87</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Tweet coded as a -3 or extreme emotion favoring the No campaign (McCabe).</td>
<td>89</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Tweet coded as a 3 or extreme emotion favoring the Yes campaign (Ashworth)</td>
<td>90</td>
</tr>
</tbody>
</table>
IF IT’S NOT SCOTTISH, IT’S CRAP!: WHAT TWITTER AND THE SCOTTISH REFERENDUM CAN TELL US

Marcella Harrison Morris, M.S.
George Mason University, 2015
Thesis Director: Dr. Thomas E. Flores

On September 18, 2014 after nearly two years of intense campaigning, a referendum was held in Scotland for voters to decide if Scotland would declare independence. After a contentious campaign, 55% of Scottish voters rejected independence. This thesis tackles the question of what Twitter can tell us about the Scottish referendum. This is achieved by breaking down the question in two parts: first, can Twitter have predictive qualities and had we looked in the right place could we have seen the divisions between 45% and 55% found in the results beforehand? Second, what can Twitter show an observer about national identity construction in the lead up to and wake of a referendum of this nature? On the question of predictability, we find that there is little ability for Twitter to accurately predict an election with a strong silent majority, but that Twitter does allow for superb observation of the public discourse and allows for observers to understand the construction and changes in identity as individuals are faced with serious questions that
could radically change the status quo. Through this study, we begin to unpack the hugely powerful tool of Twitter and gain otherwise invisible insights into what Scottish identity definitions may now include.
INTRODUCTION

Scottish Background

On September 18, 2014, political life in Scotland changed. People took to the polls to share their voice to answer a 300-year old question—should Scotland become an independent nation? The story of Scottish independence arguably begins in 1701 with the unification of Scotland and the rest of the United Kingdom through the Act of Union where the Kingdom of Scotland joined with the United Kingdom and the first Scottish Parliament was disbanded. While the history of Scotland dates back centuries before that and the push for Scottish independence for some began with the Act of Union, the Scottish National Party (SNP) formed in 1934 with the joining of the Scottish Party and the National Party (About Us). From the party’s formation through to Scottish devolution in 1997, the SNP was generally a fringe party—never a big player in British politics, but still standing candidates for election and sending some to Westminster (Schwartz).

In Scotland, British politics have always had the upper hand even though the SNP became the biggest party elected in Scotland with approximately 35% of the Scottish vote in 1968 (Schwartz 197). This English upper hand is displayed throughout political history in the UK especially with the sweeping victory Margaret Thatcher and the Tories saw in the 1970s, which extended north into Scotland at the same time a push for devolution and additional Scottish power for Scotland failed. The sweeping victories of the Tories and
then shortcomings of the independence movement ended the hope of establishing a sub-
state Scottish Parliament for nearly thirty years (Black: Profile).

At this time, with the SNP pushing toward national sovereignty through
independence, it would be logical, even expected, to have seen violence associated with
the push for independence. However, the SNP did not condone, let alone advocate for,
vigilante violence to meet these goals (Schwartz 4). Many may note that this political, non-violent
process is in stark contrast to the movements in Northern Ireland against the union of
Great Britain occurring at this time. In fact, John Shwartz reports from surveys conducted
of local SNP leaders, “most felt that any member who did use violence ought to be
expelled from the party” (498).

The SNP continued to see electoral struggle through 1980s and 1990s, during
which time the party leadership, including Alex Salmond, worked hard to pull party
members into shape and to push for devolution and independence. It was nearly ten years
after devolution in 1997 that the SNP won the status of opposition government in 2007. It
was not until 2011 that the SNP was elected to majority government status (Broughton,
About Us). It was after victory in 2011 that SNP leader, Alex Salmond, declared a call
for a referendum on independence to be held within five years of the SNP’s victory
(About Us). This long awaited referendum occurred on September 18, 2014.

This study focuses on two questions of Twitter spurred by the referendum: could
we have predicted the results had we looked at traffic on Twitter, and how did people
interact on Twitter when making sense of the results of the election, particularly when it
comes to Scottish identity construction? However, it is important to begin by knowing how things played out during the referendum.

On the morning of September 18, 2014, individuals, over sixteen years of age, living in Scotland, and pre-registered to vote, took to the polls to answer the question, “Should Scotland be an independent country?” (“Scottish Independence Referendum - GOV.UK”). The opportunity and the legal provisions for it to occur and have binding implications\(^1\) for the future of the United Kingdom were established on October 15\(^{th}\), 2012 by Alex Salmond, First Minister of Scotland, and David Cameron, Prime Minister of the United Kingdom of Great Britain and Northern Ireland. These events of independence referendum and campaigns were spurred by the SNP victory and majority government status won in 2011 (“Scottish Independence Referendum - GOV.UK”).

By the end of the day on the 18\(^{th}\), 3,619,915 voters had cast a ballot.\(^2\) 84.6% of the eligible population voted in the referendum. This level of turnout broke UK records held since universal suffrage laws were enacted dating back to 1950. Further, this level of turnout is high, but it does not measure up to the Eritrean vote for independence in 1993, which reached 98.5% or the vote to establish South Sudan, which was measured at 97.5%\(^3\) (Parmelee, Kron and Gettleman).\(^4\) In Scotland, turnout ranged from 91%\(^5\) to 75%

\(^1\) The Scottish referendum, had it passed, would have led to Scotland’s process of independence which would have begun immediately, unlike the recent votes seen in Catalonia.

\(^2\) There was very little instance of ballots being rejected in the referendum, election administrators rejected 0.1% of ballots (3,429 ballots, to be exact) (“Scotland’s Independence Referendum”).

\(^3\) Although, some precincts in South Sudan reported over 100% turnout.

\(^4\) While the Eritrea and South Sudan numbers are at least 10 points higher than those from Scotland, it is important to consider the differences in the conflicts spurring the votes in these contexts
across Scotland. Glasgow and Dundee, generally seen as Yes Campaign strongholds, saw turnout on the lower end (75% and 78.8% respectively) (“Scottish Independence Referendum”).

This high turnout translated into a result of 44.7% vote for Yes and 55.3% vote for No, or 1.6 Million Yes votes and voters, and 2 million No votes and voters. The close nature of the referendum vote was interestingly split in individual council votes. The Yes campaign won between 33.4% \(^6\) and 57.3% \(^7\) of every council but they won only 4 of the 32 Councils outright (“Scottish Independence Referendum”, “SCOTLAND VOTED ‘NO’ ON INDEPENDENCE”).

At 6:18 am on September 19\(^{th}\), 2014, Alex Salmond conceded the referendum and nearly two years of constant campaigning was over (“Scottish Independence Referendum: Final Results in Full”).

While this seems like an abrupt ending, the conversation immediately shifted from the results to the future and the promises of future devolution of powers to Scotland. Further, pundits stressed the importance of the spectacular turnout and enfranchisement of 16 and 17 year-olds in the political process (“Scottish Independence Referendum: Final Results in Full’’). This analysis of the role democracy and enfranchisement played

---

\(^5\) 91% turnout was seen in East Dumbartonshire (“Scottish Independence Referendum”).

\(^6\) The Scottish Borders council reported the lowest amount of support for the Yes campaign, 33.4% (“Scottish Independence Referendum”).

\(^7\) Dundee council reported the highest amount of support for the Yes campaign, 57.3% (“Scottish Independence Referendum”).
in this referendum was reflected in the Twitter traffic analyzed in this study, as we will see below.

**Other Independence Referenda**

While the concept of voting for independence is not foreign idea in modern politics, a fully recognized and accepted vote is rare even in countries with long standing histories of referendums on the topic of independence. Since 1999, few votes on independence have occurred that were close to successful and recognized by the parent state. Notable examples include East Timor in 1999 in their push for independence from colonial hold overs (BBC News), Montenegro in 2006’s final step in the slow, peaceful divorce from Serbia (Wood), and South Sudan’s 2011 end to a bloody civil war for independence (Gettleman). There are recent examples of votes for independence in territories and cities in Ukraine during the current conflict there (“Referendum Results in Donetsk and Lugansk Regions Show Landslide Support for Self-Rule”), and of course the November 9th vote in Catalonia for independence from Spain (The Economist).

However, neither of these examples was sanctioned by the necessary powers as was seen in Scotland in 2014.

With these in mind, we turn to the events in Scotland. Here, we see an established democracy with consent from all sides coming together to open the polling places and ballot boxes to establish the will of the people outside of a conflict setting, or really even a sizeable political need for independence. The United Kingdom has consistently ranked highest score possible, a score of 10 on the Polity Index (Marshall). Thus, rather than working to throw off a yoke of colonialism or an oppressive, authoritarian regime, the
Scottish independence referendum took place in and tested one of the world’s renowned democracies at a time of relative stability. When considering the Scottish vote and results we are struck with two natural comparison groups—the group with other nationalist movements actively making strides toward independence and the group of wealthy democracies with regionalist movements rising to positions that are more prominent in society. In this case, the comparison is clearer, while still not perfect, with the processes of sovereignty in Puerto Rico and Quebec.

With the Quebec process toward independence, two referendums have been held. The first was in 1980, where 40% voted for succession. The second vote, in 1995, saw more success with a 50%-49% vote for union (Bryant). In the case of Quebec—like that of Scotland—monetary policy, economic issues, and other important policy questions were never truly answered before either vote and led to uncertainty in the impact of a successful secessionist vote (Bryant). With the 1995 referendum, it became abundantly clear that there was a lack of path forward for a secessionist vote in Canada. Hèrbert reports that voters were never quite sure what the over-100-word-referendum question actually meant let alone what it would result in policy wise. Some voters apparently

---

8 While the Scottish referendum is framed in stability, there are obvious parallels with other more violent pushes for self-determination in the United Kingdom and Western Europe. As seen with the Catalonian interest and support represented in the Twitter traffic, we can see others looked to Scotland for inspiration and a glimpse into the future. Without a doubt, the Kurds, Northern Irish, Basque, and Tamil populations, just to name a few, were paying close attention to the results and process. A key distinction here however, is the history of violence. While there was terrorism in Scotland associated with independence, the activities of the ‘Tartan Terrorists’ or the Scottish National Liberation Army pale in comparison to those of the IRA, PKK, or LTTE.

9 Independence movements in Puerto Rico or Quebec, for instance.
believed it would spur a second referendum to verify the first (Hèrbert). Further, the
devolution and additional rights and recognition of heritage in Quebec seen in the 1960s
and 1970s, “appears to have dampened the separatist spirit” (Bryant). In the case of the
1995 referendum, we see similar politics in the run up to the vote. In September 2014,
leadership from Westminster and other British political parties converged on Scotland for
one final push and final decisions and rulings on important policy issues were announced.
In Canada, Quebec saw politicians make promises and voice intention to meet some
concerns of the region immediately before the vote (Quebec Referendum (1995)).

In Puerto Rico, votes on independence have been cast and counted more recently,
but they have been similarly split on the issues. In 2012, the United States supported a
two-question referendum for statehood in Puerto Rico and it passed by a small majority
leaving the island staunchly divided (Fox and Coto). While President Obama promised to
honor the will of the people of Puerto Rico as voiced through the referendum, what form
that support would take is still unknown as congressional approval would be required for
the options for additional sovereignty or independence (Fox an Coto). Part of the
confusion was that the referendum in Puerto Rico asked two questions. First, voters were
asked whether voters wanted to maintain the status as a Commonwealth of the United
States—a proposition that 52% of Puerto Ricans voted against (Planas). Second, voters
were asked if they wanted, “to become a U.S. State, an independent country, or a freely
associated state” (Planas). Here, 61% opted for statehood, but so many left the question
blank intentionally, the actual count was around 45% of voters opting for statehood—
leaving the referendum without a clear majority on a preferred path forward (Planas). To
add confusion to the question of statehood and independence in Puerto Rico, at the same
time voters answered the referendum question, they also voted out the pro-statehood
government that held power at the time on the island (Fox and Coto).

While we have similarly comparable cases for the Scottish one, with Canada and
the United States, we see clearly that the Scottish referendum on independence went the
furthest in guaranteeing independence with a yes vote on September 18, 2014. Also,
interestingly in this brief comparison, there are similarities in the political promises and
concessions made by the national governments to maintain the status quo, similar
unresolved currency and policy questions, and interestingly—similar vote tallies. What
these three cases leave an observer with is an understanding that a vote can happen
peacefully, but that societies are deeply divided on these questions of independence.

**Why Twitter?**

Building off the understanding that the democratic nature of the Scottish
referendum played a critical role in the process as compared to other pushes for
nationhood in the world, we are drawn to the role of the individual. In the Scottish
system, the individual is the one deciding which box to check in the ballot box. Here, the
individual is the unit of analysis for decision makers in a democracy even when dealing
with a nationalistic question.  

In the process of identity construction, individuals define and redefine what a certain identity means to
them. This is negotiation process between the individual and broader group identity. In the case of
Scotland, in the time immediately surrounding the announcement of the result, individuals are grappling

---

10 In the process of identity construction, individuals define and redefine what a certain identity means to
them. This is negotiation process between the individual and broader group identity. In the case of
Scotland, in the time immediately surrounding the announcement of the result, individuals are grappling
In years past, the best option may have been newspaper editorials or opinion polling across the board as they capture at least some individuals who are engaged with the important changes in society. However, as times and technologies change, so should our research methods. With the advent of social media and the micro blog site, Twitter, we are able to directly observe commentary and snippets of dialogue between individuals engaging with the events of the referendum in real time—something traditional polling could never offer. Thus allowing for a better sense of how and when public discourse shifts on a certain topic which, in turn, allows a more robust understanding of the impact of events on sentiment and time series analysis of this data.

