PsychOpen GOLD

Original Research Reports

From I to We: Group Formation and Linguistic Adaption in an Online Xenophobic Forum

Emma A. Bäck*[a], Hanna Bäck[b], Marie Gustafsson Sendén[c], Sverker Sikström[d]

[a] Department of Psychology, Gothenburg University, Gothenburg, Sweden. **[b]** Department of Political Science, Lund University, Lund, Sweden. **[c]** Department of Psychology, Stockholm University, Stockholm, Sweden. **[d]** Department of Psychology, Lund University, Lund, Sweden.

Abstract

Much of identity formation processes nowadays takes place online, indicating that intergroup differentiation may be found in online communities. This paper focuses on identity formation processes in an open online xenophobic, anti-immigrant, discussion forum. Open discussion forums provide an excellent opportunity to investigate open interactions that may reveal how identity is formed and how individual users are influenced by other users. Using computational text analysis and Linguistic Inquiry Word Count (LIWC), our results show that new users change from an individual identification to a group identification over time as indicated by a decrease in the use of "I" and increase in the use of "we". The analyses also show increased use of "they" indicating intergroup differentiation. Moreover, the linguistic style of new users became more similar to that of the overall forum over time. Further, the emotional content decreased over time. The results indicate that new users on a forum create a collective identity with the other users and adapt to them linguistically.

Keywords: identity formation, group formation, social adaption, linguistic analysis, online material

Non-Technical Summary

1. Background
All humans have a fundamental desire to belong to social groups, and create collective identities with groups. Groups online, such as chat forums, may also constitute a social group in a psychological sense. This means that individuals joining an online forum may start to identify with the other users, and hence adapt to these users. Moreover, users may start to distance themselves from other groups, so called outgroups. This could be problematic, for example if the forum is anti-immigrant and these other groups are immigrant groups.

2. Why was this study done?
This study was conducted to explore how processes of group identification and adaption can be studied in an online forum. Specifically, this is of importance if such forums are used to spread anti-immigrant messages or to recruit people to radical groups.

3. What did the researchers do and find?
In order to explore how identification processes and adaption takes place in an online milieu, we have to use the text generated by the users in an online forum. Much of these processes can be studied analyzing how language is used. We use a computerized method, which basically counts the use of specific types of words, and we follow individual users over time from the time they join the forum. In the present analyses, we used text data comprised of about 60 000 000 words. The forum that we analysed in this study is an online forum in Sweden, which has over one million users. The sub-forum that we focused on is under the theme Immigration and integration. Even though the online forum is not

explicitly racist, it is heavily dominated by anti-immigrant sentiments. The linguistic analysis revealed two major findings. First, we see a change in the use of pronouns over time. The use of 'I' decreases, while the use of 'we' increases when an individual has participated for a longer time in the discussions in the forum. In addition, the use of 'they' increases over time. Second, we see that the linguistic style of new users becomes more similar to the linguistic style of the forum as a whole over time, indicating that individuals adapt to the other users in the forum.

4. What do these findings mean?

These findings indicate that processes of group identification and adaption take place in online forums in a similar way as in real life settings. The decrease in use of 'I' and increase in use of 'we' signals a collective identity formation within the forum. The increased use of 'they' signals increased distancing to one or more outgroups. The linguistic adaption is also a signal of normal group processes, showing that individuals in online forums want to be part of the group and hence adapts to the norms of the group.

*Corresponding author at: Department of Psychology, Gothenburg University, Box 500, SE-405 30, Sweden. E-mail: Emma.Back@psy.gu.se

Humans are inherently social and desire belongingness to social groups (Baumeister & Leary, 1995; Tajfel & Turner, 1986). According to social identity theory, identifying with a group provides psychological benefits (Tajfel & Turner, 1986). Group identification constitutes the basis on which individuals build their sense of self (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), and entails psychological investment in the group (Tajfel & Turner, 1979; Tajfel & Turner, 1986). Ingroup identification necessitates the existence of one or more outgroups (Brewer, 1991; Gustafsson Sendén et al., 2014b). Hence, as an individual starts to identify with a group they should also show signs of differentiation to other group(s).

When an individual starts identifying with a group, the group also exerts influence on the individual members (Kelman, 1958). Social influence is broadly defined as any change - emotional, behavioral, or attitudinal – that has its roots in others' real or imagined presence (Allport, 1954). There are many different forms of social influence defined in the literature, such as conformity, compliance, obedience, internalization and identification to mention a few. What differs between these forms is the reason for the observable adaption. The reason may be a genuine attitude shift, such as when an individual internalizes attitudes, which could be based on identification with a group (Deutsch & Gerard, 1955; Miller & Prentice, 1996). It could also be public compliance displayed for increased acceptance that is not accompanied with a genuine attitude change (Goldstein, Cialdini, & Griskevicius, 2008). Regardless of why an individual displays an observable behavioral change that is in line with group norms, social identification with a group is the basis for the change.

In social psychological terms, a group is defined as more than two people that share certain goals (Cartwright & Zander, 1968). However, research also shows that even the most arbitrary shared attribute has the power to

elicit group processes (Tajfel & Turner, 1986), known as 'minimal groups' (Tajfel, 1970). An example of this is that splitting people arbitrarily into a red and blue group leads to intergroup differentiation (Frank & Gilovich, 1988). In line with this, much research has established that attitudes have the power to function as group boundaries (Bäck et al., 2011; Biernat & Vescio, 1993; Kenworthy & Miller, 2002; Mackie, Devos, & Smith, 2000; Simon, Hastedt, & Aufderheide, 1997). Hence, it is possible to think of online communities, where the members may share certain attitudes, as psychological groups that provide individuals with the benefits of being part of a group similarly as any real life group would. Moreover, such groups should exert the same sort of social influence as other groups (Tajfel & Turner, 1986). Processes of social identification, intergroup differentiation and social influence have to date not been studied in online forums. The aim of the present research is to fill this gap and provide information on how such processes can be studied through language used on the forum.

