NORDIS

NORDIS – NORdic observatory for digital media and information DISorders

Task 2.3 Emotional landscapes of misinformation spread

Date: 28-02-2023

Final version

Action No:	2020-EU-IA-0189
Project Acronym:	NORDIS
Project title:	NORdic observatory for digital media and information DISorders
Start date of the project:	01/09/2021
Duration of the project:	24
Project website address:	https://datalab.au.dk/nordis
Authors of the deliverable	Marina Charquero-Ballester Jessica Gabriele Walter Astrid Sletten Rybner Ida A. Nissen Kenneth Christian Enevoldsen Anja Bechmann
Activity number	Activity 2
Task	Task 2.3

Reviewers:
Mervi Pantti – University of Helsinki
Minna Aslama Horowitz – University of Helsinki
Matteo Magnani – Uppsala University

Funding:
The research was funded by EU CEF grant number 2394203.

NORDIS

1.0 Executive Summary

In February 2020, only a month after the Covid-19 pandemic was declared, the World Health Organization (WHO) announced that the crisis was accompanied by an 'infodemic' of misinformation (WHO 2020). Whereas previous public health crisis also affected millions, constant media coverage regarding COVID-19 and extended use of social media turned this scenario into an unprecedented situation with two co-occurring but different types of crises: Covid-19 and an aggravated misinformation crisis. When investigating crises in a context of resilient democracies, as is the case of the Nordic countries, questions around collective action arise. In particular, it has been argued that democracy poses a problem to collective action, since for every individual citizen, the cost of productive political engagement often outweighs the additional policy benefits to be gained from such behaviour. In the case of the Nordic countries, this effect might be further strengthened by their comprehensive healthcare systems, high levels of education and high levels of trust in organizations and the media. Nevertheless, situations in which a threat arises, might break the resistance to step out of the individual comfort and motivate citizens to organise for collective action. How does that happen? A strong body of research supports the importance of emotions in that context.

In this study, we use Twitter data from the Nordic countries during the second wave of the pandemic and select tweets containing hashtags pertaining to one of each of the following categories: misinformation-related hashtags, highly used Covid-19 hashtags and highly used general hashtags. We opted for this approach based on previous work arguing for an additional function of hashtags to act as linguistic markers indicating the target of the appraisal in the tweet, that is, who is the user addressing to with their interpretation of a situation, going beyond their organizational and topic-making function. The use of hashtags has been argued by Zapavigna to have an additional function to upscale the call to affiliate with the values - or share the emotions - expressed in the tweets, which can be further associated with states of action readiness potentially leading to collective action.

Which emotions appear more predominant in tweet appraisals referring to crises in the Nordic Welfare system?

Our study shows that fear and sadness appeared consistently higher in both types of crisis appraisals (Covid-19 and misinformation) across all the Nordic countries taking part in this study. The only exception was Denmark, in which we did not find any significant differences in the amount of fear expressed in the Covid-19 vs. general appraisals. While the effect was present in both types of crisis appraisals, shown as significant results in both cases, the effect sizes were large for the comparison between misinformation and general tweet appraisals, but only small for the comparison between Covid-19 and general tweet appraisals. In other words, we do not see a large increase in the use of fear and sadness expression in appraisals that refer to the Covid-19 crisis. In relation to collective action, these two emotions often have an inhibiting effect, which – if we were to take these results in isolation - would be stronger in the misinformation crisis.

In regards to the expression of anger and disgust in appraisals, we found a significant and very large effect when comparing the misinformation vs. the general tweet appraisals in all Nordic countries. This is interesting because anger, and the related emotion disgust, are strong predictors of collective action. As such the data would indicate a very fertile context for collective action around misinformation. In the case of Covid-19 vs. general tweet appraisals, results were not as consistent across countries, and the effect size was small even in cases where significant differences were present. In particular, only Swedish and Norwegian tweets had higher expression of anger and disgust in Covid-19 tweet appraisals, whereas the oppositive effect was found for Danish and Finnish tweets. As such, we would expect anger to act as a potential driving force for collective action only in the misinformation crisis. However, the accompanying high expression of fear and sadness might play an attenuating effect.

Before diving into the interaction of anger, fear and sadness, and the potential for collective action when taken into combination, it is worth mentioning that the presence of joy and optimism was consistently higher in the non-crisis appraisals, both Covid-19 and misinformation appraisals. However, as with the case of negative emotions, the effect sizes were much larger when comparing misinformation vs. general appraisals than when comparing Covid-19 vs. general appraisals, indicating that the emotional landscape of misinformation appraisal tweets differs strongly from the general emotional landscape in all its components.

How do anger, fear and sadness interact in crisis appraisals?

Given the importance of the interaction between negative emotions for collective action, we further investigated the relationship between all pairs of negative emotion expression: Anger-fear, anger-sadness and sadness-fear. In theory, collective action can be facilitated by high levels of anger and low levels of fear and sadness. As such, a change in the ratio between anger and the other two emotions, where fear and sadness become more prominent, might indicate decreased chances for collective action. This is what we found in the Covid-19 vs. general tweet appraisals. On the other hand, a change in the ratio in which fear and sadness become less prominent, leaving anger dominate, might lead to increased chances for collective action. We found a trend towards this effect in the misinformation vs. general appraisals. Further research is needed to investigate how this quantifies in online and offline collective action.