The social media site, Twitter, was designed to answer the question, “‘what is happening?’ in 140 characters or less,” has quickly expanded its reach—in numbers of users, in number of tweets, and in the opportunity for research based on these micro-blogs since it’s opening in March of 2006 (Murthy 2, “Milestones”). Through individual accounts of what is happening, observers and researchers are able to watch the individual voter or onlooker engage with the political system, campaign for a cause, recruit others to their cause, and even grapple with the results as they are announced and the reality of what has transpired becomes real to the individual user. Twitter allows for all of this to be understood and analyzed at the individual, granular level. No longer must we rely on editorial pages, which are chosen by editorial staffs and not fully representative of the population who have the ability to sway society through the ballot box. While there are
issues with representative sampling of a population through Twitter, it is a fundamentally more inclusive and accessible system than the editorial pages of the ‘old media’. For Twitter, only character count limits your message, Messages, on average, include 11 words (O’Connor et al.). Users can send as many 140-character messages as they wish, and as many users can join—as long as they have internet connectivity. That is to say, there are no column inches to track, nor a limited numbers of writers per edition of the paper or magazine.

Who Uses Twitter?

When unpacking the black box that is Twitter, it is important to understand the demographics behind the system. According to beeolve.com’s extensive analysis of global Twitter use published in 2012, we find that 53% of Twitter users are female and 47% are male. Additionally, females not only outnumber males in user profiles, they also send more tweets overall. Further, 73% of Twitter users who reported their age were between 15 and 25 years old. 26-35-year-olds make up the next higher percentage at nearly 15% with older groups dropping precipitously after 35 years of age. While these figures likely track with expectations, the reporting of age, like other self-reported demographics, are hardly reliable measures, less than one-half of one-percent of Twitter users report their age on their Twitter profiles. When it comes to usage by country, the rates vary as well, the United States tops the list with 50% of world Twitter usage, and

---

11 Twitter, like other online platforms, struggles with ‘bots’ or trolls, who operate as spammers to jam the airways with hateful messages or ones that promote their cause through automated systems. This study dives into the Twitter traffic that discussed the Scottish referendum in totality, rather than skimming out accounts suspected of trolling and spamming. Here we are able to see a complete picture of the Twittersphere. Additional research could take different approaches, including looking only at original Twitter messages rather than retweets to help account for more organic messages.
the UK comes in at 17%, based on bee evolve.com’s 2012 report. This usage lines up with the projections of Twitter penetration in the United Kingdom reported by eMarketer.com, who reported 28% of the world’s usage of Twitter originating in the UK in 2013, up from just 13% in 2011.12

Users, ranging from celebrities to popular brands and run of the mill individual citizens, share a myriad of important and less than important developments in their lives. User profiles, which are customizable and make tweets sent by the user easily accessed, allows for individual expression of the self (Murthy 4). Content of tweets vary widely in importance. They can include anything from how terrible the user’s morning commute was to how they voted in a referendum that day. This study fits in the midst of all this. People, including individual voters, talked about Scotland and the prospects for independence, shared news, jokes, opinions, and poll results about the referendum. This data are rich for analysis to understand the minute-to-minute changes in what people think and feel about new changes and challenges.

**How is Twitter Different?**

Here, we have a social media platform that allows for public communication mixed with conversation and other forms of broadcasting of personal ideas. Jan-Hinrik Schmidt puts forward in his chapter, *Twitter and the Rise of Personal Publics* that Twitter and other platforms form “a particular communicative space which is affording the emergence of a new type of publicness: the “personal public”” (Schmidt 4). For Schmidt

12 To put these global usage rates in perspective, with the US around 50% and the UK around 20%, the next country on the list was Australia with barely 4% of all Twitter traffic, showing a larger corner on the global market going to the UK, even if the US is still bigger overall (Beevolve.com).
‘personal publics’ should constitute a new sphere that contrasts with the ‘traditional publics’ of old media, to included news media and other traditional news communication mechanisms. New, ‘personal publics’ are defined by three elements according to Schmidt. First, individuals select content “of personal relevance” rather than relying on journalists and ‘old media’ or news providers (4). This personal relevance relates also to content creation according to Tsou et al. who argue that Twitter not only allows for the exchange of ideas it also allows for ideas to span the globe “without physical geographic impediments” (Tsou et al. 337). Further, the information promoted on Twitter that is no longer tied down by geography reflects the user’s personal preferences or political biases. While no authors argue a tweet equates directly to a vote, Ko et al. base their study of the South Korean election on the concept that conservative and liberal biases of Twitter users is directly reflected in the types of articles they share or retweet (143).

Next, with of Twitter’s shift from traditional publics to personal publics, the content selected is shared with a purposefully selected audience (Schmidt 4). In the case of Twitter, this audience is the Twitter user’s ‘followers’ or other Twitter users who self-select to ‘follow’ other users so as to be included in what that individual shares on Twitter rather than the undirected general news media’s mass appeal with articles for their broad reader or viewer base. With the added mechanisms that allow for the direction of a message to another user, the personal public of Twitter allows for direct contact and connectivity with two types of users, the active and the passive, as discussed by
Huberman, Romero, and Wu, who delineate between ‘followers’ and ‘friends’ in the Twittersphere.

Finally, Schmidt includes that Twitter content is presented conversationally. This conversational aspect of Twitter is reinforced in Shamma et al.’s work where they distinguish “what is conversational and what is a comment” through analysis of Twitter traffic American Presidential debates. Twitter is interactive; users can direct a message at others and link to trending topics and current events (Schmidt, Shamma et al.). Axel Mairder and Julian Ausserhofer discuss similar topics in detail in their chapter in *Twitter and Society; “Political Discourses on Twitter: Networking Topics, Objects, and People*” where they argue, “[o]n Twitter, however, political news were not only reported on, but also interpreted and actively connected to other topics by the users at the same time” (307). This is in direct contrast to Schmidt’s ‘traditional publics’, a concept that stems from the “traditional Two-Step-Flow of Communication model” describing the “interrelation of interpersonal and mass communication” (Mairder and Ausserhofer 307). Within the Twittersphere, we see that the traditional print and consume process of ‘traditional publics’ is no longer a complete process because “reports by news media and interpretation by the personal social networks become part of the same news stream” resulting in a hybrid of news and opinion in each tweet (Mairder and Ausserhofer 307).

---

13 This distinction becomes important when considering the commentary and conversational aspects of Twitter dialogues. For Huberman, Romero, and Wu ‘followers’ are the traditionally defined by the Twitter platform functions, but ‘friends’ are individuals who have directly mentioned each other in tweets multiple times, indicating a deeper connection and heightened engagement than just passively reading messages of the other.

14 Discussed by Katz and Lazarsfeld in 1955.
Overall Methodology
This study draws from the same dataset of tweets but employs different coding schemes and methods of analysis for each of the following chapters. Here, I discuss the formation of the dataset and the overall methods that apply across the board to the study as a whole. Later, in each chapter I dive deeper into the specifics that comprise the individual analysis.

Filtering tweets from the Firehose
While Twitter provides constant updates from users on ‘what’s going on’ it also provides a window into the past discussions happening across the globe. This snapshot of the past is what fuels this study. The data here was provided through Gnip, a Twitter owned company that provides custom historical and real-time access to the full Twitter “firehose”. Based on the sampling rules I created and provided to Gnip, roughly 2,000,000 tweets met the criteria set out. The search returned English language tweets from 02:00AM September 17, 2014-Midnight September 19, 2014 UTC.

The inclusion criteria for Gnip searchers are based on Boolean logic, which allows for basic yes or no inclusion. For the search, I included the most common hashtags associated with the referendum that were non-partisan: #indyref and #scottsdecide

---

15 From: 2014-09-17 01:00 UTC; To: 2014-09-20 00:59 UTC ((Scotland independence) OR (Scottish Independence) OR (Scotland UK) OR ("Northern Ireland" Scotland vote) OR (EU Scotland vote) OR (Oil Scotland vote) OR (Scotland "yes vote") OR (Scotland "no vote") OR ("#betterscotland" Scotland) OR (#nothanks Scotland) OR (#indyref) OR (#Scottsdecide) OR (#Scotland) OR (#the45) OR (#45percent)) (lang:en)

16 Gnip and Twitter operate on UTC or Coordinated Universal Time, which is closely related to Greenwich Mean Time. Conversions were made in the dataset to align the samples with BST or British Summer Time (+1 GMT/UTC), as this was necessary as the referendum fell while the isles were observing BST. Additionally the conversions based on time zones were made after Gnip provided the corpus of tweets. This resulted in 23 hours of tweets for September 17 rather than 24 based on the miscalculated time frames when the data was culled.
capture the discussion as a whole. I also included #yesvote, #45percent, and #the45 on the pro-independence side and #novote, #bettertogether, and #nothanks on the pro-union side in order to best include all related tweets. I included logic for tweets not hashtagged but that contained relevant combinations of key words such as “Scotland” and “independence”, “Scottish” and “Independence”, “Scotland” and “UK” to further skim the Twittersphere for the whole of the conversations surrounding this referendum.17

**Sampling from the Corpus**

The Boolean logic I provided Gnip was then applied to the Twitter firehose18—the entire body of public tweets sent through the Twitter platform. Applying the logic criteria discussed above and delineated in the footnote filtered the firehose down to only the most relevant tweets posted about a subject. There are difficulties finding every single tweet on a topic, however the criteria I used was intended to cast a wide filter over the firehose to understand how users utilize Twitter and how what they had to say on the topic of the Scottish referendum reflected in the polls and in identity construction. All tolled; Gnip returned over 2,000,000 tweets fitting the criteria provided.

---

17 Axel Bruns and Hallvard Moe discuss the shortcomings of relying only on hashtagged conversations to understand the full context of the Twitter conversation in their chapter in the book, *Twitter and Society* Chapter 2. My data here shows this is a valid concern as based on the data here between 50 and 100 tweets from each day within this sample contain zero hashtags.

18 This is an historic filtering system for all of Twitter’s history. Gnip’s system allows for targeted selection of Twitter history to be culled (for a price) by researchers, companies, and others. This system allows a researcher to look further back than 10 days, which is the standard for the Twitter API.
From the corpus of 2,000,000 tweets, the metadata for each of those tweets\textsuperscript{19} was sorted based on the pieces that lined up with this study, many of the variables utilized in the analysis at the heart of this paper were created based on the initial data and the public user profiles and tweets themselves. Random samples of 500 tweets from each day were pulled and coded by hand in order to address the research questions posed in this study.\textsuperscript{20} This paper looks specifically at the time directly after the referendum—the 24 hours of September 19—the day after the polls closed\textsuperscript{21} although the dataset includes tweets from September 17-19, 2014.

While the corpus includes over two million tweets and the random sample for this study includes 1500 total, each tweet was coded, verified, and cleaned by hand. Hand coding allows for additional analysis of complex messages. Nuance and sarcasm are important aspects of political discourse and they are challenges to mechanized sentiment analysis. Further, with the uniquely defined meaning of Yes and No in the case of a referendum, it is important to have contextual judgment in place to properly attribute meaning to the traditionally common words.

\textsuperscript{19} Special thanks go out to Brian Wingenroth, Research Software Architect for the National Consortium of the Study of Terrorism and Responses to Terrorism for converting the metadata to more usable files and being super nice about it.

\textsuperscript{20} Here with the abundance of ‘yes’ and ‘no’ traditional sentiment analysis would struggle identifying the real meaning behind the tweets themselves unlike how previous studies could focus on party or candidate names. In the Scottish case, we see yes and no indicating consent and descent, as they traditionally indicate in the English language, as well as support for the Yes campaign and No campaign respectively.

\textsuperscript{21} Of the tweets collected over 900,000 were sent on September 19\textsuperscript{th}. The other tweets that were pulled by Gnip are the focus of future work.
Coding by Hand

Hand coding requires a tradeoff, individually interpreting and coding tweets requires time, energy, and the undivided attention of an individual coder. Mechanized systems can process more tweets with less dedicated time but the caliber of analysis can omit critically important details when compared to hand coding analysis. In this study, I opted to hand code each of the tweets included in the analysis. This research involved not only the building of the dataset, like any study would require, but also the analysis of individual tweets multiple times to ensure uniformity and accuracy. If the intent behind the message in the tweet was unclear or did not translate well from the initial form of the data (i.e. emoticons and other symbols or acronyms used generally came through as jumbled strings of characters) I followed each link to individual tweets to see what the initial message included on Twitter itself. This type of thorough review of content and analysis would not be processed through a machine-coded dataset. Additionally, I cannot stress enough the potential for confusion surrounding the conflation of Yes indicating support for the Yes Campaign and a yes in the traditional sense of affirming a statement or opinion. The ability to take a nuanced approach to analyzing each message included in the study allowed for a different type of sentiment analysis, as well as quality control, even though it resulted in a smaller number of cases analyzed.

Location and Time Zones

An overarching issue with Twitter as a mechanism for research is the reliable reporting of the location of tweets or Twitter users. As research has found, “most tweets are not associated with a geographic location” (Wilken 161). While ‘location sparsity,’ a
concept coined by Cheng et al. (2010, 739), is challenging within the Twitter platform, there are a number of ways to include your location as a user or for a researcher to ascertain a location from a tweet each with its strengths and weaknesses. Unfortunately, the most precise location information provided through users checking the “add my location to my Tweets” option within Twitter is the least used. Another option is to use the self-reported location function within a user’s profile page on their Twitter profile. This, as Rowan Wilken discusses, is a “notoriously inaccurate measure of location” (157). On this tact, Wilken reports that a study conducted by Cheng et al. in 2010, “21% [of profiles] included a location as ‘granular as a city name’” (161). Companies that work closely with Twitter and sell or repurpose data employ workarounds that aggregate information to devise a location feature to include for each tweet they collect (161). For Gnip, the workaround for location is a user-generated, categorical, variable called “time zone.” This information is present for nearly 80% of the tweets in the database. Further, time zones are sliced further, delineating London from Edinburgh, and Dublin from the rest.

---


23 According to a study done by Reips and Garaizar in 2011, 0.6% of all tweets contain the geographic location which reports latitudinal and longitudinal coordinates for each tweet (Wilken 161). Within my full dataset for this study (pre-sampling), this held approximately true in that the geo-location was very rarely included with the tweets.