Social Networks as Arenas for Identity Formation

The popularity of social networking sites has increased immensely during the last decade. At the same time, offline socializing has shown a decline (Duggan & Smith, 2013). Now, much of the socializing actually takes place online (Ganda, 2014). In order to be part of an online community, the individual must socialize with other users. Through such socializing, individuals create self-representations (Enli & Thumim, 2012). Hence, the processes of identity formation, may to a large extent take place on the Internet in various online forums.

The political arena today is becoming increasingly polarized and right-wing populism is increasing across the globe. One of the most important political issues is immigration, where populist parties create anti-immigration sentiments increasing intergroup differentiation (Bäck & Bäck, 2017; Mudde, 2007; Stanley, 2008). Studying how anti-immigrant ideas are discussed and evolved among the population have been difficult for several reasons. First, individuals sympathizing with right-wing populist parties are difficult for researchers to reach. Second, because anti-immigrant attitudes have been strongly disapproved of, the arenas where such sentiments can be voiced have been closed to outsiders.

It is commonly assumed that the Internet has significantly changed the forms of political engagement, especially for radical political groups (Thompson, 2011). The Internet has provided radical groups with important platforms for exchange of ideas, ideological development, and resources for marketing (Edwards & Gribbon, 2013; Neumann, 2013). In addition, as many of these forums are open for public view they provide an important source of insight into how radical politics is justified, motivated, and narrated (see Owens & Palmer, 2003; Caiani & Parenti, 2009). This is knowledge that has previously been hard to attain due to the lack of access to radical groups and their arguments.

In order to explore processes of social identification, intergroup differentiation and social influence in an existing online setting, one is limited to the written text provided by the users. Of crucial importance is how such processes can be captured and analyzed using only written language available from the posts on the forum. Previous research indicates that language has the potential to provide such information, for instance through the analysis of the use of personal pronouns, and other word choices.

Personal Pronouns as Markers of Identity and Group Differentiation

Personal pronouns are linguistic elements that indicate how individuals view their roles in social groups (Campbell & Pennebaker, 2003; Chung & Pennebaker, 2007; Pennebaker, 2011a). Personal pronouns substitute and represent social categories, but do not explicitly refer to particular groups. As such, they are used relative to the speaker's

point of view and provide a non-reactive way to explore social psychological phenomena. Past research has shown that pronoun use is correlated with status, group identity and social relations (Chung & Pennebaker, 2007; Gustafsson Sendén, Lindholm, & Sikström, 2014a, 2014b; Kacewicz et al., 2013). For example, a shift from 'I' to 'We' was found to reflect a change from an individual to a collective identity (Brewer & Gardner, 1996; Fitzsimons & Kay, 2004; Gustafsson Sendén et al., 2014b). Social status is also related to the extent to which first person pronouns are used in communication. Low-status individuals use 'I' more than high-status individuals (Dino, Reysen, & Branscombe, 2009; Kacewicz et al., 2013; Pennebaker, 2011a; Slatcher, Chung, Pennebaker, & Stone, 2007), while high-status individuals use 'we' more often (Kacewicz et al., 2013; Slatcher et al., 2007). This pattern is observed both in real life and on Internet forums (Dino et al., 2009). Hence, a shift from "I" to "we" may signal an individual's identification with the group and a rise in status when becoming an accepted member of the group.

Ingroup identification also implies differentiation to one or more outgroups (Brewer, 1991). Ingroup identification and formation is also affected when there is perceived competition between groups (Brewer, 1991). In the classical experiments by Sherif and colleagues (Sherif et al., 1961), the ingroup became more cohesive when competition with other groups was salient. In addition, social exclusion functions to strengthen the ingroup (Williams, 2007). In a political setting, populist and anti-immigrant arguments build on the differentiation between an ingroup and one or more outgroups. For instance, linguistic analyses of American Nazis have shown that use of third person plural pronouns (they, them, their) is the single best predictor of extreme attitudes (Pennebaker & Chung, 2008).

Taken together, previous research shows that it is possible to study group formation and differentiation through linguistic signals. Especially, this should be shown in a shift from 'I' to 'We' (Brewer & Gardner, 1996; Fitzsimons & Kay, 2004; Gustafsson Sendén et al., 2014b; Pennebaker, 2011a), but also in an increased differentiation to one or more outgroups as signaled by increased use of 'they' (Pennebaker & Chung, 2008). Drawing on this literature, we present a first hypothesis saying that *the use of 'I' will decrease over time and the use of 'we' and 'they' will increase relative to the overall usage of pronouns (H1).*

Being part of a social group also entails that the individual member will change their behavior and attitudes to fit with the group (Tajfel & Turner, 1986). Regardless of why an individual displays changes in behavior or attitudes due to social influence from being a member of a group, such changes are markers of group influence. Norms can be created and sustained through group interactions and provide new members with guidelines to which they can adjust in order to be a valued member of the group. Because language can be seen as behavior (Fiedler, 2008), it may be possible to study processes of social influence through linguistic analysis. Thus, our second hypothesis is that *the linguistic style of new users will become increasingly similar to the linguistic style of the overall forum over time (H2).*

Finally, it is possible to detect changes in content of posts over time. There is an ongoing debate about how online communities function in their way of censoring or promoting development of ideas. Sunstein (2001) argues that online communities are so called 'echo chambers', where the same arguments are reinforced and repeated, and that such communities are closed for deviating ideas, which are censored (Jamieson & Capella, 2009). If an online forum functions as an echo chamber, we would not suspect much variation in the content of the posts over time, besides for perhaps the linguistic style. Because of their nature, echo chambers are detrimental to evolution of ideas, gaining knowledge and new perspectives.