Overall, we found differences in emotional expression when comparing two different types of crisis oriented tweet appraisals in the Nordic Twittersphere. Taking into consideration the Nordic context, with their resilient democracies and high trust societies, emotions have been suggested to be particularly important in organising for collective action (Groenedyck, 2011). This study suggests that the misinformation crisis would be more likely to present a fertile environment for collective action in the Nordic countries than the co-occurring crisis around Covid-19. However, more research is needed to investigate the degree to which this translates into collective action both in online and offline behaviour.

2.0 Introduction

In February 2020, only a month after the Covid-19 pandemic was declared, the World Health Organization (WHO) announced that the crisis was accompanied by an 'infodemic' of misinformation (WHO 2020). While previous public health crisis also affected millions, constant media coverage regarding COVID-19 and extended use of social media turned this scenario into an unprecedented situation with two co-occurring, but different types, of crises. We refer to misinformationa as a crisis, as misinformation reaches more people and spreads further than true information (Vosoughi et al. 2018) and is potentially harmful in various aspects such as by affecting health-protective behavior (Allington et al. 2020). While the two crises had an impact worldwide, here we focus on four of the Nordic countries (Denmark, Finland, Norway and Sweden), referred to as the Nordic countries in the remainder of this text for simplicity purposes. The reason for doing so is that the Nordic countries share characteristics such as healthy democracies (Transparency International, 2019), comprehensive healthcare systems (OECD, 2021b), high levels of education (OECD, 2021a) and high levels of trust in organizations and the media (Delhey & Newton, 2005; OECD, 2020). It has been argued that democracy poses a problem to collective action, since for every individual citizen, the cost of productive political engagement often outweighs the additional policy benefits to be gained from such behavior (Groenendyk, 2011). In this case, the comprehensive healthcare systems and high levels of education and trust in the media may accentuate even further the individual cost for political engagement. Nevertheless, even in the scenario of a strong, resilient democracy, some situations break this resistance and motivate citizens to disengage from the individual comfort and initiate collective action. In this respect, a strong body of research supports the importance of emotions in that context (e.g., van Zomeren for review). Modern psychological theories suggests that emotions perform a key role in human functioning (Lazarus, 1991; Scherer et al., 2001, van Zomeren et al., 2012). The cognitive appraisal of the environment leads to the experience of concrete emotions (e.g., anger or fear), which can be further associated with states of action readiness potentially leading to collective action.

Signatures of public emotions have been shown to be present throughout social media platforms, such as Twitter data originating from different countries (Xue et al., 2020; Zhang et al., 2022). Here we take the creation of content on Twitter around the two crises (i.e., Covid-19 and misinformation) as manifestation of the cognitive appraisal, which further carries the emotion and potential for collective action. We build this from Zappavigna's work on hashtags use for appraisals. Zappavigna (2011) argues that, in addition to its topic-marking function, Twitter hashtags acts as a linguistic marker indicating the target of the appraisal in the tweet, which can be used to upscale the call to affiliate with the values - or share the emotions - expressed in the tweet. We regard this affiliation as the basis for collective action, with the creation of content around these two issues already consisting of a form of online collective action (e.g., Lundgaard & Razmerita). In particular we focus on emotions expressed in three different sets of tweets grouped by the hashtags they contain: one set consisting of hashtags related to Covid-19, one set consisting of hashtags related to mis- and disinformation, and one set consisting of highly used, non-Covid-19 related (i.e., general) hashtags.

We refer to emotional landscapes since we set no restrictions on the primary emotions evaluated. Our initial model was trained on 11 emotions, from which we selected those that the model showed to be able to reliably estimate from our data (see Methods), 'painting a landscape' of the emotional expression on Twitter. We then visualise these emotions using violin plots, which allow us to gain some insights into the distribution of the emotion expression in each subsample of tweets. If we were to overlap all these violin plots, the figure would resemble the mountains and valleys of a landscape, with accumulation of tweets with similar probability scores for a given emotion creating 'mountains' of different sizes.

We chose to conduct our study using a large Twitter dataset collected between August 2020 and March 2021, that was then further filtered down using the hashtags as specified in the Methods section. We then investigate quantitative differences between the general and the two crisis-oriented pools of tweets (i.e., Covid-19 and misinformation). We also investigate the relationship between negative emotions within and across languages, and how this relationship differs in the different appraisal contexts. Building up on the results from our previous report on the dominant public emotions present in this time period, we go a step further, zooming into crisis appraisals taking two examples, Covid-19 and misinformation. By comparing the two crisis and establishing any differences in emotion between them, we aim to gain a deeper understanding of the role that emotions play in organising for collective action online in the context of highly resilient and democratic systems.