24 In order to be able to measure the location as accurately as possible, the samples of 500 tweets were drawn only from those with the time zone variable listed. This allows for the analysis of Edinburgh’s tweets as they compare to London’s, and the rest of the world, which in this case is extremely important.
**What Lies Ahead**

Based on the literature that surrounds the concepts of identity, we can recognize nuance from 140 character tweets that allows us to apply what we know. In the following sections, I will explore two questions: first—could we have predicted the results based on Twitter traffic?, and second, how are people talking about these events—particularly how are people interacting with those who disagree with them on the referendum question?

In the first chapter, I look specifically at the content of the tweets, breaking down tweets into discrete categories based on what they support. Drawing distinctions between Yes Vote support, No Vote support, comments on the successes of a democratic systems and vote, as well as general information—just to name a few—allows for us to analyze and understand what is going on in the public sphere of Twitter. This analysis allows us to see what is driving the discussion and debate, as well as the regional divisions in content and timing of tweets. In this chapter, the question asked is–could we have predicted the results based on the Twitter traffic? Through analyzing the content by region and by day, we find that there are still significant challenges in the predictive capabilities of Twitter, and that, at present, it serves more as a megaphone of public discourse rather than a traditionally conducted poll of public opinion or sentiment.

In the second chapter, we dive into the questions of national identity and the emotionality of the Twittersphere surrounding the Scottish referendum in two ways. First, I explore the overall emotionality of Tweets based on a unique coding scheme that analyzes how individuals talk about their ingroup and outgroup as it relates to how they view the referendum question. This analysis allows us to see that the day after the referendum was the most emotional overall as people come to terms with the realization
of the 45-55% split in the country. Additionally, we see that the most emotional tweets came from Scotland itself, which aligns with findings from the previous chapter that the conversation is driven by Scotland itself. To address the question of what these findings mean for national identity within Scotland, I analyze individual tweets to deconstruct the identity markers included in their content and to help make sense of what the individuals themselves are saying.

Overall this study addresses two questions: what can Twitter tell us about the role of election forecasting in this case, and what it can tell us about the negotiation and redefinition of national identity in the wake of a failed nationalistic question. In the following pages, I will clean what is possible from these questions and topics from the established literature and analysis of the Twitter traffic.
REFERENDUM PREDICTION AND RESULTS WITH TWITTER

Twitter, established in 2006, is a burgeoning source for social science research. From its advent, the debate about its role in society and politics erupted. Two camps formed around election prediction with a couple different methodologies in the fray. Some believe Twitter can be a gold mine for prediction and it should be harnessed to analyze elections. Others believe there are significant shortcomings and those who believe or direct research otherwise are misguided. Additionally, there are discussions on the feasibility, necessity, and functionality of sentiment analysis as it relates to Twitter and large-scale data analysis that comes along with it. In this section, I hope to contextualize the comparison between traditional polling and Twitter based analysis of public opinion, summarize the debate as it stands in the field and position this research in to speak to the current gaps in the field by adding additional methods and findings for the case of the 2014 Scottish referendum.

Traditional Election Prediction—To what Twitter-Based Research is Compared?

“Interest in pre-election polls has always been based on the expectation that they can provide accurate advance indications of election outcomes” (Crespi 4). This focus on predicting electoral outcomes has been around and continually refined over centuries. The field of polling, sampling, and election prediction has ranged from great successes in the modern day that are now so ubiquitous in politics, to failures in election prediction
like the 1936 *Literary Digest* presidential election poll. The questions driving polling in politics seem to stem from the question asked by Crossley after his 1936 poll failed, “‘[i]s it possible to sample public opinion sufficiently accurately to forecast an election, particularly a close one?’” (Hillygus 963). Here, we see these questions still plague the careers of pollsters and politicians. However, public opinion surveys and polling data have claimed a place in our understanding of politics so firmly that the field is continually working to hone the practice. Now, with technologies like Twitter where opinions are publically shared, the field sees new opportunities for innovative polling and sampling techniques based in social media. With these new opportunities being tested, it is important that we first understand the traditional polling tools to which Twitter-based methods will be compared.

**A Brief History of Polling (in American Politics)**

The 1824 American presidential campaign marked the first published ‘straw-poll’ which was a “[n]on-binding expression of a group’s stance on a pending decision” (Carpini 285). These polls included individual voters and mirrored a shift in newspaper focus where the press focused on the upcoming election from a citizen’s perspective for both civic and economic benefits. These polls continued with little change methodologically until the 1930s (Carpini 285-286). It was not until the 1936 presidential election where significant changes in survey methodology were attempted. The first issue of the Journal *Public Opinion Quarterly* was published in 1937 and it contained an essay comparing the flawed *Literary Digest* poll—one of non-representative random sampling based on phone and automobile records—and the findings of George
Gallup’s results which were based on a quota-controlled, more generalizable, survey (Hillygus 962).

The field of election prediction has changed drastically since the inaugural issue of *Public Opinion Quarterly*, but the progression forward has two mechanisms. First, as the field has advanced, reliance on polling data has increased and as reliance on polling data has increased, it has driven advances in the field (Hillygus 962, Carpini 295). This cyclical relationship truly came into its own in the 1960s and 1970s when “surveys became a fixture of political campaigns” (Hillygus 963). However, even though polling became a critical factor in political life, there were still significant challenges to reliable prediction. Professional pollsters still struggle with the questions at the core of election prediction. Between 1982 and 1984 presidential election, there was significant discussion surrounding the ‘art’ of polling that focused on the lack of scientific prediction capabilities at the time (Crespi 2). The biggest question when it comes to accurate polling is adequately representing ‘likely voters in the sample of the population polled—strategies for doing this have evolved over time, but still can unravel an otherwise solidly founded poll.

The other driving force behind advances in polling and survey-based measurement of public opinion stems from advances in technology. While telephone polling was conducted earlier than the 1936 *Literary Digest* poll, telephone penetration in the United States did not reach 90% until the late 1960s. As telephone rates increased, survey methods were refined, and in the 1980s telephone surveys were the preferred polling techniques (Carpini 295). This technology became dominant because most
Americans had phones in their homes; it became easier to use the technology to contact a representative sample of the population. However, as technology and survey methods progressed telephone surveys changed too. Moving from random digit dialing to interactive-voice-response and web-based platforms, we see the direct connections between technological advances and the revision of polling strategies to utilize new technology to measure public opinion more quickly, accurately, and cost-effectively (Carpini 295, Hillygus 963). Considering the exponential changes we see with the advent and growth of internet polling and data mining, it seems obvious that the next step in attempting to understand public opinion will rely on sites like Twitter. However, we are still refining survey methods for these new technologies.

**Challenges of Traditional Techniques and Sampling**

The largest flaw in the *Literary Digest* poll was the sample from which the survey responses were drawn. In today’s understanding of failed sampling strategies, this 1936 example serves as the example of flawed sampling and as a case study of how not to sample. However at the time, survey methodology and drawing representative samples was far from the established level we know today. A large strength of the 1936 *Literary Digest* polling case study is the built-in comparative success of Gallup’s in predicting the election, thus showing that if done properly, survey-based polling could have a place in political discourse (Carpini 290-291).

When it comes to refining survey and polling techniques, it seems the field of public opinion measurement is still working through the process for ‘old technologies’. Carpini chronicles advances in the field of survey methodology that now are
commonplace in political science based survey research but took serious consideration and time to establish as the standard practice. These advances included promotion of probability-based area sampling, standardized documentation, and data reports, as well as clarifying wordings of questions including the difference in responses based on open- and closed-ended questions (293-294). However, even with the great advances in sampling and survey construction that many have worked to establish over the years, there are still significant challenges in even determining what makes a successful or accurate poll. Both Hillygus and Berinsky discuss challenges still facing sampling and polling broadly. First, Berinsky contributes to the significant discussion on the role of non-response in a representative sample diving into both the full non-response where individuals opt-out of participating, and the more nuanced, question level omission of individual answers. He finds that there is still more for researchers to consider on these topics of missing results because, “[s]urvey non-response threatens the very foundation of the survey research enterprise” (345). Adding fuel to the fire of non-response in surveys, we are then faced with phenomena like ‘social desirability bias’ and the “Bradley effect” where respondents indicate preferences base on what is socially acceptable rather than how they actually will vote or act. This is particularly pronounced in questions of race in American politics through the “Bradley effect” where polls skew support for African American candidates higher than they should because respondents falsely report that they would be comfortable voting for an African American candidate (Hillygus 996).

On the other hand, Hillygus and Traugott add to the discussion of what is important for polling. When considering a successful poll, does it accurately predict the
winner of the election, or does it report the exact percentages of the vote that decides the election (Traugott 318-319)? Hillygus adds, “the very task of assessing accuracy raises questions about how best to measure it” (964). Further, there is a significant lack of transparency in modern polling methods because they are seen as industry secrets and are highly protected by companies that conduct the polls (Crespi 9, Hillygus 967). This shortage in information sharing or publication, contributes to a serious lack of literature on polling methodologies, but has lead to literature focused on “identifying and analyzing the determinants of voter decisions and voting behavior” (Crespi 9). While there is movement on many fronts, one development becoming more popular in election prediction comes from aggregating a number of different polls. This method of aggregation was first seen on a website—Real Clear Politics—but has become more and more common and is often seen to have been popularized by the work of Nate Silver (Traugott 327, Walker). These aggregations of polls do exactly what it sounds like they do, they take the average (or another more complicated measure) of multiple different sources over similar time frames asking the same question and merge the findings of different methodologies and samples together to a new aggregate estimation of public opinion (Traugott 327). These systems have found recent success and seem to be an indication of where part of the field of election prediction may be headed. However, others are turning the focus to new technologies and new modica of mining public opinion through publicly available sources like Twitter.

With the over a century of refinement of public opinion polling, we find that while there are still challenges to measuring accurately the support for a policy or to
forecast and election. However, this field of election prediction has seen reliable successes and these abilities to forecast the future has become integral to the political culture. It is these standards to which Twitter, or social media-based polling will be held.

**What Do We Know About Twitter Research and Elections?**

Analysts have toyed with message tracking and electoral campaigns on Twitter since at least the 2008 presidential campaign in the United States, to some success. Shamma, Kennedy, and Churchill and Tumasjan et al. both are able to find significant connection between Twitter traffic and the measure of public opinion. They set out match through social media analysis. Also, importantly, they prove that people are using Twitter for political discourse. However, the jury is still out. This is especially true for Gayo-Avello and others who argue the current field of research looking at Twitter for public opinion mining is overly simplistic and does not take into consideration the necessary nuances in sentiment analysis and election dynamics to produce adequate results. While there is debate on the topic, what is generally agreed up is that there is room for more work in this field.

**We Can Predict Elections on Twitter**

Shamma, Kennedy, and Churchill set out to understand the impact Twitter can have on media events leading up to an election, specifically a 2008 American presidential debate. They find that Twitter captures the users reactions to the events, or “Twittering debate-like events actually tracks take-away reactions” from the broadcast events (7). These evaluative reactions lend insight into the important aspects of political decision making for individuals or voters. These results are confirmed to a degree by the analysis
of Skemp surrounding the Massachusetts Democratic Senatorial special primary in 2009 with Martha Coakley (Skemp). In Shamma, Kennedy, and Churchill’s study focused on the 2008 Presidential debates in the United States, we see that during the event individuals took to Twitter to discuss or share topically-relevant reactions to the events playing out in the political sphere. Further on public opinion analysis, O’Connor et al.’s research links Twitter sentiment with public opinion through analyzing consumer confidence and approval ratings for President Obama in 2008-2009 compared to sentiment analysis of Twitter traffic over the same times scale. They conclude that simple Twitter-based sentiment analysis “replicates consumer confidence and presidential job approval polls” (7).

Ko et al. ran a similar study on the elections in South Korea in 2011, finding that Twitter traffic related to the election is an indicator of public attention being paid to the election in which people generally promote their own political leanings (142-143). While this is something to note, it is also of critical importance for politics, where we expect that the support of an idea or set of policies generally equate to a vote in favor of those supporting that idea.

Here, we see the direct connections with the events in Scotland. As the campaigns for independence ramped up, individuals wrestled with the ideas surrounding independence, and with Twitter is a platform for observation of these types of reactions to politics. Based on the two-year campaign process, and the attention paid to the

---

25 Skemp’s study looked at the Senatorial primary that selected Martha Coakley as the Democratic candidate to face Scott Brown (R) in the special election to fill Ted Kennedy’s seat in the Senate.
question of independence for Scotland in the media and everyday life in Scotland, we can expect to see a great deal of traffic surrounding the referendum and more importantly, traffic that reflects the levels of support for the Yes and No votes in the campaigns.

When it comes to the political discussions on Twitter, the inherent questions involve who is talking and whether it matches the current political landscape, importantly the outcomes of elections. Tumasjan et al. find that the Twitter traffic surrounding the 2009 German national election, “plausibly reflects the offline political landscape” (178). This landmark study in the direct connection between Twitter mentions of candidates or parties and election predictions show that simple focus on candidate or party name mention counts and sentiment analysis reflect the end vote counts. They additionally find that, “while Twitter is used as a forum for political deliberation, this forum is dominated by a small number of heavy users” (Tumasjan et al. 181). The differences in types of users indicate that there are at least two distinct types of Twitter user—those who are active, and those who are more passive. Paring this with the robust results found in the study still support the findings that Twitter traffic on political events represent the results of ballots. Skoric et al. found similar, but more cautioned, successes in their exploration of the 2011 General Election in Singapore. Skoric et al. found that mentions of political candidates or parties on Twitter allowed for prediction of the percentages of the vote won by different parties in the General Election in 2011 (2583). They find that this prediction mechanism was better than chance in predicting these vote shares, but did not meet the same levels of confidence as Tumsajan et al.’s study. However, Skoric et al. find the relationship in Singapore is stronger for national trends rather than constituency-level
trends, which they suggest means Twitter is better suited for “macro-level assessments of political sentiment” rather than individual elections in a particular district (2589). They also suggest considering the restrictions on media and free speech in Singapore, considering other factors like the election competitiveness, restrictions on free speech, or how stable a democracy the country is to indicate how well Twitter will represent the results before they are finalized by electoral commissions. Concluding, they find, “Twitter is a more likely to represent a valid indicator of national (macro) political sentiment” in stable, democracies with freedom of the press (2590).