However, this view of the Internet as being composed of echo chambers has been criticized (Brundidge, 2010). For example, Garrett (2009) argues that the Internet does not have the perfect conditions for avoiding unwanted

information. On the contrary, it is difficult to completely avoid politically diverse information. Internet forums provide individuals with opportunities to discuss, develop ideas, and get new information from other members (Edwards & Gribbon, 2013; Neumann, 2013), and most likely they will meet contradicting arguments to which they need to respond. In fact, Karlsen et al. (2017) find that most people who participate in online discussions with like-minded others, also discuss with people who disagree with them, and very few people state that they are never contradicted in online discussions. Moreover, most people state that encountering contradicting arguments leads to increased strength in their prior position, and almost half of the participants stated that they learned something from such contradicting arguments (Karlsen et al., 2017).

Taber and Lodge (2006) show that when individuals encounter evidence that supports or contradicts their prior political beliefs, they uncritically accept supporting arguments, but also actively counter contradicting arguments. Karlsen et al. (2017) show that this also happens in online political environments and that this results in strengthened attitude positions. The fact that one needs to counter contradicting arguments, may force individuals to more elaborate thinking. This indicates that the content of the posts in an online forum may also change over time as arguments become more fine-tuned and input from both supporting and contradicting members are integrated into an individual's own beliefs. This is likely to result (linguistically) in an increase in indicators of cognitive complexity. Hence, we hypothesize that *the content of the posts will change over time, such that indicators of complex thinking will increase (H3a).*

Finally, it is possible that there will be a decrease in negative emotional content. This idea is primarily based on the nature of the specific forum under scrutiny. This forum is dominated by anti-immigrant sentiments, which may allure to individuals seeking an outlet for negativity surrounding immigration where this is perceived to be acceptable. Garcia and colleagues (2016) show that the content of individuals' expression depends on their valence, and they show that their arousal significantly decreases afterwards as a regulation mechanism. This result indicates that after having expressed negativity in the forum, the need for such expressions should decrease. Hence, we expect that *the content of the posts will change such that indicators of negative emotions will decrease, over time (H3b).*

Through analyses of word use in posts and how they change over time, it is possible to capture both complex thinking and emotions. Words that reflect cognitive complexity include both content and function words. Content words for example include *understand*, *realize* and *because*. Function words for example include *but, without, never, to* and *for*. For instance, it has been shown that prepositions (e.g., *to*, *for*) signal that the speaker is providing more complex information about a topic (Fernández et al., 2013; Tausczik & Pennebaker, 2010). Higher use of these types of words has been related to the ability to juggle complex ideas (see Tausczik & Pennebaker, 2010, for a review).

The Present Study

The present study investigates how processes of group identification, group differentiation, and social influence change over time for a new user in on online xenophobic forum. Moreover, we investigate the change in content of posts over time. We do so by using computerized text analysis techniques that make it possible to analyze large text corpora, and that minimize the possibility of bias in the interpretation of texts.

The specific forum that the present research utilizes is a xenophobic forum, but it is not advertised as a purely racist or white supremacist forum. Instead the forum is presented as a "very liberal forum", where people are able to express their opinions, whatever they may be. This "extreme liberal" idea implies that there is very little censorship,

which has resulted in that the forum is highly xenophobic. Nonetheless, due to its liberal self-presentation, the xenophobic discussions are not unchallenged. For example, also anti-racist people join this forum in order to challenge individuals with xenophobic attitudes. This means that the forum is not likely to function as a pure echo chamber, because contradicting arguments must be met with own arguments. Hence, individuals will learn from more experienced users how to counter contradicting arguments in a convincing way. Hence, they are likely to incorporate new knowledge, embrace input and contribute to evolving ideas and arguments.

Methods and Data

Computational Text Analysis and Linguistic Inquiry Word Count

The analyses presented here are computational analyses of large-scale text data, which can be used to investigate a wide variety of psychological phenomena (Chung & Pennebaker, 2011; Pennebaker, 2011b; Pennebaker, Booth, & Francis, 2007). There are several reasons for why large-scale textual analyses may be relevant for the present purposes. For instance, the Internet provides access to large-scale data, and these data are so abundant that they are also found in rather specialized areas. Natural language analyses of anonymous social media forums also circumvent social desirability biases that may be present in traditional self-rating research, which is a particular important concern in relation to issues related to outgroups (Maass, Salvi, Arcuri, & Semin, 1989; von Hippel, Sekaquaptewa, & Vargas, 1997, 2008). The to-be analyzed media uses "aliases", yielding anonymity of the users and at the same time allow us to track individuals over time and analyze changes in communication patterns.

Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2007; Chung & Pennebaker, 2007; Pennebaker, 2011b; Pennebaker, Francis, & Booth, 2001) is a computerized text analysis program that computes a LIWC score, i.e., the percentage of various language categories relative to the number of total words (see also www.liwc.net). LIWC consists of dictionaries representing either content words or function in different categories. Function words include grammatical words such as pronouns and prepositions, and reflect the speaker's style more than content. Function words include LIWC categories such as *exclusive* words (e.g., but, without, exclude), *negations* (no, not, never), and *prepositions* (to, for). Content words carry meaning by definition and can reflect aspects such as positive (happiness, joy) and negative emotions (anger, fear); or cognition by dictionaries reflecting *insight* (understand, realize), or *cause* (because, cause).

In the present research, we use a Swedish translation of the LIWC 2007 dictionaries, and investigate changes in both function and content words. All analyses were conducted using the semanticexcel.com online tool for analyzing text, which is developed by our research lab. The LIWC dictionaries that we have included in this study focus on group identity and differentiation, emotions and cognitive complexity.

Group identity and differentiation dictionaries: Three function dictionaries including personal pronouns will be analyzed: 1) First person singular (I, me), 2) First person plural (we, us), and 3) third person plural (they, their).

Emotion dictionaries: Four content dictionaries will be analyzed: 1) negative emotions, 2) anger, 3) swear words, and, 4) sexually oriented words. Anger is a subset of the negative emotion words. The negative emotion dictionary includes negative feelings such as *depressed* or *scared*, but also generally negative words such as *fail* or *ugly*. The subset of anger words includes words such as *anger* and *hit*. In addition, because hostility toward other groups

can be expressed by swear words and sexually related words, these two dictionaries are included as predictors of emotional change.

Dictionaries related to cognitive complexity: In total, seven dictionaries will be used to assess cognitive complexity. Five dictionaries are related to function words: 1) prepositions (*over, to, about*), 2) articles (*a, an, the*), 3) inclusive word (*and, plus, with*), 4) exclusive words (*if, not, without*), 5) conjunctions (*until, when*). Two dictionaries are related to content words; 6) cause (*cause, force*), 7) insight (*belief, understand*).

## The Internet Forum Flashback

The forum investigated here is one of the largest Internet forums in Sweden, called Flashback (www.flashback.org). The forum claims to work for freedom of speech. It has over one million users who, in total, write 15 000 to 20 000 posts every day. It is often criticized for being extreme, for example in being too lenient regarding drug related posts but also for being hostile in allowing denigrating posts toward groups such as immigrants, Jews, Romas, and feminists. The forum has many sub-forums and we investigate one of these, which focuses on immigration issues.

The total text data from the sub-forum consists of 964 Megabytes. The total amount of data includes 700,000 posts from 11th of July, 2004 until 25th of April, 2015. Since the focus of the present research is to analyze change over time, we applied a set of selection criteria for users to be included in the analyses. First, posts used for LIWC-analyses should include more than 50 words (Petrie, Pennebaker, & Sivertsen, 2008). Second, because the study focuses on change and identification, participants should have been active during a six-month period, and should have posted at least 10 times during this period. This leaves us with 654,041 posts, 11,751 unique users

Empirical Results

Descriptive Results

Descriptive results are presented as word clouds in Figures 1a, 1b and 1c. Due to limitation in computational power, this analysis was based on a smaller subset of the data consisting of 5,000 posts. Figure 1a illustrates words overrepresented on Flashback as compared to natural Swedish language. It clearly shows that the content of the forum is related to immigration (e.g., Somalis, Muslims, Africans, immigration). The high representation of denigrating words (e.g., parasites, mass-migration, illiteracy) is most likely related to the ethnic groups such that it is these groups that are described with the denigrating labels. Figure 1b shows words that are overrepresented in posts written early on the forum. More precisely, it shows words that correlate negatively with how long the users have been active on the forum. Colored words have a significantly positive correlation between word frequency and the number of days since the first appearance. Some preliminary support for our hypotheses can be detected here, such as the over-representation of individual pronouns (I, me), and the use of swear words and degrading descriptions (i.e., shit, apes, devil, packs). Figure 1c shows words overrepresented in later posts, i.e. words where the usage of the words correlates positively with how long the users has been active on the forum. The words here typically lack emotional content and are indicators of higher complexity in language. Again, this analysis provides preliminary support for the idea that time on the forum is related to more complex thinking, and less emotionality.
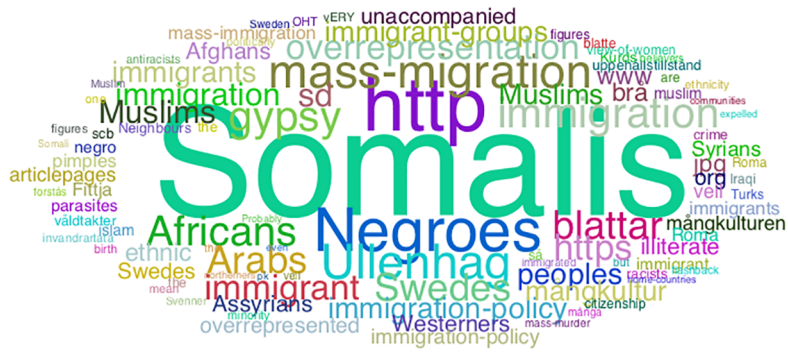
Figure 1a. Word cloud of words with a relative higher word frequency in the forum compared to the Swedish version of Google N-grams, using chi-square tests. Colored words are significant following Bonferroni correction for multiple comparisons, and grey words are significant at *p* < .05. Font size is proportional to the relative Chi-square value.