3.0 Materials and methods

We used the XLM-RoBERTa base model proposed by Conneau et al., (2019). It is a large multi-lingual language model, trained on 2.5TB of filtered CommonCrawl data and based on the initial RoBERTa model released in 2019 (arXiv:1907.11692v1). Many of the previous studies studying public emotions on Twitter from a language psychology perspective have used Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015). While this tool has the advantage of being simple and scalable, it often detects false signals as words can appear in different contexts and take on different meanings. Furthermore, the latest version (LIWC 2015) has not been translated into all languages present in this study and, as previous research has shown, within-language standardization is needed when analyzing texts from different languages (Dudău & Sava, 2021). Applying the LIWC in this dataset would have required translation of the entire dataset, with the consequent loss of information inherent to automatic translation of such short and context-dependent excerpts of text. Therefore, instead of using a lexicon-based approach, we decided to use the XLM-RoBERTa-base model (Conneau et al., 2020). This model is based on a deep Bidirectional Encoder Representation from Transformers (BERT) model (Devlin et al., 2019). The model shares the advantages of traditional transformer models, which use attention mechanisms to extract relational context and even long-range dependencies in a sentence. The availability of pretrained multilingual models, as is the case of XLM-RoBERTa, allowed us to use the same model to analyse tweets in 4 different languages. Additionally, we initially opted for a sampling method aimed at collecting as many tweets as possible without any hashtag or covid-related keyword specifications (see Methods), allowing for an overall sample representative of the Nordic Twitter, as well as the possibility to filter it down to create subsamples comparable to previous studies (refs).

Training data

We used the SemEval 2018, task 1 (Mohammad & Bravo-Marquez, 2017) as the training dataset to finetune the model for emotion detection. This dataset was manually annotated through a crowdsourcing project, with each tweet being annotated by, on average, 7 annotators (for further details, see Mohammad and Bravo-Marquez, 2017). The possible emotions were the following: 1) anger (also includes annoyance, rage) 2) anticipation (also includes interest, vigilance) 3) disgust (also includes disinterest, dislike, loathing) 4) fear (also includes apprehension, anxiety, terror) 5) joy (also includes serenity, ecstasy) 6) love (also includes affection) 7) optimism (also includes hopefulness, confidence) 8) pessimism (also includes cynicism, no confidence) 9) sadness (also includes pensiveness, grief) 10) surprise (also includes distraction, amazement) 11) trust (also includes acceptance, liking, admiration) 12) neutral or no emotion. Each of the 11 emotions were considered whether present in each of the tweets, resulting in a multilabel training dataset in which each tweet can contribute to multiple emotions. Since we needed the training dataset to be a available in the four Nordic languages of this study, we translated the English version into all 4 Nordic languages using the eTranslation tool provided by the European Commision (https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranslation_en). The English version of the dataset consists of 10,983 tweets in total, 6,838 for the training, 886 for the validation and 3,259 for the test set. The training set, used for the finetuning of the model, contains a total of 160 anxiety tweets (i.e., tweets labeled as both expressing anticipation and fear). Accordingly, we expect the fine-tuned xlm-Roberta model to be able to detect anxiety if present in our Twitter dataset.

Twitter data

In this study we analysed a large Twitter dataset containing tweets in four of the Nordic languages: Danish, Norwegian, Swedish and Finnish. The tweets in Danish, Norwegian, and Swedish were collected through the HOPE project (https://hope-project.dk/#/) using a set of stopwords in each of the languages to scrape through the Twitter API before the language tag was available. The stopwords were sourced from the Open Subtitles website (http://www.opensubtitles.org), and generated by selecting the top 100 most frequent words in the four Nordic languages of this study, full list available at the following github repository: https://github.com/centre-for-humanities-computing/stopwords-danish-distinct. For each language, words were cross-checked with the lists from any of the other Nordic languages lists relevant for this study and removed if the word was present across languages. The reason for doing so was to create a list of frequently used words in each of the languages that was still as differentiated as possible from the other Nordic languages. In the case of Norwegian, additional Nynorsk words were added to the Bokmal stopwords list after checking for their absence in the other languages. The tweets in Finnish were collected using the equivalent set of stopwords in that language. Our dataset includes tweets that were posted between August 2020 and March 2021. The Twitter dataset contained a total of 57.828.980 tweets (29.088.137 in Swedish, 6.168.893 in Norwegian, 7.369.613 in Danish, and 15.202.337 in Finnish).

Data analysis

First, we fine-tuned the XLM-RoBERTa base for emotion detection using a learning rate of 2e5 as in the original publication (Devlin et al., 2019)**.** We used a batch size of 16 and fine-tuned over four epochs, but set an early stopping callback based on the evaluation loss, a strategy commonly used to prevent overfitting of the model. For evaluation of the fine tuning on the validation set, a threshold of 0.5 was set above which a tweet was considered to have been assigned to a specific emotion category according to the model. We then evaluated the performance of the model on the validation set focusing primarily on the f1-score, choosing this metric due to our imbalanced dataset. The f1-score combines the precision and the recall metrics - the two most common metrics that take into account class imbalance - into a single metric. In our case, this was important given the unequal amount of different emotions that can be found on our Twitter training dataset used for the emotion detection fine-tuning. Performance metrics after the finetuning can be found in the supplementary material (**Supplementary figure 1**). As shown in the Supplementary material, only 5 out of the 11 emotions obtained an f1-score above 0.60, which we set as the threshold above which we consider the model to be able to reliably detect the emotion. These were anger, disgust, fear, joy, optimism and sadness.