With the Scottish case, we find intersections with Skoric’s analysis and presumptions. The question of Scottish independence was a highly contentious, inherently a national question, that impacted more than just Scotland. The press and political speech have significantly fewer restrictions in Scotland and the United Kingdom than does Singapore. Based on Skoric et al.’s findings we expect the results of Twitter analysis to mirror the results of the vote very closely.

**What we Know about the Challenges to Prediction on Twitter**

The Tumasjan et al. study spurred great debate and the study’s strong findings were called into question. Across the board, Jungherr, Jürgens, and Schoen criticized it in an article responding to the study’s methods and findings. The response points out some of the flaws in the research calling for less arbitrary choices in timeframe and other aspects of research design. These naysayers are not alone in the dialogue about whether prediction is possible and what methods should be deployed to test the predictability of Twitter as the field is in its infancy.
Contributing significantly to the debate on the limits of predicting elections on Twitter is Daniel Gayo-Avello. In his 2011 paper co-authored with Panagiotis Takis Metaxas and Eni Mustafaraj, they argue that prediction on Twitter is possible, and that it is a powerful resource but that no one has properly harnessed. In *The Limits of Electoral Predictions using Twitter* they find the highly cited studies of Tumasjan et al. and O’Connor et al. function no better than chance in predicting the outcomes of elections because of the dearth of information surrounding demographics and other traditional control variables that are accessible to professional pollsters (4). Further, Gayo-Avello discusses the challenges facing research that relies on social media—generalizability and representation of the full population accurately. He states, “just because you have a big N doesn’t mean that it’s representative or generalizable” (Gayo-Avello 2012 92). Gayo-Avello takes particular issue with selection bias, lack of reported demographic information, the potential for disconnect between tweeting something and casting a vote for it, and most importantly, the lack of standard models and methods to ask questions of the data (Gayo-Avello 2010). While Gayo-Avello thinks the predictive power of Twitter analysis in elections is overstated in previous studies, he presents caveats to consider and suggestions for future research that would pass ‘muster’ for him. First, he calls for predictions. He correctly notes that most studies ‘predict from the rearview mirror’ or, they see what the Twittersphere tells them once the results of the election are known and then decompress the data to see what can be found. Rather, Gayo-Avello advocates for somehow making and publicizing a model of analysis ahead of the vote and comparing the results to reality once a prediction was made. In this vein, he argues for a better
understanding of incumbency and the impact it can have on any election to be included in any election-based modeling, particularly one being analyzed through social media (Gayo-Avello 2010, 3). These two points would be interesting in the case of Scotland’s referendum as the vote was very close leading up to September 18th and the Yes campaign was a ‘louder’ campaign on the streets, but the news media tended to support the No vote. Without much confidence in traditional projections leading up to the announcing of the results, knowing if Twitter followed public sentiment or media accounts could provide great insights into the utility of prediction with Twitter.

Next Gayo-Avello, advocates for seeking out and using only reputable data, from social media and all internet based research. Spammers and imposters or automated systems set up to stream information are a real threat to research based on counts or mentions online.26 Here, we see a demonstration of when a tweet does not equate to a vote. If spammers and fake accounts come into play there are issues related to mention tallies employed by this study and others undertaking these studies. We see similar challenges when faced by any sort of silent majority (Gayo-Avello 2010, 3). The silent majority, a phenomenon where the majority is the ‘quieter’ campaign where the majority of voters support one candidate or issue, but the minority supporters have a message that carries the airwaves, is an easier case to make, and/or has more access to media platforms.

26 See From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search by Panagiotis Takis Metaxas and EniMustafaraj in Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.
On Twitter, we can see how silent majorities could pose a problem if a study is based on Twitter traffic alone. Mustafaraj et al. unpack this question further where they distinguish between two types of Twitter users, the “vocal minority (users who tweet very often) and the silent majority (users who tweeted only once)” (1). While their study focused on spammers and fake accounts making up most of the vocal majority in a Massachusetts Senatorial campaign in 2010, it is still a distinction between how individuals engage on Twitter. In considering the silent majority, however, is a tricky topic when it comes to observation-based public opinion measures focused on Twitter or other polling techniques.

There is a fundamental difference between a tweet and vote and it is critical to moving Twitter based election literature forward. Tweets have low barriers to entry and low impact while votes have higher barriers and higher impact. However, these barriers apply to traditional polling too. Traditional polling relies on measures of public opinion of likely voters to measure how well a party or candidate might do in an election. Here, we employ similar logic, to predict what will happen in the ballot box based on public commentary. Here it is based on Twitter rather than other polling measures.

When considering the rationale behind participation in a silent majority and vocal minority, many circumstances come to mind ranging from Yes campaign fighting to reestablish a new Scottish identity rather than maintaining the status quo. These competing interests bring both the spirit of campaigning for change and identity politics into the mix for the Yes campaign. However, also, present is the phenomenon of ‘wishful thinking’ where “people usually expect their favorite candidate to win” (Krizan, Miller,
and Johar 140). In their study based on the 2008 presidential election in the United States, Krizan, Miller, and Johar found that “people’s preferences shaped their expectations” and that this phenomenon is further enhanced because an individual’s disappointment was shaped by their expectations regarding the election outcome (140). This paried with the findings from Bollen, Pepe, and Mao that indicate that “[r]egardless of their contentment and intended use, tweets often convey pertinent information about their author’s mood status” (1). This reinforcing cycle of preference for a side, firmly believing it will succeed, and more disappointment when those expectations are not realized, are critically important to understanding parts of the Scottish referendum. And if a Twitter user is tweeting about these things, their mood will be reflected in their messages. Especially when we consider an additional finding of the Krizan, Miller, and Johar study: that to a modest level, “rousing overoptimism (sic) among the candidates and supporters can exacerbate the negative reactions should the candidate [or campaign] lose” (146). In other words, the increased confidence put forward during the lead up to the referendum by the campaign organizations were setting themselves up for increased disappointment. Considering that, we already had a more active and vocal Yes campaign on Twitter, and if we consider the inflated levels of disappointment from the ‘wishful thinking’ effect, we see even more emotion and reason to take to Twitter in the aftermath of the two-year build up to September 18th.

This study allows for the unique analysis in the immediate lead up to the Scottish referendum and the time surrounding the announcement of results that individuals wrestled with the results of this decisive vote. By using the unique case of a referendum
with a strong silent majority, we see that Twitter mirrors the campaigns, not the results of the referendum. Through hand coding—something currently missing from the literature—we see that the nuance is critical to understanding the way the tweets represent the public opinion of the Twittersphere. And with such a competitive and national vote with a free press, this study directly challenges Ko et al.’s inference that Twitter will better track public sentiment when it comes to these conditions.

**What are the Tradeoffs between Polling and Twitter?**

When looking to understand what is going on in the run up to an election, traditional polls focus on building a representative sample of the ‘likely voter’. This was where 1936 *Literary Digest* poll failed and traditional polling still works hard to capture a representative sample of the population that will vote on Election Day in order to generalize their findings to the broader electorate. Twitter based research focuses on easier to collect and more timely publicly available information from Twitter users. The struggle with this stems from the demographics of Twitter users, as the population of Twitter users does not correspond with a representative sample of the electorate. From this, Twitter based research runs into issues like the silent majorities and vocal minorities as discussed by Mustafaraj et al..

While Twitter based research currently struggles with generalizability past the population of active Twitter users, there are a number of benefits that this research can provide with which traditional polling still struggles. First, analyzing public Tweets can be done in real time and it can provide spectacular understanding of real time changes in the political realm. This can provide robust analysis of shifts in public discourse as they
happen and allow for better understanding of when and how changes in sentiment occur. Second, as Twitter grows and becomes more and more accessible, the population using Twitter will likely better represent society more broadly, thus allowing better generalizability and representation of the broader population of likely voters. Pursuing this research while Twitter is still young is far from futile as it allows for developments methodologically for future use, just as we have seen in traditional polling and sampling mechanisms.

**Research Design**

Based on the challenges put forward by previous studies discussed in the literature review, this study was designed around a straightforward consideration that we could learn a great deal from the analysis of Twitter in the case of the Scottish referendum. The Scottish case presents a timely and contentious example to explore the ability of Twitter analysis to predict the results of a referendum question put to a vote in a stable democracy with higher than average Twitter use. The results gleaned from the use of Twitter in this political context could allow for insight into the participation in both Scotland, the UK and the world broadly, the types of messages that dominated the Twittersphere, and how this compares to the final tally of the referendum.

In order to harness the content of messages from Twitter, additional work and coding is required. Here, I discuss the coding mechanisms to clarify process and contextualize results. The coding schemes utilized here are unique to this paper, as all

---

27 As compared to the rest of the world, Beevolve.com ranks the United Kingdom as #2, behind the United States in Twitter usage.
Coding and sentiment analysis was done by hand rather than through mechanized sentiment analysis to provide insight into a different method of analysis to studies of this kind.  

**Coding Rules**

After sampling the data from the corpus of 2,000,000 tweets provide by Gnip, a variable was created to help digest the time zone classifications. The data provided from Gnip delineates between Edinburgh, London, and other global cities. In order to understand the tweets originating from Scotland and the UK, I combined all Edinburgh time zones into a Scotland category. All the London tweets were combined into the UK category, and all other tweets were categorized as “World”. This re-categorization of data was done in order to analyze the difference in sentiment and in volume of the tweets surrounding the referendum inside Scotland, the United Kingdom, and the world as will be displayed in the data below.

<table>
<thead>
<tr>
<th>Region</th>
<th>9/17/14</th>
<th>9/18/14</th>
<th>9/19/14</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotland</td>
<td>125</td>
<td>99</td>
<td>46</td>
<td>270</td>
</tr>
<tr>
<td>UK</td>
<td>141</td>
<td>129</td>
<td>131</td>
<td>401</td>
</tr>
<tr>
<td>World</td>
<td>234</td>
<td>272</td>
<td>323</td>
<td>829</td>
</tr>
</tbody>
</table>

---

28 Hand coding, as discussed in the introduction allows for additional analysis of complex messages. Nuance and sarcasm are important aspects of political discourse and they are challenges to mechanized sentiment analysis.

29 There were no other UK or Scottish cities included in the time zone data. Dublin was broken out as a separate entity—it was added to the World category along with all other cities of the Commonwealth.
Next, the tweets were classified into a broad range of categories based on the content of the message. If the intent behind the tweet was unclear and a hyperlink or photo was included in the tweet, those were considered when categorizing the tweet. Categories included, yes vote and no vote,\textsuperscript{30} as well as tweets asking people to accept the results peacefully, and tweets looking toward future change after the referendum. Tweets discussing other priorities either on voting day or as a critical rationale for the user’s support of the yes or no campaign, were coded separately into an ‘other priorities’ category. And tweets focused on the democratic process or nature of the vote, (including the role of 16 and 17 year olds for the first time, commentary on the nature or existence of a referendum of this nature, and the high turnout) were also categorized separately, under the ‘General Democracy’ header. Additional categories were created for tweets sharing information, results, or general commentary, humor, and for those that were unclear in their intent.\textsuperscript{31} Table 2 shows the distribution of the categories across the categories and regions.

\textsuperscript{30} An ‘undecided’ category is included for tweets sent on 9/17/14 and 9/18/14. These tweets specifically indicated that the Twitter user had not made up their mind about how to vote or which campaign they support in the referendum. This is a different categorization from “unclear” where there was intent voiced, but it was not clear which campaign was had the support of the Twitter user.

\textsuperscript{31} If a tweet was not clear in intent, any links, photographs, or other included content to help determine intent.
Based on Table 1, above, we see significantly more people are tweeting in the world than in the UK or Scotland specifically, as expected based on the world population and distributions of Twitter users. But comparing that to Table 2’s data, we see a clear focus of Twitter traffic on information sharing. Across the board, the heavy lifting in these categories focus on the sharing of information or results as they are announced on September 19th. Here, we find people answering the ‘what is happening’ question that Twitter asks with general information and the sharing of results. These results indicate direct engagement with the process as it is happening. This type of engagement is mirrored in the nearly 65% of the tweets from Scotland itself that reference the democratic process. The general democracy category included reports of the level of turnout for the referendum, which far surpassed turnout for a traditional parliamentary

---

32 This includes Tweets coded, Yes Vote, No Vote, and General Democracy from the Scotland time zone.
election. This level of engagement is impressive for a country where voter turnout is non-compulsory.

Once the coding was complete, the data was analyzed over the timeline of the election proceedings. That is to say, this study pays particular attention to when the polls opened and closed,\textsuperscript{33} when results began rolling in from councils, and when Alex Salmond of the SNP conceded the race.\textsuperscript{34}

**Discussion and Results**

**When are people tweeting and what can it tell us?**

When considering the data we first want to know when people were tweeting over the course of the days leading up to the vote and how those trends relate to the key turning points in the timeline. Looking at Figure 1, below, we see expected ebbs and flows of total Twitter traffic. Looking at the trend over time, we see the low counts in the early morning hours on the 17\textsuperscript{th} and 18\textsuperscript{th}, with a sharp spike at 7am when polls open in Scotland and others are beginning their day in the morning (or at least beginning their day on Twitter). Further, there is a peak again on September 18\textsuperscript{th} between 8pm and 10pm leading to the polls closing. Notably on the night of the 18th, we do not see the drop in activity in the early hours of the morning one would expect and that was seen on the previous night as the conversation continues in anticipation of the announcement of the results. In fact, the data peaks early in the morning of September 19 between 4am and 6am when result information started spreading, leading up to Salmond conceding around

\textsuperscript{33} Polls opened at 7am and closed at 10pm across Scotland on September 18\textsuperscript{th} 2014 (What Time Do the Scottish Referendum Polls Open?).