Figure 1b. Word cloud showing words with significant negative (indicative of early usage of the words) correlations between word frequency and time of writing on the forum for individual users (i.e., the logarithm of the number of days plus one since an individual user makes the first post on the forum). Colored words are significant following Bonferroni correction for multiple comparisons, and grey words are significant at *p* < .05. Font size is proportional to the relative correlation.

Figure 1c. Word cloud showing significant positive (indicative of late usage of the words) correlations respectively, between word frequency and time of writing on the forum for individual users (i.e., the logarithm of the number of days plus one since an individual user makes the first post on the forum). Colored words are significant following Bonferroni correction for multiple comparisons, and grey words are significant at *p* < .05. Font size is proportional to the relative correlation.

Correlation between LIWC dictionaries. Table 1 shows how the LIWC scores from different LIWC dictionaries correlate with each other. First, the pairwise correlation between all LIWC scores for each participant was calcu-

lated. Second, these correlations were averaged across participants. The results are summarized in Table 1, showing Pearson correlation coefficients for each pair that are significant following Bonferroni correction for multiple comparisons. Cells with non-significant correlations are omitted. As can be seen, many of the correlation coefficients are significant. For instance, all pairwise combinations of negative emotions correlate with each other. The measure for cognitive complexity yielded mixed results, where nine combinations correlate positively, and five negatively. As expected, the plural pronouns correlate positively with each other, whereas the plural and the singular forms correlate negatively with each other.

Table 1

Mean Correlations Over Participants Between all LIWC Categories and Pronoun Use, (n = 11751)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. 1st singular		-.062	-.060		.043	.020				-.010	.067		-.025	.062
2. 1st plural			.009	.027	.010		.024		.019	-.024		.240		-.013
3. 3rd plural					.034	.025	.027	.009	.065	-.065	.016	.043	.037	.018
4. Certainty					.013	.014	.017	.009	.031	-.011	.025	.023	.015	.013
5. Neg emotions						.276	.493	.122		-.034	.353	.016	.017	.040
6. Anger							.210	.277						
7. Swear words								.215	-.016		-.013	.041	.060	.068
8. Sexual										.009	-.012			
9. Prepositions										.030	.072	.473	.458	.342
10. Articles											-.028	-.013	.351	
11. Exclusion												-.075	-.022	.024
12. Inclusion													.225	.283
13. Causation														.476
14. Insight														

Note. The table shows mean values of Pearson correlation coefficient scores over participants (*N* = 11785) between pairwise LIWC scores. Mean values that are not significantly different from zero using two-tailed *t*-tests (*df* = 11784) are shown as empty cells. Bonferroni correction for multiple comparisons were used.

Main LIWC Analyses

The change of linguistic behavior over time was tested by analyzing correlations between LIWC scores for the posts and the number of days that a user had been on the forum. This correlation was first conducted for each user. The statistics was then calculated by testing the hypothesis whether the mean correlation values, from all selected users, was different from zero (except for the first hypothesis), using two-sided *t*-tests. Table 2 includes a summary of all dictionaries. Due to the increased risk for type I-error because of multiple significance testing, only *p*-values below .001 are considered significant.

The first hypothesis was concerned with group identity and intergroup differentiation as measured by pronoun use. Specifically, we expected that the use of 'I' would decrease over time and the use of 'we' and 'they' would increase relative to the overall pronoun use (which decreased over time, *r*(654041) = -.0360, *p* < .001)[i]. To evaluate the first hypothesis, we performed two-tailed *t*-tests assessing whether the correlation coefficients of the tested pronoun differed from the mean correlation of all measured pronouns. In support of the hypothesis, the correlation for first singular pronouns (i.e., 'I') decreased, whereas the mean correlation of first and third plural forms of pronouns

(e.g. 'we' and e.g., 'they' respectively) increased, relative to the mean correlation of all pronouns. The results are shown in Table 2.

Table 2

Summary of Tested LIWC Dictionaries Grouped Into the Categories 1) Group Identification and Differentiation, 2) Emotional Change, and 3) Cognitive Change, n = 11751

Categories	Dictionaries	Example words	Mean *r*
Group differentiation	First person singular	I, my, me	-.0103 ***[a]
	First person plural	We, our, us	.0115 ***[a]
	Third person plural	They, them, their	.0081 ***[a]
	Certainty	Absolutely, sure	.0016 NS
Emotion	Negative Emotion	Sad, loose	-.0117 ***
	Anger	Anger, hate	-.0121 ***
	Swear words	Hell, jackass	-.0002 NS
	Sexual words	Fuck, bitch	-.0083 ***
Cognitive complexity	Prepositions	If, on, of	.0074 ***
	Articles	The, an,	.0002 NS
	Exclusive words	But, except, exclude	-.0094 ***
	Inclusive words	And	.0014 NS
	Cause	Effect, because, why	.0017 NS
	Insight	Know, think	-.0021 NS

[a]Here we tested whether the correlation differed from the mean value of all pronouns. See the text for details.

***$p < .001$. NS = not significant.

The second hypothesis was that the linguistic style of new users would become increasingly similar to other users on the forum over time. This hypothesis is evaluated by first z-transforming each LIWC score, so that each has a mean value of zero and a standard deviation of one. Then we measure how each post differs from the standardized values by summing the absolute z-values over all 62 LIWC categories from 2007. Thus, low values on these deviation scores indicate that posts are more prototypical, or highly similar, to what other users write. These deviation scores are analyzed in the same way as for Hypothesis 1 (i.e., by correlating each user score with the number of days on the forum, and then *t*-testing whether the correlations are significantly different from zero). In support of the hypothesis, the results show an increase in similarity, as indicated by decreasing deviation scores (Figure 2). The mean correlation coefficient between this measure and time on the forum was -.0086, which is significant, $t(11749) = -3.77$, $p < 0.001$.