Before applying the fine-tuned model, we preprocessed the Twitter dataset by removing mentions and URLs. We then decoded emojis present in the text using Demoji 1.1.0 (https://github.com/bsolomon1124/demoji). After the text preprocessing was done, we estimated the probabilities for each emotion to be present in each tweet. Tweets with all emotional probabilities below 0.5 were considered neutral tweets and excluded from further analysis. For each tweet, we selected the highest probability across the 6 emotions that the model is able to reliably estimate, and set this as the main emotion expressed in the tweet.

We separated original tweets from retweets. In order to create a general and a Covid19 subsample of tweets, we computed hashtag counts to determine which were the most used hashtags in our dataset.

The 17 most often used Covid-related and 17 most-often used covid-unrelated hashtags were used to select two subsamples of tweets. Both subsamples were the same size with 21869 tweets.

We additionally used the misinformation-related hashtags (#fakenews #misinformation #falskanyheter #feilinformasjon #falskenyheter #vääräätietoa #valeuutisia) to create the third subsample. We used both the terms fake news and misinformation translated to the nordic languages in addition to the english terms. Of note, the definition for disinformation by Buning et al. (2018) refers to "false, inaccurate or misleading information designed, presented and promoted to intentionally cause public harm or for profit". This term is interchangeable with the term "fake news" or "rumor" that have been carefully conceptualized in different studies (e.g. Bechmann and O'Loughlin 2020, Kalsnes 2018, Tandoc et al. 2018). Additionally we use the term misinformation, in contrast to the term disinformation defined by Buning et al., (2018), which makes no assumptions about the intention. As such we aim for a more comprehensive subsample of tweets, possibly containing a wider range of expressed emotion, while still focused on information disorders. The tweet extraction approach in this case did not include the selection through most used hashtags because the prevalence of misinformation related hashtags in Nordic Twitter is very low. Consequently, the amount of misinformation-related tweets was lower, in particular: Sweden (579), Norway (190), Denmark (354), Finland (694).

4.0 Results

The management response to the Covid-19 crisis varied widely across these countries, with some of them adopting more restrictive measures, specially in the early stages of the pandemic, (e.g., Denmark) and others being particularly lax in the implementation of such restrictive measures (e.g., Sweden).

As detailed in the *Methods* section, evaluation of the performance of the XLM-RoBERTa model on the validation set revealed high variability in the f1-scores used to assess the ability of the model to detect each individual emotion. We decided to focus only on those emotions that the model could detect best by setting a threshold at f1 > 0.60, above which we consider that the model can reliably detect a particular emotion in a tweet (see Supplementary **table 1**). Emotions with an f1-score of 0.60 or higher were, in descending order, joy, anger, disgust, optimism, fear and sadness.

In order to gain a deeper understanding of emotions in crisis oriented vs. non-crisis oriented emotion discourses, we look at three subsamples of tweets; one sampled through Covid-related hashtags, one sampled through misinformation-related hashtags and one sampled through general hashtags. On a descriptive level, we can observe a difference between the distribution of the emotion probabilities in the misinformation oriented and the other two subsamples, specifically for the negative emotions (i.e., fear, sadness, disgust and anger) **(Figure 1).** This difference seems to be characterized by a concentration of the probabilities around the middle range for fear and sadness and around the higher range for disgust and anger in the misinformation subsample in contrast to the other two subsamples. Probability distributions of positive emotions such as joy and optimism were not as markedly different across the three subsamples, although we still observe a more even distribution of high and low probabilities in the general and Covid-19 subsamples than in the misinformation subsample.