\textsuperscript{34} Alex Salmond conceded the referendum at 6:18 am on September 19\textsuperscript{th} 2014 (Holehouse).
6:20am on the 19th. Thus, indicating that Scots stayed up awaiting results—and this is mirrored in individual Twitter user posts as well. From about 7am onward there is a significant drop-off of Twitter traffic about Scotland and the referendum. While still present, it trails off as the results settled and people move onto other topics.

![Tweets Over Time](image)

**Figure 1: Count of all tweets over time.**

If we break down the data by region we see the trends from the aggregate data closely mirrored in each region studied here. Figure 2 displays the same data as Figure 1 but breaks it down by region and time. Here we see generally the same peaks and valleys over time. Interestingly this suggests the Twittersphere, while global, is responding in real time in Scotland and the United Kingdom, as displayed when comparing Figures 1
and 2 with Figure 3, which breaks out only the Twitter traffic in Scotland. If individuals were fully engaged globally with this referendum we would expect to see steady levels of world chatter with peaks and valleys in the Scottish and British lines following the rhythm of the day that we see with all the data. But with the mirrored trends, we can conclude that for this time-period, events in Scotland and the information, originating in local time, surrounding them drove the timing of the Twitter traffic on the referendum.

Figure 2: Tweets over time broken down by region.
Of additional note, we actually see a decrease in the overall number of tweets sent from inside Scotland on September 19th and an increase in the tweets sent from the rest of the world, while the UK says relatively constant over the three-day span, as seen in Figure 4. Interestingly here, one would expect to see this trail off Scottish and UK tweets because the campaign-based social media profiles and information sharing on voting logistics would be winding down. However, based on the sum totals of the entire corpus of the data made available for this study, more tweets were sent on September 19th than any other day in the study, showing that the conversation once the results were announced shifted from Scotland and the UK to include more of the broad Twittersphere.35

35 September 17 yielded 421,295 tweets over 22 hours, September 18 yielded 710,690 tweets over 24 hours, and September 19 yielded 915,187 tweets over 24 hours.
What are people tweeting and what can it tell us?

Turning from looking at when people are sending messages in relation to other referendum-related events to what people are saying; similar trends are confirmed from above. First, Twitter is Twitter. There is a lot of general information sharing going on online regarding the referendum. Looking at Figure 5, these trends present themselves in the data from the content of each day’s tweets. Here, we see significant numbers of tweets from each day sharing information about the referendum. On the 17th and 18th some of these messages were information about where and when to vote, where to find more information about the referendum itself and the positives and negatives associated with the two choices. While on the 19th, this category was dominated by sharing of the results as different councils announced their votes and people shared them widely in the early morning hours but shifted quickly once the results were announced.
Across the board, the Yes Vote category has the next highest count. Here, we see the impact of the campaign and the reach of the message of independence; this was true across all three regions, as shown in Figure 5. Of note, however, is the comparison between the ‘General Democracy’ counts and the ‘Yes Vote’ and ‘No Vote’ category counts. These comparisons, particularly on September 19th are clear. The Yes campaign supporters begin to trail off, and while the information sharing continues, tweets about the process, turnout, and inclusion of 16 and 17-year-old Scottish voters begin to spring up prominently. These tweets also include the messages of support for each campaign having ‘fought gallantly’ as the referendum campaigns represented a great showing during the long push toward the September 2014 vote that stared years prior.

Of additional note when breaking down the divisions of content distribution over the three regions, the Yes-No divide in Scotland tweets is always very strong as expected when comparing Tweets originating in Scotland and those coming from the UK. However, digging into the distributions, the gap between tweets supporting the No campaign and the Yes campaign stemming from the United Kingdom dwindles with each day, resulting in more Yes support on September 19th than No. This is particularly noteworthy, considering the overall distribution of critical tweet content over time, as displayed in Figure 6.
Figure 5: Tweets broken down by day, region, and content of the message.
Figure 6: Count over time of important categories of tweets over time.

In Figure 6, we see again the count of all of the tweets over the time considered for this study, however this time tweets are broken down by content analysis to include the categories critically important to this chapter. As seen previously, the lion’s share of the tweets share information or results. The peaks of that content mirror the same distributions of the overall tweet counts with peaks in the late evening before the referendum and from polls closing on September 18th through to Salmond’s concesion.

---

36 Only tweets coded as General Democracy, Information/Results, Yes Vote, No Vote, or Unclear in content are included in Figure 6 in order to avoid clutter in the graph.
speech on the morning of September 19th. The exception here is visible at the time the polls open, at 7am on September 18th. Here, the Yes and No votes rise together. While Yes still out paces No at the time polls open, just after the results are finalized, between 5am and 7am on September 19th, the No Vote count surpasses the Yes Vote. Yet, even with the spike in support for the No Vote on Twitter, immediately when results became final, September 19th was the only day where UK support for the Yes vote out paced UK support for the No vote.

**What do the tweets say? Yes or No?**

Looking into the meat of the discussions that directly represent support for the two sides of the campaign. The Yes campaign out Tweeted the No campaign. As displayed in Figure 7, the division is nearly 70% Yes and 30% No. This question can be analyzed in different ways; if we look at the results from the day before the election (September 17), this distinction skews further in the Yes campaign’s favor but over the next two days reflects more of the No sentiment as seen in the results (Figure 8). While the Twitter traffic never matched the poll results, the increases in support for the No Vote brought the distribution closest to what the polls

---

37 Undecided was not an included as an option for coding for September 19th as the referendum had already taken place and results were announced early that morning.
Figure 7: Total proportions of Yes/No and undecided tweets over the three-day period.

Figure 8: Total proportion of Yes/No and undecided tweets broken down by day.
reflected that day. Interestingly the results that came the closest to what the polls reflected were the United Kingdom’s total numbers from all three days. Figure 9 presents the counts of all Yes Vote, No Vote, and undecided tweets from all the days. Here, we see that tweets sent from Scotland are very supportive of the Yes campaign, while the United Kingdom’s Twitter traffic better mirrors the lead up to the referendum in which the electorate was highly split but still too close to call, and the world’s tweets are caught between the two.

![Pie charts showing the proportion of Yes/No and undecided tweets by region.](image)

**Figure 9:** Total proportion of Yes/No and undecided tweets broken down by region.

**Discussion**

As seen in the review of the literature and where election prediction work stands presently, these topics are not simple mechanisms, to deploy on a whim, to cleanly and
neatly analyze election results. In this study, a significant benefit was the ability to hand code the tweets to harness the true intent behind the messages and display it alongside the others. However, this work limited the size of the samples. While the analysis here is convincing that Twitter could not have predicted anything near the 45%-55% results seen in Scotland, greater strides could be made toward mechanically sorting out the informational or general Twitter noise from the more personal tweets that may equate directly to a vote. However, it should be noted that even traditional polling methods surrounding the Scottish referendum had difficulty projecting the result. It was seen as a toss-up, until the results were finalized.

Finding a way to identify tweets that more directly equate to a vote is critically important. Perhaps this could be done through culling only tweets sent by the second type of users identified by Mustafaraj et al. as those who tweet only once or twice on a subject rather than the more active users. Scrubbing the data to find more individualized messages that reflect the voting intention of an individual beforehand coding could harness two powerful analysis mechanisms.

This study provides a challenge to the assumptions regarding the role of free societies and the contention of an election presented by Ko et al.. The United Kingdom, a stable democracy with free media and political discourse, held a national vote in Scotland and Twitter did not successfully predict the results even in the immediate run-up to the ballot count. This raises the question of a potential disconnect between public discourse, public sentiment, and referendums. Based on the analysis of the Twitter traffic, we see a clear case of a silent majority at work in Scotland. But, how do we determine what leads
to this when it comes to something as charged and important as a referendum on independence? Answers to these questions are likely to be multifaceted and complex, thus they run the risk of being lost in a sea of 140-character messages.

Importantly, this analysis confirms that people are talking about politics, particularly about the referendum in Scotland, on Twitter. But in this case, we saw the reflection of public opinion and the broader conversation in the Twitter conversations, not the prediction of election results. It is not until after results are announced that the support for the No vote is strongly represented on Twitter throughout the referendum campaign, the Yes vote had the easier job selling their side. Through this, it is made clear that Twitter mirrors the discussions, debates, and conversations going on in the political sphere, inclusive of the role a silent majority. Our mechanisms are not designed to notice what is not tweeted.

**Conclusions**

This analysis illuminates three key areas of the current Twitter based election prediction literature. First, we see that the discussion surrounding the Scottish referendum globally was driven very directly by content produced on the ground in the United Kingdom (or at least in that timezone). This was displayed through the time-based analysis of content in which the tweets not from Scotland and the United Kingdom followed the same time-based trends in peaks of discussion-indicating the local influence Twitter has on the global conversation surrounding localized events.

Second, this analysis confirms the existence of a silent majority and vocal minority within the Twitter traffic surrounding the Scottish referendum. With the
significant divergence between the results of Twitter traffic shown here and the results of the referendum, we clearly understand that this phenomenon is present in this case. We also see grounds for the compounding impact of the wishfulthinking constructs argued by Kirzan et al. As shown by the Yes vote being more vocal and more dissapponited and shocked by the results, leading to the longer lived conversations in the Yes catgory of content after the results were finalized and announced.

Third, and most importantly, we see through this analysis that for Twitter-based research, when compared to the established field of election forecasting, there is still work to be done. The Twitter- and social media-based methodologies are still in the development phase, for concrete forecastings, and what we can see through this study is that Twitter is a place for discussions, including political ones, as seen in much of the literature. However, Twitter does a better job mirroring the public discourse and conversation being had among Twitter users, or the relatively young (and slightly more female) population, rather than the public opinion that is targeted in traditional polling mechanisms. However, in this vein, this study calls into question assumptions put forward in Scoric et al.’s work on the Singaporean elections where they posited that in a more national electoral question and in a nation with a freer and fairer press and public space than what is seen in Singapore, Twitter-based research would align more closely with the election results (Skoric et al. 2590). While a referendum may function differently in Twitter discussions in the political realm, it seems that the national question in Scotland and the free and fair election and press did not allow for the type of analysis they predicted.
Through this work we see the opportunity to continue the discussion of the role of Twitter in political forecasting. While there is still a long way to go before social media-based analysis has the same grounding that traditional polling sees in politics, the continued push to explore and refine is required of any new field. Understanding Twitter’s place in the analytical world is important to understanding society and the impact of individual people sharing their views and what we may be able to learn from this new, live streaming of 140 characters.
HOW PEOPLE WRESTLE WITH NEW IDENTITY CONSTRUCTS ON TWITTER

On September 18, 2014, democracy in the United Kingdom and nationalism in Scotland were put to the test; Scotland, as a nation, was directly asked if it wanted to become independent. For some, this was a question 300 years in the making, for others, it took a back seat to other basic considerations like an agreed-upon currency and status within the European Union. On September 19th, the results were announced. Through Twitter, the world saw immediate reactions to the outcomes. Using new media and our understandings of how they function, we are able to gain fresh, real-time insights into ‘what is happening’—the question at the heart of the Twitter platform.

Twitter allowed us to watch, at the individual level, as the events unfolded in Scotland. Individuals discussed the upcoming vote and shared last minute pleas attempted to convert undecided voters to their side on September 17th. On September 18th, Twitter users shared information about polling locations and how to get there if someone needed a lift and shared their own rationale for their vote. On the 19th, they grappled with this new division in Scottish identity between Yes and No votes on Twitter, in 140 characters or less. Taking these messages from Twitter and the emotions including pride, anticipation, anger, and disappointment, contained in the Twitter traffic, on September 17th-19th, we see how individuals interact with the Twitter platform when it comes to political referenda, how people react to the news, and particularly, we see the divisions
between those who voted yes, those who voted no, and those who stayed home from the polls, within the generally unified ‘Scottish’ identity.

Building off the analysis conducted in the previous chapter, now we turn to the questions relating to identity construction and reconstruction as it occurs during the lead up to the referendum and during the announcement of the results on September 19th. This analysis allows us to address the identity questions that are often associated with national identity, for the Scottish case. “‘Who are you?’ ‘Who are we?’ and ‘Who are they’….their answers are identities—always assertions, always contingent, always negotiable, but also always consequential” (Tilly, 209). Charles Tilly, a stalwart in the realm of identity politics and nationalism, delineates the key points regarding identity succinctly. These three questions include the individual, group, and other, and with the progression used, he alludes to the interconnectedness and the important distinctions between the different levels where identity becomes salient. In the case of the Scottish referendum, these points are all at play, and the answers are of great consequence for current Scotland and the status of national identity.

Identity Broadly

According to Tilly, “[i]dentity belongs to that potent set of social arrangements in which people construct shared stories about who they are, how they are connected, and what has happened to them” (209). These range in scope and salience, as we will explore below, but they share four components: boundaries between us and them, set norms within and across the boundary, and history or stories of inter and intra boundary experiences (209). As for the Scottish case, changes realized through the referendum in
some of these four categories present challenges within the identity group come to the surface. First, there seems to be a schism between the ‘us’ and the ‘them’ within the established Scottish identity. When considering the changes in the United Kingdom consider that there is now not only the Scottish/English/Welsh/Northern Irish divide, now, as we will see, there is a highly relevant Pro-independence Scottish and Anti-independence divide which this referendum brought squarely into the public sphere.
With this new divide in an identity group, it adds to the types of relations between groups and types or relations across groups in new and difficult ways. However, what is rightfully missing from this set of requirements for identity laid out by Tilly is any piece that withholds or categorically rejects changes in the arrangements set forward; in other words, identities and relationships between identity groups are malleable (8).