Hypotheses 3a and 3b related to changes in content. Specifically we expected that indicators of cognitive complexity should increase and indicators of negative emotions should decrease over time. The analysis to evaluate these hypotheses was conducted as described above. There was no clear pattern for the hypothesis regarding changes in cognitive complexity over time: prepositions increased significantly over time, whereas exclusion words decreased. No significant results were found for *articles, inclusive, conjunctions, causal*, and *insight* LIWC categories. However, results showed a decrease in negative emotion words as indicated by significant decreases in anger, swear words and sexual LIWC categories, which supports the hypothesis about emotional content (H3b). The results are presented in Table 1.
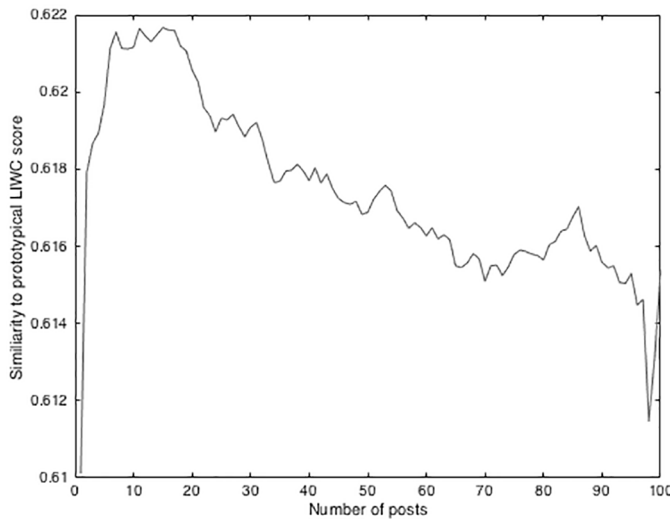
Figure 2. Deviance of posts to the content of forum as a function of the time spent on the forum, *n* = 11751.

Note. The y-axis represents deviations between posts presented at a certain time and all posts. Each LIWC score was first z-transformed, so that each LIWC score had mean of zero and standard deviation of one. Deviations from the normal scores are then calculated as the absolute values of summed over all LIWC scores. The data is smoothed over 20 time periods.

General Discussion

The present research has investigated language use on an online xenophobic forum. The aim was to investigate processes of identity formation, intergroup differentiation and possibly changes in emotional and cognitive content of the posts. To fulfil this aim, we traced individual users' posts over their first six months as members of the forum, and used a computerized text analysis technique called *Linguistic Inquiry Word Count* (Chung & Pennebaker, 2011; Pennebaker et al., 2007), which was adapted to a Swedish setting.

We expected and found changes in cues related to group identity formation and intergroup differentiation. Specifically, there was a significant decrease in the use of 'I' and a simultaneous increase in the use of 'we' and 'they'. This has previously been related to group identity formation and differentiation to one or more outgroups (Brewer & Gardner, 1996; Gustafsson Sendén et al., 2014b). Increased usage of plural, and decreased frequency of singular, nouns have also been found in both normal, and extremist, group formations (Seyle, 2007). There was a decrease in singular pronouns and a relative increase in collective pronouns. The increase in collective pronouns referred both to the ingroup (we) and to one or more outgroups (they). These results suggest a shift toward a collective identity among participants, and a stronger differentiation between the own group and the outgroup(s).

Because individuals form identities online and because we see this in the use of pronouns, we also expected to see tendencies of social influence and adaption. This effect was also found, such that individuals' linguistic style became increasingly similar to other users' linguistic style over time. Past research has shown that accommodation of communication style occurs automatically when people connect to people or groups they like (Giles & Ogay,

2007; Ireland et al., 2011), but also that similarity in communicative style functions as cohesive glue within a group (Reid, Giles, & Harwood, 2005).

The specific forum we investigated is characterized by discussions and exchange of ideas, which should be related to cognitive elaboration. In the terms of Karlsen et al. (2017), this particular forum should function as 'trench warfare', where arguments are countered by other participants. Still, the results could not confirm an increase in cognitive complexity. It is difficult to determine why this was not observed even though a general trend to conform to the linguistic style on the forum was observed. A possible explanation is that the linguistic technique for analyzing posts requires that the posts are of a certain length. This biases the sample of posts available for analysis already from the beginning. It is plausible that longer posts also contain more cognitive complexity leading to a kind of 'ceiling effect' for these LIWC categories.

Because the political opinions of xenophobia are stigmatized in society, we expected that individuals when joining the forum may be frustrated and feel suppressed. This could lead to an initial outburst of anger related content in the posts at the beginning of an individual's time on the forum. In relation, we found that anger words were more prevalent in the beginning and decreased over time. This idea would also be in line with previous research that has shown that expressing oneself decreases arousal (Garcia et al., 2016). Moreover, because the forum is not explicitly racist, individuals may have simply adapted to the social norms on the forum prescribing less negative emotional displays. Finally, a possible explanation for the decrease in negative emotional words might be that users who are very angry leave the forum, because of its non-racist focus, and end up in more hostile forums. An interesting finding that was not part of the hypotheses in the present research is that the third person plural category correlated positively with all four negative emotions categories, suggesting that people using for example 'they' express more negative emotions (see Table 1).