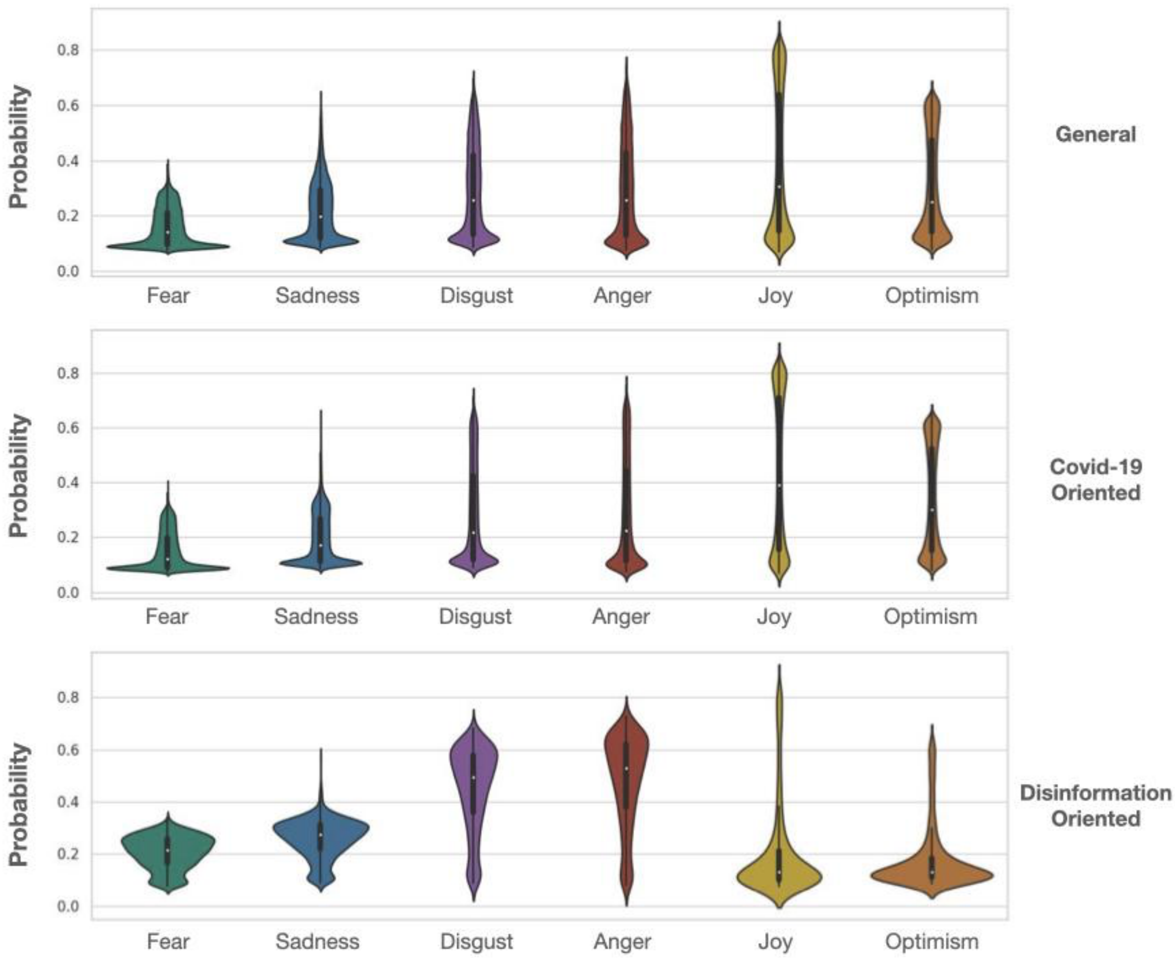


Figure 1. Distribution of the emotion probabilities in the three subsamples. Starkest differences found in the misinformation-oriented subsample in comparison to the General and Covid-19 oriented ones, especially around negative emotions (i.e., fear, sadness, disgust and anger)

We then compare each of the crisis-oriented subsamples (Covid-19 and misinformation) against the non-crisis oriented (general) subsample.

In the Covid-19 vs. general subsample, the most consistent finding across countries concerns sadness, which had higher probability scores in the crisis-oriented subsample for all countries (Danish p=0.0002; Swedish p= 0.0002; Norwegian p=0.0002, Finnish p=0.0002) **(Figure 2).** The probability scores for fear were also consistently higher in the crisis-oriented subsample (Swedish p= 0.0002; Norwegian p=0.0002, Finnish p=0.0002), with the exception of the Danish data (p=0.2458), which showed no significant differences between crisis-oriented and non-crisis oriented tweets for fear. Joy and optimisms showed the opposite, yet consistent patterns across countries, where the non crisis oriented subsample showed higher probability scores (Joy: Swedish p= 0.0002; Norwegian p=0.0002, Finnish p=0.0002) (Optimism: Swedish p= 0.0002; Norwegian p=0.0002, Finnish p=0.0002), with the exception of Denmark (Joy p=0.4866; Optimism p = 0.5128). The less consistent findings across countries concern anger and disgust. Here we observe that Norway and Sweden have significantly higher probability scores for anger and disgust in the crisis-oriented subsample (Anger: Norwegian p= 0.0002, Swedish p = 0.0002 ; Disgust: Norwegian p= 0.0002 , Swedish p = 0.0002), whereas Denmark and Finland have higher probability scores in the non crisis-oriented sample (Anger: Danish p= 0.0002, Finnish p = 0.0002 ; Disgust: Danish p= 0.0002 , Finnish p = 0.0002).