Consider Anthony Smith’s addition to this sentiment in his works on identity by stating, “human beings have multiple collective identities, whose scope and intensity will vary with time and place” (Smith 175). Further, for Tilly, “every political process includes assertions of identity, including definitions of relevant us-them boundaries” (8). This assertion of Tilly will come back to us through the analysis of the Twitter results surrounding the results of the Scottish referendum, as it takes into account the individual’s ranking and rationale for that ranking of identities and their importance vis-a-vis political decisions. Through the progression of identity, we understand how an event, like the referendum in Scotland, can increase the significance of formerly less important distinctions within the Scottish identity. When we delve deeper into categories of identity, we can understand further dynamics and interplays.
Inherent in the lack of rigid structure is an opening for additional identities to take hold and grow in importance depending on topics and situations. For us, this flexibility and infinite number of identities becomes particularly important. Identity is an individual phenomenon at its core while there are distinctions within groups, there are also distinctions and overlaps within one individual, leading to the ability to float between identities at different times. This can be a fluid transition between identities, or it can be an excruciating process and the differences in these processes can be reflected in how the changes or feelings are discussed. Now, with Twitter, people are sharing these transitions and how they see themselves and the world in a public forum. Observers can see how individual align themselves with certain ingroups, and how they interact with or attack other outgroups in real time and in comparison to what is currently going on in society—like an independence referendum.

Identity Categories at Issue

With identity conflicts, we are forced to consider different levels of identity in a hierarchical structure based on situational relevance and individual passion for a certain individual identity. We can see more tangibly how one could prioritize identities, or fundamental parts of themselves. In the case of Scotland, we are looking not at different identities, per say, but different prioritizations of facets of the “Scottish” identity and the percentages of the populations that were counted under these new banners through the referendum. Theorists have attempted to categorize these levels of identity. Of particular interest in this paper is the distinction drawn between social identity and national identity (and all the baggage that goes along with national identity).
Social Identity

Social identity, as described by Karina Korostelina includes “a feeling of belonging to a social group, as a strong connection with social category and as an important part of our mind that affects our social perceptions and behavior” (15).

Additionally for Korostelina, social identity provides five critical psychological assets to an individual: self-esteem, social status, safety, protection and support, and status in an ingroup (185). Within this structure, she allows for expansion, evolution, and socially based flexibility for constructs of social identity, which allows for shifts in the prioritization of one group identity over another depending on situational changes (17).

Social Identity, for Korostelina fulfills two critical purposes for individuals, first, identity “provides protection and confidence” through the psychological functions she delineates, and second, identity places “a person into a system of social relationships” (18-19). With these two functions, we see the importance of social identity but also its inherent context-based flexibility. Through the protection and confidence and the inclusion into social relationships, we can easily see how social identities can encompass a group’s full being if threatened. These feelings result in a sense of “we-ness” within the identity group (Korostelina 19).

National Identity

While social identity provides insights for us into the nature of identity, national identity takes additional forms that are critically important to our understanding of certain phenomena. Anthony Smith traces the beginning of the concept of nationalism, as we know it today, to Friedrich Meinecke’s 1908 discussion of Kulturnation, a “largely passive cultural community” (Smith 8). However, in current geo-politics and theory,
national identity ranges from passive culture to active, sometimes violent, expansion (or protection) of culture or cultural dominance.

Through these colorful descriptions, we see connections between national and social identity, sometimes leaving little to distinguish one from the other. However, within national identity, there is an association with or drive for an overarching governing body recognized by the international system as somewhat autonomous (Cederman et al. 2011, Wimmer 2013). Korostelina delineates notable distinctions for “the acquisition of national identity” which includes distinctions that establish shared national customs or norms, a shared definition of the self, and other elements that distinguish the unique nature of the new nation (182).

It should be noted here, the distinctions between social and national identity are flexible, just as distinctions between all identities are. In the case of Scotland, there are individuals who link Scottish-ness with a need to be independent from the British Crown, while others link it to what type of whisky they drink. What is critical to note is that there is a distinction between the two and it generally surrounds the need to govern a nation representative of its people as an individual unit or sovereign state. However, both national and social identities are flexible in construction and are constructed by individuals with in the social group or nation. These identities are manifested individually but lead individuals into a broader dialogue within groups established by this shared identity. This individual negotiation of identity within a shared group identity, leads to the potential for disconnects between individuals over how a shared identity should be
expressed—and in the case of Scotland, an expression of how individuals prioritize their identity was through how their vote was cast.

**Nationalism as a Manifestation of National Identity**

When considering national identity, there are different embodiments of what it means to be a nation through history. In the Scottish case, this ranges from warring clans, to the independent Kingdom of Scotland, and the post-Act of Unification Scotland seen today as a part of the United Kingdom. We can understand this in the abstract. However it quickly devolves from an abstract concept to real, tangible, fuel for division among peoples. Walker Connor in his article that attempts to separate ethnicity from nationalism and participation in a nation from participation in a state, he argues that nations are, a self-selecting process can include many, and just as easily exclude others. In order to ‘acquire’ a national identity, one must accept the defining facets of said national identity. Gellner argues that cultural foundations for a nation can be constructed from the past, thin air, or a combination of the two (56). The choices in how to construct or revive cultures are the decisions draw new distinctions between groups to allow for inclusion, or to actively exclude those who are unwanted.

While the question from the referendum ballot, “Should Scotland be an independent country?” is theoretically straightforward, when presented to voters the externalities of the potential change from union to independence are quickly realized (Scottish Independence Referendum - GOV.UK.). A more accurate question that voters were asked to answer would look like the following, ‘do you think Scottish independence is more important than economic, regional, and political stability?’ or ‘does your Scottish
identity require the realization of full, regional, political autonomy?’. This juxtaposition of phrasing helps to solidify some of the implications that voters juggled when voting on the referendum question—and helps to ground the theoretical question of independence, which took a geo-political debate over autonomy into the realm of identity politics, and what it really means to be Scottish.

The considerations of implications at question in Scotland in September 2014 reflect the circumstances that may lead to the recalculation of priority associated with identity in Korostelina’s discussion of “satisfaction by fulfillment of [identity] functions” (185). She argues for the psychological, basic needs, delineated earlier, that she argues social identity fills. She continues that if the national identity overtakes a social identity in filling these psychological needs “it can lead to the quick disappearance of other social identities” (185). This dominance of national identity when paired with conflict or an external threat to the full realization of the national identity can lead to manipulation and violence.

Gellner discusses this slippery slope of national identity when it comes to conflict, where he posits, “nationalism is a theory of political legitimacy, which requires that ethnic boundaries should not cut across political ones” (1). It is also at the heart of the question asked of Scots on September 18th, when they were asked if Scottish nationalism outranked other social identities. Andreas Wimmer’s research, published in 2013, shows “[m]ost of today’s more prominent and protracted wars are also associated with the national principle” (2). The national principle, for Wimmer mirrors that of Gellner and Smith, statehood should equate to nationalism and nationalism equates to a homogeneous
national population. Wimmer finds, when national borders and nationalist populations are not aligned seeds of conflict are sewn. For Wimmer, these seeds of conflict sprout twice as often if nationalism has gained a platform in politics (5).

These questions are not simple ones for individuals to sort through. Juggling multiple identities can be an unconscious facet of humanity, or it can cause internal and external strife. And, when politics get involved, it muddies the water further. These questions were internal battles for some individuals; many others discussed how they answered these questions on Twitter as for them it answered Twitter’s foundational question of ‘what is happening?’. In Scotland, we see these factors at odds; some are satisfied with the social identity of Scotland within the broader constructs of the United Kingdom at least in comparison with the potential tradeoffs associated with declaring independence, while others hold a stronger national identity and harbor nationalist goals in establishing a Scottish state separate from the United Kingdom. Through Twitter activity from surrounding the referendum, we see Scots, Brits, and the world react to the new distinctions Scots made within their own identity group.

**Role of the Self and Others**

As has been hinted at so far, there is a critical divide between the identification of the self (or ingroup) and the other (or outgroup) when it comes to identity—and in the case of the independence referendum, this distinction is even more pronounced. As “[s]ocial identity is usually described in terms of group similarities and reflects shared interests, values, and beliefs, while individual identity is defined as a set of individual features and provides a basis for differentiating an individual from a group of people”
(Korostelina 35). There are distinctions between the roles the self and the other play at individual and group levels, which are critical to delineate in the case of the Scottish referendum.

**The Scottish Self**

One of the first questions one must ask around self-identity is in regard to the unit of analysis for the identity in question. Are you considering individuals, or groups? It is critical that we address both, but we will begin with the individual. Here with unique individuals, we quickly can see how this directly translates to unique identities of individual selves. It is through the interactions of individuals with other individuals that we see differences in identity form.

When considering the frameworks of self-identity, the significant part of this study speaks to the ‘Scottish self’, what it is, and how it came to be. Sociologists at the University of Edinburg dove into both the individual and group levels of Scottish identity and how it is defined and redefined from person to person. McCrone et al. find that for most, “Scotland refers to ‘a sense of place’ rather than ‘a sense of tribe’” displaying what would be classified as “civic rather than an ethnic form of nationalism” (1998, 33). While, in the Scottish case, we are dealing in terms of civic national identity, we see the same types of boundaries, self-definitions, and personal flexibility of identity (and acceptance of identity) that we expect to see in other identity groups. The study allowed Scottish identity to be deconstructed to researchers in interviews by Scots themselves. Through this, we understand how individuals define their claim to Scottish identity and how they classify others in terms of legitimacy of Scottish identity. For the authors, the
interview results boil down to three facets that are strong enough to support a claim of Scottish identity independently. These strong markers of identity are: being born in Scotland, having lived in Scotland for an extended period, and having Scottish ancestry (1998, 1999, 2001).

To rank the importance of these qualifiers, they report at least 65% of respondents to rank each birth, ancestry, and residence as important in their study (1998). At the time this research was conducted, Scottish identity, bounded by both firm borders and straightforward timelines of union with the United Kingdom, is described as “largely uncontentious and unproblematic” (McCrone 1998, 632). This uncontentious nature of Scottish identity, the authors attribute to the key identifiers—birth, length of residence, and ancestry—are difficult to hone in on based on appearance, so there is generally an air of ‘assumed Scottish-ness’ which allows for social interactions to flow smoothly (1998, 650). However, McCrone et al. do draw attention to “Scots who go elsewhere or make disparaging comments about Scotland”. The authors point to interview transcripts that point to this as a valid reason to question or outwardly reject someone’s Scottish nationality (2001, 40).

Mc Cone et al.’s work displays Scottish identity as established and defined, but relatively settled and uncontentious as many of the markers are not immediately obvious to outside observers or strangers passing on the street. But identities are flexible by definition and events like an independence referendum can call individuals to consider what their identity means to them and heightened competition around identity group and

---

38 The authors base this on day-to-day life, which excludes sporting rivalries (34, 2001).
meaning making in politics creates opportunities for re-negotiation of identity. In the case of a group identity like what we have in Scotland, it can cause not only individuals to recalculate what their identity means to them, but also how they classify other’s claims to their shared identity and allows for new markers, as McCone et al. deem them, to arise in society. Statements to this effect are present on Twitter as discussed below.

The Other

By definition, the importance of these commonalities and benefits drawn from them are products of the characteristics being unique to that group which directly transfers to excluding those who do not share in these in-group features. Therefore if there is a group identity, there must be an ‘other’. This stems from the concept that, “social identity is connected not only with the perception of similarities within an ingroup…but also the perception of differences between this group and the members of the other groups” (Korostelina 24). Considering the concept of ‘the other’ on an individual level is relatively natural. We are individuals and there are other individuals with whom we interact at differing levels of intensity. These distinctions in conceptualization and reconceptualization of ‘the other’ are important to this study as there are ‘other’s’ who do not share the same common history or core identity, and then there are those who do not share the same values or beliefs toward independence within the ingroup. As shown in McCrone et al.’s work-the definition of what it means to be Scottish changed when individuals discussed their friends or co-worker’s claims to the Scottish identity (2001, 40).
In the case of Scotland, we see multiple constructions of the self and the other. In some analysis, we can see Scotland as the self, with the UK and the rest of the world playing the role of the other, while in other cases, it’s those favoring independence globally including Scots, Catalonians, and other vocal supporters of self determination with the other being those fighting for the union of Great Britain. In this analysis, of the ‘other’ we will focus on the newly established division between Scots who are pro-independence and those who are pro-union, in other words, the Yes voters and the No voters in the referendum.

Building off McCrone et al.’s work, we find connections with Anna Triandafyllidou’s which tackles part of the question of how these concepts of ‘the other’ impact national identity. Triandafyllidou brings to light the concept of, “‘significant others’ to investigate the ways in which others may condition the formation or lead to transformation of the identity of the ingroup” (593). In her article Triandafyllidou, accepts the premise delineated with regard to nationalism that, “opposition to the other is taken as an intrinsic feature of nationalism” (596) but continues that both nationalists and those who study nationalism as an absolute; either people see themselves as a nation or they do not. Triandafyllidou challenges this mentality as misleading as the benefits from national identity have relative value both to their view of foreigners and themselves at that time as national identity is made meaningful in contrast to other nations (599).

Triandafyllidou distinguishes between other nations as those who are relatively benign and those who “have influenced the development of its identity [the nation at question] by means of their ‘threatening’ presence [the other nation]” (600). These
threatening nations are deemed ‘significant others’ and they “pose a threat to the existence of the nation [the nation at question]” (600). When considering the role of the other in Scotland-post referendum, one can see how this relates to the case at hand, Triandafyllidou relates the significant other concretely to a threat to the nation, “the significant other may also be a group that threatens to blur the distinctiveness of the ingroup” (600). In the case, of Scotland, we find the role of the ‘significant other’ played on both sides of the vote for independence as those pro-independence Scots see those No voters as a direct threat to their ability to tangibly realize their nationhood. On the other had, pro-union Scots see the Yes campaign as a direct threat to their idea of nationalism thus causing friction within the formerly unified Scottish identity.