Limitations and Future Research

As a first attempt at understanding the interactions taking place in online forums, there are some limitations to the present research that are important to note. For instance, the criteria to select which users to analyze can be discussed. Because the focus here was changes occurring over time, members who only posted a few posts where not included. Moreover, for the LIWC-analysis to be successful there must be a minimum number of words, which means that short replies and interactions were not included. This may have biased the sample of posts in favor of more complex posts.

Another limitation is that linguistic behavior cannot be correlated with explicit attitudes to the ingroup or the outgroup. This is a consequence of using natural language behavior and online forums. Knowing how explicit attitudes toward immigration and immigration groups change over time would add information to how the change in linguistic behavior should be understood.

Conclusions

Our study highlights the importance of analyzing open online milieus when examining the processes of identity formation, social adaption and intergroup differentiation. The results indicate that forum participants form groups online and start to differentiate between their own group and other group(s). In line with social identity theory (Tajfel & Turner, 1986), we also observe linguistic adaption to the group. Hence, our results indicate that processes of identity formation may take place online.

Notes

i) Due to that the overall use of pronouns decreased over time, pronouns are not significant in the word clouds of late use, Figure 1c.

Competing Interests

The authors have declared that no competing interests exist.

References

Allport, G. (1954). *The nature of prejudice.* Cambridge, MA, USA: Addison-Wesley.

Bäck, E. A., & Bäck, H. (2017). Radikaliseringens politiska psykologi – Betydelsen av hot och polarisering. In C. Edling & A. Rostami (Eds.), *Våldsbejakande extremism: En forskarantologi* (Governmental Public Investigations 2017:67). Retrieved from http://www.sou.gov.se/wp-content/uploads/2014/10/SOU-2017_67_webb.pdf

Bäck, E. A., Esaiasson, P., Gilljam, M., Svenson, O., & Lindholm, T. (2011). Post-decision consolidation in large group decision-making. *Scandinavian Journal of Psychology, 52*, 320-328. doi:10.1111/j.1467-9450.2011.00878.x

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497-529. doi:10.1037/0033-2909.117.3.497

Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology, 29*, 166-202. doi:10.1006/jesp.1993.1008

Brewer, M. B. (1991). The social self – On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*, 475-482.

Brewer, M. B., & Gardner, W. L. (1996). Who is this "we"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*, 83-93. doi:10.1037/0022-3514.71.1.83

Brundidge, J. (2010). Encountering "difference" in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks. *Journal of Communication, 60*, 680-700. doi:10.1111/j.1460-2466.2010.01509.x

Caiani, M., & Parenti, L. (2009). The dark side of the web: Italian right-wing extremist groups and the Internet. *South European Society & Politics, 14*, 273-294. doi:10.1080/13608740903342491

Campbell, R. S., & Pennebaker, J. W. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science, 14*, 60-65. doi:10.1111/1467-9280.01419

Cartwright, D., & Zander, A. F. (1968). *Group dynamics: Research and theory*. New York, NY, USA: Harper & Row.

PsychOpen GOLD

Chung, C. K., & Pennebaker, J. W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343-359). New York, NY, USA: Psychology Press.

Chung, C. K., & Pennebaker, J. W. (2011). Using computerized text analysis to assess threatening communications and behavior. In National Research Council, Division of Behavioral and Social Sciences and Education, Board on Behavioral, Cognitive, and Sensory Sciences, & C. Chauvin (Eds.), *Threatening communications and behavior: Perspectives on the pursuit of public figures*. Washington, DC, USA: The National Academies Press.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology, 51*, 629-636. doi:10.1037/h0046408

Dino, A., Reysen, S., & Branscombe, N. R. (2009). Online interactions between group members who differ in status. *Journal of Language and Social Psychology, 28*, 85-93. doi:10.1177/0261927X08325916

Duggan, M., & Smith, A. (2013, December). *Social media update 2013* (Pew Research Center report). Retrieved from http://www.pewinternet.org/2013/12/30/social-media-update-2013

Edwards, C., & Gribbon, L. (2013). Pathways to violent extremism in the digital era. *The RUSI Journal, 158*, 40-47. doi:10.1080/03071847.2013.847714

Enli, G. S., & Thumim, N. (2012). Socializing and self-representation online: Exploring Facebook. *Observatorio Journal, 6*, 87-105.

Fernández, I., Igartua, J., Moral, F. E. P., Acosta, T., & Muñoz, D. (2013). Language use depending on news frame and immigrant origin. *International Journal of Psychology, 48*, 772-784. doi:10.1080/00207594.2012.723803

Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science, 3*, 38-47. doi:10.1111/j.1745-6916.2008.00060.x

Fitzsimons, G. M., & Kay, A. C. (2004). Language and interpersonal cognition: Causal effects of variations in pronoun usage on perceptions of closeness. *Personality and Social Psychology Bulletin, 30*, 547-557. doi:10.1177/0146167203262852

Frank, M. G., & Gilovich, T. (1988). The dark side of self and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology, 54*, 74-85. doi:10.1037/0022-3514.54.1.74

Ganda, M. (2014). *Social media and self: Influences on the formation of identity and understanding of self through social networking sites* (University Honors thesis, Portland State University, Portland, Oregon, USA). Retrieved from http://archives.pdx.edu/ds/psu/11971