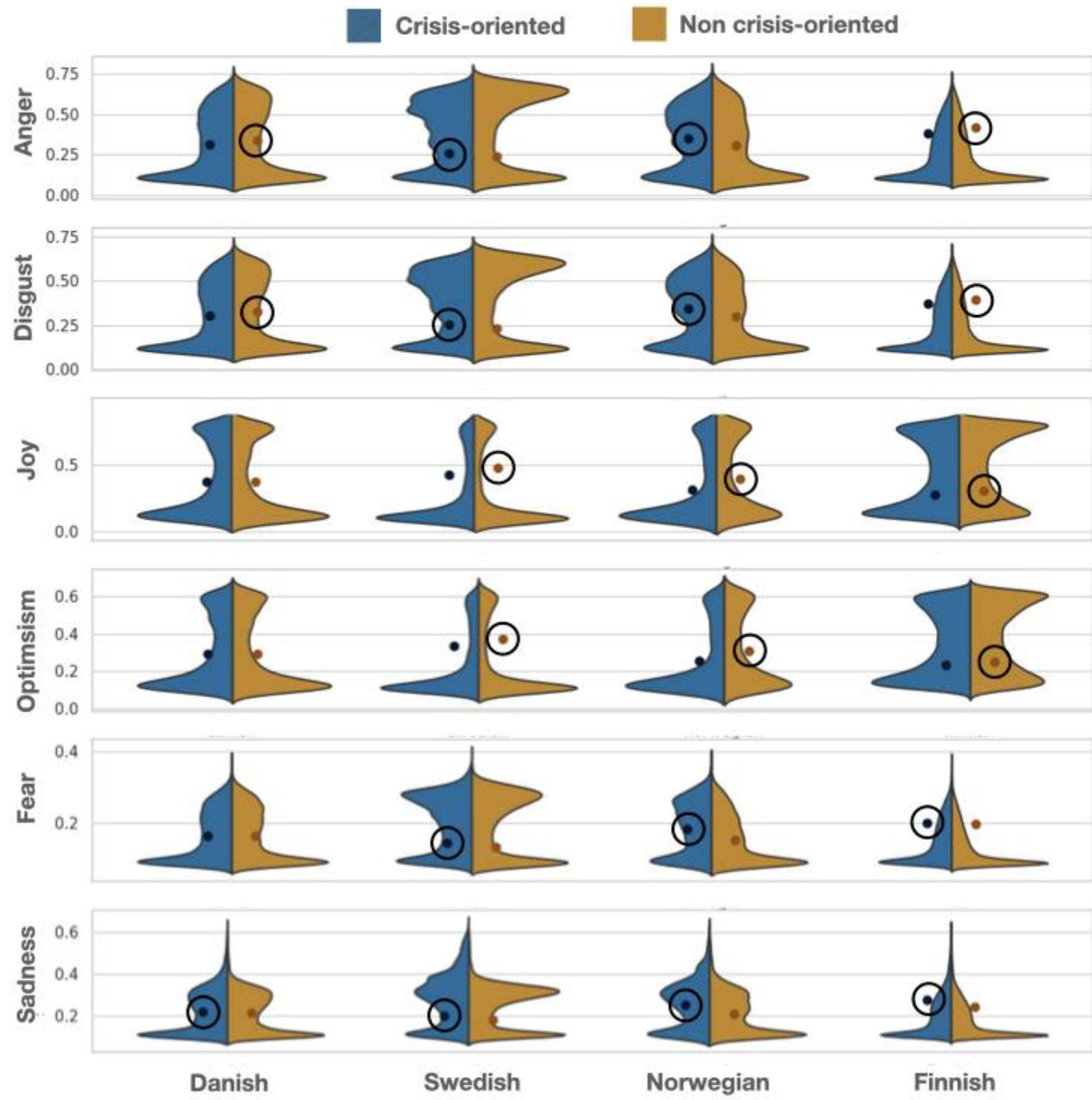


Figure 2. Sadness and fear are consistently higher in the crisis-oriented subsamples, while optimism and joy are consistently lower in the crisis-oriented subsamples. Anger and disgust were higher in the crisis-oriented sample in Sweden and Norway, and lower in Denmark and Finland.

	Country	Emotion	Mean(Non crisis)	Mean(Crisis)	SD(Non crisis)	SD(Crisis)	P−value	Standardized Effectsize(Cohen's d)
1	DA	Anger	0.336619	0.313374	0.201242	0.183158	0.0002	0.120590
2	DA	Disgust	0.324223	0.305722	0.184123	0.168720	0.0002	0.104628
3	DA	Joy	0.369570	0.371270	0.265441	0.257794	0.4866	0.006498
4	DA	Optimism	0.293681	0.292535	0.184960	0.178926	0.5128	0.006295
5	DA	Sadness	0.212526	0.219266	0.092137	0.100213	0.0002	0.069981
6	DA	Fear	0.163265	0.162500	0.069447	0.068075	0.2458	0.011113
7	FI	Anger	0.237412	0.255778	0.154021	0.149636	0.0002	0.120735
8	FI	Disgust	0.232640	0.253086	0.140352	0.137834	0.0002	0.146598
9	FI	Joy	0.479348	0.424479	0.254596	0.246875	0.0002	0.217509
10	FI	Optimism	0.370278	0.335560	0.178156	0.173103	0.0002	0.196699
11	FI	Sadness	0.177904	0.199355	0.085214	0.095615	0.0002	0.235213
12	FI	Fear	0.130387	0.144025	0.054110	0.057960	0.0002	0.241474
13	SV	Anger	0.417055	0.382023	0.225188	0.188291	0.0002	0.168184
14	SV	Disgust	0.397367	0.373099	0.206096	0.173980	0.0002	0.126990
15	SV	Joy	0.304659	0.278426	0.269214	0.235061	0.0002	0.103665
16	SV	Optimism	0.249628	0.231861	0.185969	0.164212	0.0002	0.101144
17	SV	Sadness	0.242449	0.275146	0.097923	0.119364	0.0002	0.296203
18	SV	Fear	0.195264	0.198991	0.082365	0.076330	0.0002	0.046919
19	NO	Anger	0.308310	0.350753	0.187001	0.183694	0.0002	0.227504
20	NO	Disgust	0.298375	0.344077	0.172432	0.169929	0.0002	0.264637
21	NO	Joy	0.392199	0.314069	0.265426	0.248319	0.0002	0.300547
22	NO	Optimism	0.308722	0.256740	0.185156	0.173305	0.0002	0.286880
23	NO	Sadness	0.210305	0.249812	0.095400	0.108767	0.0002	0.379187
24	NO	Fear	0.152347	0.182835	0.065698	0.072392	0.0002	0.430706

Table 1. Means, standard deviations, p-values and effect sizes of the Covid-19 vs. general subsample of tweet appraisals.