**The Use and Functionality of Sentiment Analysis**
When considering the meat of the analysis of the studies of public opinion through Twitter, most studies focus on sentiment analysis as a tool to understand the true meaning behind messages. Sentiment analysis, or “the task of identifying positive and negative opinions and evaluations” began at the document level and as the character count limits on our media shrank, the challenges in establishing sentiment, feeling, or mood in the media have increased (Wilson, Wiebe, and Hoffman 347). Computer science and the social sciences began tackling these challenges, as Brendan O’Connor et al. argue, “[e]xtracting the public opinion from social media texts provides a challenging and rich context to explore computational models of natural language,” (1). While this may be interesting for computational linguistics, the work that is being done is directly
influencing the work of political science and other social sciences interested in seeing how events, like elections or referendums, unfold over time and impact public discourse.

A traditional method of establishing sentiment is to establish the polarity of word choice as defined externally and performing a count of positive and negative sentiment within a document, blog post, or now, tweet. Researchers looking specifically at Twitter sentiment have found that sentiment analysis of the Twittersphere functions extremely similarly to sentiment analysis of other media when things specific to Twitter are included in the lexicons for analysis (e.g. acronyms, emoticons) (Agarwal et al.). Further, O’Connor et al. were able to replicate the same measures of “consumer confidence and presidential job approval polls” with basic sentiment analysis applied to Twitter suggesting the huge potential for Twitter based polling or supplemental public opinion analysis based firmly in content and sentiment analysis.

While Agarwal et al. and O’Connor et al.’s work is promising when it comes to sentiment analysis, there are challenges to this field of analysis. Wilson, Wiebe, and Hoffman urge for caution and additional work when it comes to phrase-level sentiment analysis, which is where Twitter falls on the spectrum. Their study finds that their system, which looks at the phrase level “contextual polarity” adds significant clarity to the results. Further, Gayo-Avello urges additional care to be taken with any type of social media sentiment analysis, particularly when politics are involved. He argues, “[s]entiment analysis is a core task, and simplistic sentiment analysis methods should be avoided one and all” (2012 93). This he argues because of the types of humor, including puns, sarcasm, and double-speak that is so common in political analysis and everyday
conversations. While, Gayo-Avello does not completely dismiss the notion that Twitter analysis of elections or sentiment analysis as a part of that analysis will not be fruitful, he does urge caution across the board when it comes to simplistic, or untargeted and unrefined, methods of analysis of social media and Twitter data.

Of note all these studies contributing to the current understanding of the functionality of sentiment analysis, they relied on mechanized sentiment analysis and formalized lexicons of polarized words (Agarwal, O’Connor, Wilson, Wiebe, and Hoffman). This mechanized analysis may be a reliable source for a large-N Twitter based study, however, it does not allow for comparison with hand-coding of Tweets based on an understanding of the events occurring and a person actively decomposing the humor, sarcasm, and external content included in the tweets, as this study relies on for its analysis. It should be noted, however, that there is an inherent tradeoff with hand coding. Processing tweets through computer software runs automatically and thus, allows for a large sample size, while had coding requires individual attention to each word and phrase included by Twitter users—resulting in a small sample size. This tradeoff is between number of tweets analyzed and the depth of analysis conducted. This study opted for the depth of analysis with a smaller number of cases because of the benefits provided considering the unique constraints. Particularly in the case of the Scottish referendum where sarcasm and regionally specific language are extremely present, we also see non-traditional meanings associated with the words ‘yes’ and ‘no’ where they became more than showings of consent or rejection, and they lie the very heart of the argument, where candidates names or parties would traditionally fall. All of these factors complicated the
analysis but were easier to overcome with a smaller-N for the study and a person-rather than a machine-making the coding decisions.

This study allows for the unique analysis of the Scottish referendum in the immediate lead up to the referendum and the time as individuals wrestled with the results to this decisive vote. By using the unique case of a referendum, with a strong silent majority, we see that Twitter mirrors the campaigns, not the results of the referendum. Through hand coding, we see that the nuance is critical to understanding the way the tweets represent the public opinion of the Twittersphere. And with such a competitive and national vote with a free press, this study directly challenges Ko et al.’s inference that Twitter will better track public sentiment when it comes to these conditions.

**Research Design**

Taking what we see in the identity construction literature and merging it with the sentiment analysis literature from social media analysis is what motivates this research design. The three days of tweets included in this study, were analyzed for emotionality, as described below, as well as content as described in the previous chapter. This pairing of measures allows for the analysis of individual sentiment and level of support for the side of the referendum supported. These measures will be analyzed over the same timeframe of the referendum as in the previous chapter. However, in order to seek out the construction and reconstruction of national identity and what it means to be Scottish, this chapter concludes with a deep dive into the data from September 19. Analyzing September 19th specifically allows for observing identity re-definition as it occurs on Twitter with the announcement of the results of the referendum.
Coding Rules
In order to harness the overall messages from Twitter, additional work is needed. Here the coding rules for this study are discussed briefly to clarify process and contextualize results. Coding schemes are unique to this paper as all coding and sentiment analysis was done by hand rather than through mechanized sentiment analysis.

For the previous chapter, the 1500 randomly sampled tweets were categorized into a broad range of categories based message content,\(^{39}\) I then turned to code degree of emotionality. To analyze the emotionality in the messages, tweets were ranked on a 7-point scale to measure the emotional charge and content of the message. The scale used plots neutral or strictly informational or results based tweets at zero. If a tweet contained a slant or statement of support, they were categorized on a scale of 1 through 3. The scale proceeds in each direction in a linear fashion from zero, tweets supporting the yes vote are coded 1, 2, or 3 each with clear distinctions on how they progress up the scale in levels of emotionality (1 being least intense emotionally and 3 being most intense emotionally). Tweets voicing support of the no campaign proceed through -1, -2, and -3, following an identical progression as the yes vote tweets just in the opposite direction (again, 1 being least intense emotionally and 3 being most intense emotionally). This results in a -3 to 3 range, including zero, for the data in this section.

\(^{39}\) If the intent behind the tweet was unclear based solely on the text of the tweet, any hyperlinks or photos included in the tweet those were considered when categorizing the tweet. Also, if sarcasm was suspected, the links and photos were considered to clarify.

\(^{40}\) Table 2, displayed in the previous chapter, shows the distribution of the categories across the categories and regions.
When approaching the coding of this section, delineations were included for reliability. Tweets included in the neutral category mostly included sharing of results or expectations of when results would post. These tweets were seen to be information sharing or analytically based with no detectable slant to either side whatsoever.

Tweets coded 1 or -1 demonstrate some level of support for the respective side of the campaign, but they go no further. For the yes vote side of the population (the 1’s), there may be general support for the Scottish National Party or reference to the #45percent. These tweets commended the hard fought campaign and congratulated the democratic process including the size of the turnout, or speculate regarding what could have happened if the results turned out differently. On the No Vote side (the -1’s) similar things were present, including congratulatory posts or retweets about the campaign and process with a focus on moving forward. Further, tweets that were results based that used capital letters in writing only YES or NO were coded as 1’s and -1’s respectively, as these were not only announcing the final result, they announced verified results from certain cities or precincts indicating the Twitter user’s support for one of the campaigns.

The tweets coded 2’s and -2’s there are greater distinctions between the Yes and No camps but they represent equivalent emotions. For a tweet to gain a designation of ‘2’, they had to reference cheating or rigging of the referendum or vote count, or provide a ‘warning’ regarding the future of politics in Scotland or Great Britain. Further, they condemned the coverage of demonstrations with some level of sarcasm toward the ‘better together’ slogan indicating their disdain for the other-the no voter. These tweets also

---

41 #45percent was a popular hashtag that arose after the results were announced.
clearly defined an us/them distinction, a division that was not present in the 1 and -1 tweets. This distinction could have been made by distinguishing the role of ‘the other’ as those individuals who ‘chose incorrectly’ or that the Twitter user ‘cannot understand how someone would choose to vote no’ to gain the distinction of a 2. Further, these tweets expressed remorse at the way the results turned out or attempted at humorous commentary on the results but clearly conveyed the disappointment of the user. On the -2 side, these tweets supported the no campaign and the results in a celebratory manner. These tweets expressed pride in Scotland or Great Britain for ‘choosing properly’ or ‘supporting the union’. Tweets that categorized the UK or Scotland into ‘ingroup’ and ‘outgroup’ were coded here. Additionally tweets that indicated the user felt the United Kingdom was stronger because of the referendum process were categorized here.

Tweets categorized as 3’s or -3’s express sentiments similar to what is included in the 2’s and -2’s but are extreme. In the 3 and -3 categories, there is a clear us/them distinction, but there is also a claim of superiority for the Twitter user’s side of the referendum. For a 3, the user references violence, uses capitalization of full words or phrases for emphasis, or are vulgar, or identifies the ‘other’ as someone who chose differently and advocates for some for of retribution or action against them, which could take the form of a verbal attack. -3’s have similar responses. Tweets in this category refer to direct attacks, use vulgarity, or demean the ‘other’ or refer to them as ‘rebels’.

---

42 Being those who voted no as an ‘ingroup’ and those who voted yes as the ‘outgroup’—rather than a unified ‘us’.
For the deep dive into emotionality and identity construction, we look at only the data from midnight September 19 to midnight September 20, 2014 BST. As discussed above, this sample is a random sample of 500 tweets from this 24-hour portion of the corpus of tweets made available to this study that included time zone classifications. In order to understand the tweets originating from Scotland and the UK, all Edinburgh time zones were categorized into a Scotland category. All the London tweets were combined into the UK category, and all other tweets were categorized as ‘World’. This re-categorization of data was done in order to analyze the difference in sentiment and in volume of the tweets surrounding the referendum inside Scotland, the United Kingdom, and the world as will be displayed in the data below.

Results

The Overall Numbers

Dipping into the distributions of emotionality over time, we see interesting trends. As seen in Figure 10, we see peaks each day in the neutral category, as expected. Additionally, as expected we find that the distributions drop off as we progress up the scale to more and more extreme emotion. Figure 10 displays to us that most people are using Twitter as an information sharing system, which is confirmed in the previous

---

43 British Summer Time as Scotland was observing this form of daylight-savings-time during the referendum.

44 This time frame yielded 935,187 tweets that fit the inclusion criteria given to Gnip.

45 There were no other UK or Scottish cities included in the time zone data. Dublin was broken out as a separate entity—it was added to the World category along with all other cities in the British Commonwealth.
chapter when considering the large number of tweets that fall into the information/results category.

Looking at Figure 10 further, we note that September 19th by far has the most emotional tweets of any other day, followed by September 17th. Interestingly, the day of the actual vote, the day where the competition actually occurred, there were the fewest emotional tweets sent. However, the day before the vote and the day the results were announced yielded tweets that are more emotional than the day of the contest itself.

Breaking down these distributions by region is similarly enlightening, as we see Scotland’s tweets as very emotional on the side of the Yes vote but much less so on the No vote side seen in Figure 11. This type of skew is not present in any of the other regions broken out in this study indicating that the heart of the tension resided in the

Figure 10: All tweets based on emotionality by day.
Scots for this election—just like the overall trends in the Twitter chatter, Scotland drove the heart of the discussion and emotion surrounding the September 18th vote.

Figure 11: Count of emotionality over the full three days by region.
When considering all the questions one can ask of data from Twitter, it can easily become overwhelming in every way, shape, and form. However, through breaking down the data along regional lines, we can see clear patterns that confirm expectations regarding Twitter and how users utilize the site, and it illuminates specific things about the Scottish referendum’s impact on identity that are particularly interesting.

When it comes to tweet counts, it is important to consider the phenomenon of the silent majority. Here, on the day after the referendum, tweets that are clearly supportive of the yes campaign total nearly 60% of the Scottish tweets as compared to the 8% of the Scottish tweets that the No campaign sees in this data. While these differences, for the most part, equalize as we zoom out to the UK and the world sub-categories, we still are faced with the question of numbers of tweets, and how they relate to the numbers of actual voters. This is a cautionary tale as addressed in previous sections focused on the predictive capabilities of Twitter, as even retroactively Twitter is not predictive of the results of the vote, rather Twitter is a better measure of public discourse in this case.
Figure 12: This chart collapsed the overall emotions into four categories progressing up the scale of emotion. 1: neutral tweets, 2: slight emotion, 3: moderate emotion, 4: extreme emotion.

The emotionality scale established allows us to visualize the distribution of emotion reflected in the tweets as they relate to the sentiment expressed. As shown in the data, most of the tweets fall into the 0 category indicating that they have no sentiment in this regard. As shown above, nearly half the Twitter traffic captured in this sample (243 tweets are coded as information/results) is information sharing, much of that based on sharing results as they are announced from different regions. In addition to diving into the distribution across the seven-point scale of emotionality, we can also collapse the categories down to a four-point scale where tweets with equal emotions, regardless of their support for the yes or no campaign are examined together as seen above in figure 10.
First, we notice there are many discussions going on, and that those discussions across the regions of interest focuses on unslanted information sharing—as the neutral category comprises over 40% of all tweets from each region. Across the board in the graphs above the distribution of the data peaks at the 0 category, indicating, based on the coding rules, a very large percentage of the tweets sharing information, articles, results, or projections without a political slant, focused on information sharing. This peak in the neutral category reinforces the counts from the categorical variables discussed in the previous chapter. What we see is many Twitter users in Scotland, the UK, and the World following the referendum, engaging on the topic, and even providing analysis without an apparent bias to either side. But it should be noted that as the emotion progresses up the scale here, the distribution of extreme emotion in Scotland, with the United Kingdom lagging behind. Moving past the conversations regarding information or analysis, we find commentary that is more emotionally charged and indicative of what individuals

<table>
<thead>
<tr>
<th>SCORE OF EMOTIONALITY</th>
<th>SCOTLAND</th>
<th>UK</th>
<th>WORLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXTREME NO</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>MODERATE NO</td>
<td>4%</td>
<td>10%</td>
<td>2%</td>
</tr>
<tr>
<td>SLIGHT NO</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td>NEUTRAL</td>
<td>41%</td>
<td>56%</td>
<td>66%</td>
</tr>
<tr>
<td>SLIGHT YES</td>
<td>13%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>MODERATE YES</td>
<td>24%</td>
<td>11%</td>
<td>6%</td>
</tr>
<tr>
<td>EXTREME YES</td>
<td>7%</td>
<td>2%</td>
<td>1%</td>
</tr>
</tbody>
</table>
who are reacting to the results and what the new status quo means to them as an individual. Looking now to the 1-3 and -1- -3 tweets, we see interesting patterns that generally hold across regions. First, the distribution skews toward Yes supporters tweeting more often across the world. We find the ever so slight positive skew could be equated to the silent majority of the No campaign tweeting less and being quieter. This holds for the slight (1 and -1) and moderate (2 and -2) categories, but by getting to the extreme categories much of this relationship falls away. This indicates that there are those on the fringe on each side, however, extreme they may be they are few in number overall, but much more prominent in Scotland than the rest of the world as shown across the board through this study. We see September 19th as the most emotional day, and Scotland as the most emotional region on September 19th overall when looking at the aggregated numbers. With these trends holding across the board and knowing that the overall traffic on Twitter surrounding the referendum held with local traffic trends, we see clearly that this fundamentally local question was most consuming for Scots themselves and we can anticipate that the fall out politically and emotionally will be centralized in Scotland as well. 46

46 Further attention could be paid to the upcoming elections in the UK and the impact of the Scottish National Party in the overall vote and the representation of Scotland in Westminster. However, that research is outside the scope and timeline of this research.
Figure 13: Emotionality of tweets from September 19 by region.
**Results—The Words**

Turning now, deeper into the content of the tweets that demonstrate the reaction to the referendum results and resulting struggles with identity, we see another side of the Twitter messages that is not easily captured with big quantitative data analysis. When coding and categorizing the data for this section, there were a handful of tweets that stood out as particularly demonstrative of an individual struggling to make sense of what the results of the referendum truly mean and what it means for them.