Garcia, D., Kappas, A., Küster, D., & Schweitzer, F. (2016). The dynamics of emotions in online interaction. *Royal Society Open Science, 3*, Article 160059. doi:10.1098/rsos.160059

Garrett, R. K. (2009). Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication, 14*, 265-285. doi:10.1111/j.1083-6101.2009.01440.x

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293-310). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*, 472-482. doi:10.1086/586910

Gustafsson Sendén, M., Lindholm, T., & Sikström, S. (2014a). Biases in news media as reflected by personal pronouns in evaluative contexts. *Social Psychology, 45*, 103-111. doi:10.1027/1864-9335/a000165

Gustafsson Sendén, M., Lindholm, T., & Sikström, S. (2014b). Selection bias in choice of words: Evaluations of "I" and "we" differ between contexts, but "they" are always worse. *Journal of Language and Social Psychology, 33*, 49-67. doi:10.1177/0261927X13495856

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science, 22*, 39-44. doi:10.1177/0956797610392928

Jamieson, K. H., & Capella, J. N. (2009). *Echo chamber.* Oxford, United Kingdom: Oxford University Press.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2013). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology, 33*, 125-143. doi:10.1177/0261927X13502654

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication, 32*, 257-273. doi:10.1177/0267323117695734

Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *The Journal of Conflict Resolution, 2*, 51-60. doi:10.1177/002200275800200106

Kenworthy, J. B., & Miller, N. (2002). Attributional biases about the origins of attitudes: Externality, emotionality and rationality. *Journal of Personality and Social Psychology, 82*, 693-707. doi:10.1037/0022-3514.82.5.693

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts - The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*, 981-993. doi:10.1037/0022-3514.57.6.981

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology, 79*, 602-616. doi:10.1037/0022-3514.79.4.602

Miller, D. T., & Prentice, D. A. (1996). The construction of social norms and standards. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 799-829). New York, NY, USA: Guildford Press.

Mudde, C. (2007). *The populist radical right in Europe.* Cambridge, United Kingdom: Cambridge University Press.

Neumann, P. R. (2013). The trouble with radicalization. *International Affairs, 89*, 873-893. doi:10.1111/1468-2346.12049

Owens, L., & Palmer, L. K. (2003). Making the news: Anarchist counter-public relations on the world wide web. *Critical Studies in Media Communication, 20*, 335-361. doi:10.1080/0739318032000142007

Pennebaker, J. W. (2011a). *The secret life of pronouns: What our words say about us*. New York, NY, USA: Bloomsbury.

Pennebaker, J. W. (2011b). Using computer analyses to identify language style and aggressive intent: The secret life of function words. *Dynamics of Asymmetric Conflict, 4*, 92-102. doi:10.1080/17467586.2011.627932

Pennebaker, J. W., Booth, R. S., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC 2007*. Austin, TX, USA: LIWC.

Pennebaker, J. W., & Chung, C. K. (2008). Computerized text analysis of Al-Qaeda statements. In K. Krippendorff & M. Bock (Eds.), *A content analysis reader* (pp. 453-466). Thousand Oaks, CA, USA: Sage.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC): A text analysis program*. New York, NY, USA: Erlbaum.

Petrie, K. J., Pennebaker, J. W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 197-202. doi:10.1037/a0013117

PsychOpen GOLD

Reid, S. A., Giles, H., & Harwood, J. (2005). A self-categorization perspective on communication and intergroup relations. In J. Harwood & H. Giles (Eds.), *Intergroup communication: Multiple perspectives* (pp. 241-263). New York, NY, USA: Peter Lang.

Seyle, D. C. (2007). *Identity fusion and the psychology of political extremism* (Doctoral thesis, The University of Texas at Austin, Austin, Texas). Retrieved from http://hdl.handle.net/2152/3027

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. V. (1961). *Intergroup conflict and cooperation: The robbers cave experiment*. Norman, OK, USA: University of Oklahoma Book Exchange.

Simon, B., Hastedt, C., & Aufderheide, B. (1997). When self-categorization makes sense: The role of meaningful social categorization in minority and majority members' self-perception. *Journal of Personality and Social Psychology, 73*, 310-320. doi:10.1037/0022-3514.73.2.310

Slatcher, R. B., Chung, C. K., Pennebaker, J. W., & Stone, L. D. (2007). Winning words: Individual differences in linguistic style among U.S. presidential and vice presidential candidates. *Journal of Research in Personality, 41*, 63-75. doi:10.1016/j.jrp.2006.01.006

Stanley, B. (2008). The thin ideology of populism. *Journal of Political Ideologies, 13*, 95-110. doi:10.1080/13569310701822289

Sunstein, C. (2001). *Republic.com.* Princeton, NJ, USA: Princeton University Press.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755-769. doi:10.1111/j.1540-5907.2006.00214.x

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*, 96-102. doi:10.1038/scientificamerican1170-96

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA, USA: Brooks-Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago, IL, USA: Nelson-Hall.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24-54. doi:10.1177/0261927X09351676

Thompson, R. (2011). Radicalization and the use of social media. *Journal of Strategic Security, 4*, 167-190. doi:10.5038/1944-0472.4.4.8

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, United Kingdom: Basil Blackwell.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33*, 490-509. doi:10.1006/jesp.1997.1332

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (2008). Linguistic markers of implicit attitudes. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 429-458). New York, NY, USA: Psychology Press.

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58*, 425-452. doi:10.1146/annurev.psych.58.110405.085641