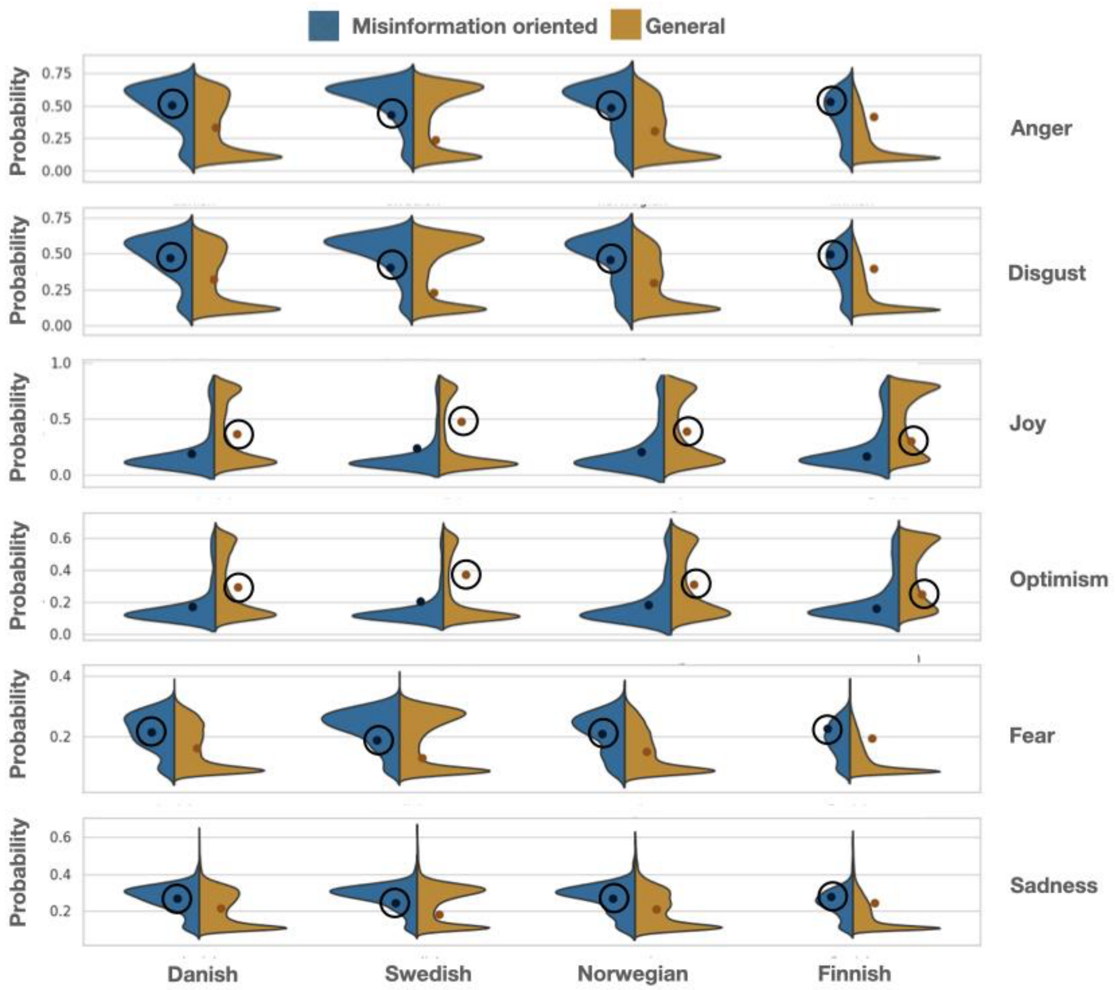


Figure 3. The means from all emotion landscapes are significantly different between misinformation-oriented and general-oriented tweets for all countries.