Here, I present seven specific tweets, one from each of the -3 – 3 categories to help demonstrate the meaning behind them, the differences between the messages, and more importantly, what this means for the Scottish identity played in this referendum. Beginning in the middle and working outward, we will dive into the intent behind the messages.

![Figure 14: Tweet Coded as a 0 or neutral message (account).](image-url)
Figure 12 represents the tweets coded 0. While a significant number of tweets in this category fell into the information/results category messages contained positive messages toward the democratic process, as presented in Figure 12. There is no distinction of the political view of Owen Jones, the Twitter user (or the retweeter), as he does not associate himself with either the downhearted Yes campaigners or the successful No campaign, but instead focuses on the system that allowed for this vote and the positive nature of the campaign—bringing the two sides together under the auspice of one, united Scottish identity. Additionally, by linking to the #indyref, Jones directly engages with others wrestling with the results and shares his message which saw great support on Twitter with 466 retweets and 358 favorites which indicates this was a message with great reach.

Figure 15: Tweet coded as a -1 or slight emotion favoring the No campaign (Pilko).

While the political views of Owen Jones are well known and publicized, for the purposes of this study, we are looking solely at the content of the Twitter messages.
Considering the two tweets together, both are coded as the slightly emotional category (-1 and 1, respectively). Through these messages, we begin to see divisions within the campaigns take form through Twitter. The Outlander Podcast\textsuperscript{48} shares the news that the results will first be announced in Gaelic (Figure 14). This implies those who keep the traditional, ancestral, language of Scotland alive, will know the official results first, proving the importance of Scottish identity.\textsuperscript{49} This encapsulates the sentiment of unity and strength in Scotland while promoting Gaelic in society at a culturally sensitive time. While Nick Pilko (Figure 13) presents his sympathies with the feelings of many No voters. Pilko, shares it would be difficult to choose something one

\textsuperscript{48} This tweet was populated in the dataset as a retweet from an individual, however it was linked as the original message from the Outlander Podcast. Thus, throughout the paper, I will refer to the Twitter user as an individual.

\textsuperscript{49} This with the included link to and external link, expands the conversation and empowers the Scottish identity while not driving the two sides further apart with the ceremonial step of announcing the results in Gaelic before immediately sharing them in English.
has never experienced with so many unknowns.\textsuperscript{50} This tweet shows thoughtfulness and consideration on a tough question that all voters faced. It shows the author sympathizes with the choices posed to the voters and he demonstrates no ill will toward the Yes campaign, but seems to say simply, ‘no thanks’ without the intensity of that now charged phrase.\textsuperscript{51} Through these messages, we see what resonates with people who make their stances known but do not feel as strongly as others do about the identity-based questions stirred up by the referendum.

Figure 17: Tweet coded -2 or moderate emotion favoring the No campaign (Miller).

\textsuperscript{50} According to Urban Dictionary, a top source on modern slang terminology, the tbh acronym stands for ‘to be honest’ (Tbh).

\textsuperscript{51} “Better Together” and “No Thanks” were two of the main campaign slogans for the No campaign.
Now we turn to the moderate emotion tweets coded in the -2 and 2 categories. Here, we see similar content to the -1 and 1 case. However these are significantly stronger in emotion. On the one hand, Miller in the -2 tweet clearly draws a line between her views and those who voted yes or her ingroup and outgroup as now defined by how a Scotsman voted on September 18th (Figure 15). She paints herself as a Scot, who voted to remain in the UK and that while she respects those who feel differently, she wants her stance known. Miller uses inclusive language, with collective pronouns, mentions of unity, and respect, but it clearly draws a dividing line between ‘us and them’ when it comes to how Miller feels regarding the #indyref. On the other hand, King, makes a different point (Figure 16). While also distinguishing between Yes and No voters, he utilizes exclusionary and negative language to make this point and distinguishing his in-group of Scots as Yes voters, and the out-group as No voters. He supplements his views with another tweet from a publication announcing that the promises of greater autonomy
and power for Scotland that were made during the campaign, have little hope of passing Parliament. King shares his views that the voters who chose to vote no because of the promise of additional powers were duped and shamefully cost the Yes voters the independence they sought and the autonomy promised if the referendum failed—all of this intended to shame the now-redefined out-group.

While the tones in these tweets are polar opposites, they mirror the entrenched nature of the views in this referendum and the difficulties in negotiating the question of national identity and social identity in one population that has now been forced to choose between the two. These two tweets demonstrate clearly the new divisions in identity made manifest by the referendum vote. Miller, claims pride in Scotland and pride in her British identity and hopes for continued unity within the Scottish people and territories that make up the United Kingdom of Great Britain and Northern Ireland. Her prioritization of national identity places British above Scottish. While King’s tweet clearly indicates that the British identity is not satisfactory and that, the Scottish nationalist cause aligns more closely with his political or economic beliefs.

Finally, this brings us to the extreme ends of the spectrum of emotions displayed in the tweets from September 19, 2014. Here, we see even clearer distinctions between the identity groups and feelings expressed toward the ‘other’ as the lines are clearly defined. The tweet by McCabe, categorized as -3, makes clear her support for the No voters, as she labels those who sided with the Yes campaign as “fools” and harbors no sympathy for the cause of Scottish independence as other No supporters indicate above (Figure 17). These distinctions allow us to clearly see the role of ‘other’ or even
Triandafyllidou’s concept of the ‘significant other’ who poses a direct threat to the livelihood or existence of the self. McCabe makes clear she would rather not have this new ‘other’ making decisions in the ballot box at all, but she concedes that regardless of independence they are all Scottish and have the same rights that she does. While, Ashworth falls on the other side of this spectrum (categorized as a 3). Here, the distinctions between the self include only those who voted Yes and the ‘other’ in this case are expanded to those who voted No and those who stayed away from the polls (Figure 18). Regardless of categorization within the role of the ‘other’, Ashworth paints with a wide brush of contempt. Scots (as specified by Ashworth), who fall into the newly formed ‘other’ are equally disgraceful, thus bringing to life the true heart of the cliché, ‘if you are not with us, you are against us”, again painting the picture of a ‘significant other’ only posed to do harm.

Figure 19: Tweet coded as a -3 or extreme emotion favoring the No campaign (McCabe).
Through these extreme tweets penned on the day after the referendum, we can see the scale of the schism within Scotland focused on the question of nationalism in Scotland. While the extreme sentiments were rare in the dataset across all regions considered, they are still there, and they still represent the views of individuals wrestling with new challenges to identity, nationalism, and political life.

**Conclusions**

What we find through this two-part analysis of tweets and Twitter traffic content, is that local Scots felt more strongly about this referendum than the rest of the world, and the majority of the emotion came as the results were announced and they began to settle into the population. This is made clear through the analysis of all the tweets, broken down by day but also through the deep dive into September 19th where we see the re-negotiation of Scottish national identity.
First, throughout all the analysis of emotionality over time and region, we find, each time that the locals in Scotland are discussing the election more in terms of what it means for them as Scots, and what it means for Scots more broadly. Rather than the rest of the world as the tweets coming out of Scotland are by far more emotionally charged than the other regions, which saw more ‘neutral’ tweets. Second, Scotland saw more overall and more emotional Yes slanted tweets across the differing levels of analysis provided here. Thus, indicating not only the presence of a silent majority, but also lending credence to the phenomenon of ‘wishful thinking’ in political campaigns as discussed by Kirzan et al. leading to enhanced disappointment for supporters of the Yes campaign and the potential for increased levels of emotionality in tweets from Yes supporters.

Further, from the individual tweets analyzed and the overall trends, the presence of a new division within Scottish identity is strikingly clear. As McCrone et al. found in their studies of how Scottish identity is defined by Scots themselves, they found the importance of length of residence, birth in Scotland, and Scottish ancestry as critically important markers of identity. While these likely have not changed within the minds of Scots, an additional layer of complexity was created through the independence referendum process. This complexity speaks to how Scots are now characterizing themselves along their referendum voting lines as a pro-independence or pro-union Scotsman. This additional marker falls in line with other divisions of Scottish identity whereby it is not readily visible, but it still exists below the surface. This is similar to the marker of being born in Scotland and can quickly arise as a distinguishing factor when
defining one’s views on who is truly Scottish and who may be a direct challenge to the realization of one view of Scottish nationalism.
CONCLUSIONS

Through this research, we find that overarchingly the Scottish referendum was advanced by local forces, concerns, and reporting. While this may not be all too shocking, the reinforcing nature of both the overall Twitter traffic matching the timeline of major turning points in the referendum and the trends in regional emotionality crystalize the locally driven nature of this referendum. The question of should Scotland be independent, was answered by Scots and the world reverberated the sentiments coming out of Scotland throughout the Twittersphere. This reverberation in the world discussions reinforces the significant role of the silent majority of the No Campaign, and Krizan et al.’s concept of ‘wishful thinking’ potentially muddied these waters further for the Yes campaign.

While the findings presented in the first chapter, do not mirror returns of the polls, the pace and content of the tweets matched the public discourse—if not public opinion—as displayed when polls opened and Yes and No tweets both spiked and when No tweets topped Yes tweets when the results were announced. These findings allow us to better understand the dynamics of Twitter use during elections and referendums and to hone methodologies based on these technologies further. Additionally, it allows us to reject the concept that in freer societies Twitter will better match electoral results as posited by Skoric et Al.
This study confirms that Twitter is a platform where individuals discuss politics, current events, and how they feel or make sense of these events. Confirming this fact, alongside the locally driven nature of the referendum-based Twitter traffic, it allows us to observe identity conversations at the individual level and to better comprehend the process of identity renegotiation and the driving forces behind individual decision making for the referendum. Here, we see a new distinguishing feature within Scottish identity—having voted Yes or No—thus adding a new marker to the identity calculus for Scottish individuals.

This individual level of comprehension allows us to understand better the mindset of the voter and the importance of different parts of the all-encompassing question of should Scotland become an independent country. These secondary, or even tertiary questions of banks, territorial security, currency, and trade are serious questions that take a great deal of weighing against the potential to realize an identity-based nation state that was possible in Scotland. Many Twitter users cited other priorities for Scotland separate from independence as rationale for their vote choice, or as at least something, they considered strongly before voting how they did. This is similarly mirrored in the debates surrounding the referendum held in Quebec where questions on trade, economic stability, and currency were cited as reasons to consider maintain the status quo (Bryant).

However, the parallels with other sovereignty movements do not stop with the similarity of questions between Scotland, Puerto Rico, and Quebec. Each referendum indicates a significantly divided society over the question of sovereignty as all three cases
present a nearly 50%-50% split between independence and union. Which raises a bigger question, what does all this mean for self-determination movements more broadly?

Scotland came the closest to actually being able to declare independence through the ballot box—when compared to the legal positions of Quebec and Puerto Rico in their referendums—as the vote on September 19th, would have definitively resulted in separation rather than the others, which were more cautioned in their legal wording. However, what is interesting in all these cases where the vote was so evenly split is that the question in the referendum was only one for change toward separation. That is to say, voters only had the choice to gain autonomy; there were no threats of losing what sovereignty was already established—in fact, in Scotland there were promises of additional powers with a No vote. In these referendums, we see established, respected, and stable identity groups asking for independence with no great threat to their survival as a group. Scotland has a devolved parliament and sends representatives to Westminster; Quebec has established language laws and participates in Canadian government equally to any other province; and Puerto Rico has many rights as an American Territory—although this comparison is the least strong in this case. But in these cases, we see the changing nature of the necessity of sovereignty to express a national identity in ways we have not seriously considered self-determination before.
REFERENCES


account, Owen JonesVerified. “Yes Campaigners Will Feel Downhearted - but They’ve Not Just Changed Scotland. They’ve Changed the Rest of Britain, Too #indyref.” microblog. @OwenJones84. N.p., 19 Sept. 2014. Web. 22 Dec. 2014.


Cederman, Widman, Gleditsch. Horizonal Inequalities and Ethnonationalist Civil War; APSR 2011


Marcella Harrison Morris, M.S. graduated from James M. Bennett High School, Salisbury, Maryland, in 2009. She received her Bachelor of Arts in Government and Politics and an Minor in Global Terrorism from the University of Maryland; College Park in 2013. She works as the undergraduate Education Coordinator for the National Consortium for the Study of Terrorism and Responses to Terrorism at the University of Maryland; College Park.