	Country	Emotion	Mean(Non crisis)	Mean(Disinformation)	SD(Non crisis)	SD(Disinformation)	P–value	Standardized Effectsize(Cohen's d)
1	DA	Anger	0.336619	0.501108	0.201242	0.165319	2e–04	0.815225
2	DA	Disgust	0.324223	0.471064	0.184123	0.150761	2e–04	0.795651
3	DA	Joy	0.369570	0.190802	0.265441	0.173129	2e–04	0.674171
4	DA	Optimism	0.293681	0.169596	0.184960	0.117657	2e–04	0.671694
5	DA	Sadness	0.212526	0.267299	0.092137	0.070812	2e–04	0.594776
6	DA	Fear	0.163265	0.215108	0.069447	0.060549	2e–04	0.744706
7	FI	Anger	0.237412	0.427409	0.154021	0.169979	2e–04	1.226102
8	FI	Disgust	0.232640	0.402127	0.140352	0.153482	2e–04	1.200625
9	FI	Joy	0.479348	0.240567	0.254596	0.194361	2e–04	0.936373
10	FI	Optimism	0.370278	0.204990	0.178156	0.131510	2e–04	0.926468
11	FI	Sadness	0.177904	0.241753	0.085214	0.073650	2e–04	0.748474
12	FI	Fear	0.130387	0.189057	0.054110	0.058478	2e–04	1.079181
13	SV	Anger	0.417055	0.530853	0.225188	0.160794	2e–04	0.506555
14	SV	Disgust	0.397367	0.496387	0.206096	0.146341	2e–04	0.481744
15	SV	Joy	0.304659	0.171981	0.269214	0.158207	2e–04	0.494878
16	SV	Optimism	0.249628	0.157835	0.185969	0.105623	2e–04	0.495746
17	SV	Sadness	0.242449	0.275605	0.097923	0.067430	2e–04	0.339985
18	SV	Fear	0.195264	0.226726	0.082365	0.060289	2e–04	0.383199
19	NO	Anger	0.308310	0.486619	0.187001	0.181571	2e–04	0.944375
20	NO	Disgust	0.298375	0.458446	0.172432	0.165218	2e–04	0.920162
21	NO	Joy	0.392199	0.208145	0.265426	0.192190	2e–04	0.693499
22	NO	Optimism	0.308722	0.181804	0.185156	0.130667	2e–04	0.685836
23	NO	Sadness	0.210305	0.264759	0.095400	0.076666	2e–04	0.571066
24	NO	Fear	0.152347	0.210037	0.065698	0.063552	2e–04	0.871110

Table 2. Means, standard deviations, p-values and effect sizes of the misinformation vs. general subsample of tweet appraisals.

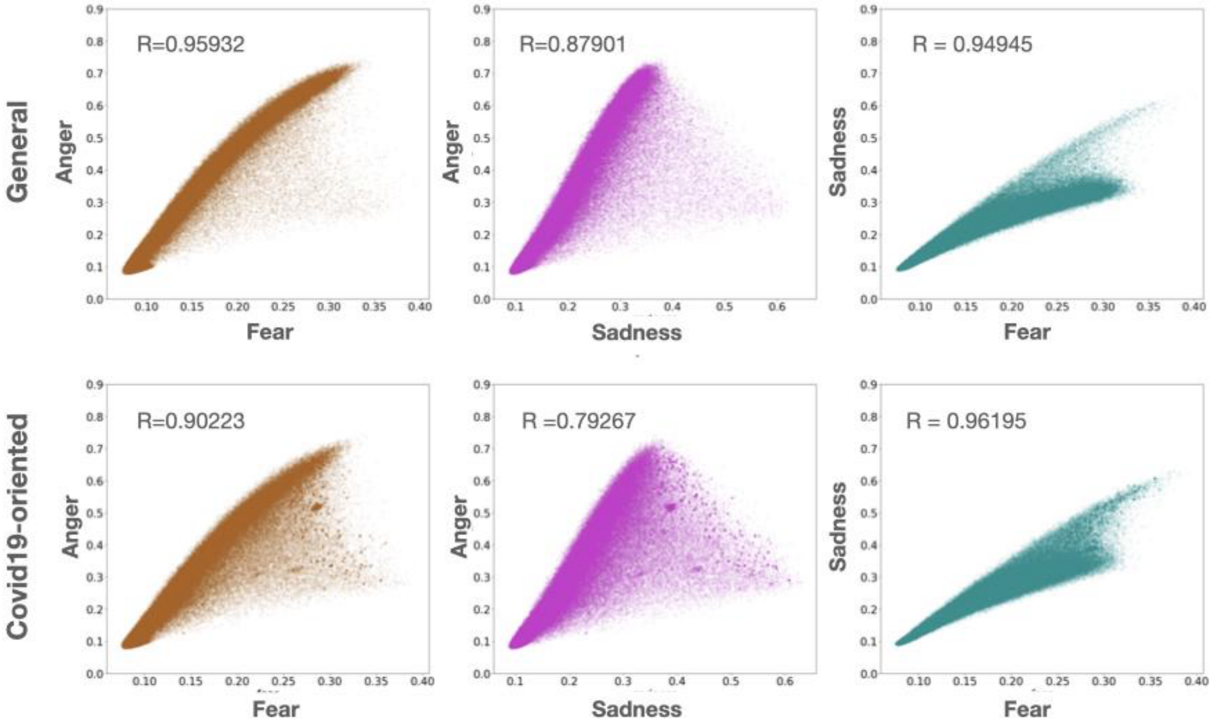
When comparing misinformation-oriented vs. general tweets, we observed significant differences for all emotions and all countries, reaching in all cases a p=0.0002 **(Figure 2)**. This is unsurprising given the results presented in **Figure 1**, where we present a descriptive overview of the differences between the emotional landscapes of misinformation vs. general appraisals across all languages and observed marked differences for all the emotions.

Lastly, we address the relationship between negative emotions in crisis oriented vs. non-crisis oriented tweets. Here we find a strong linear relationship between all combinations (anger-fear, sadness-fear and anger-sadness) across the three subsamples. However, that relationship weakened when looking at anger combinations i.e., anger-fear and anger-sadness in the Covid-19 tweets vs. general oriented tweets (anger-fear p=0,00009; anger-sadness p=0.00009) **(Figure 4a).** This weakening of the relationship in crisis-oriented tweets vs. non-crisis oriented was not present for the sadness-fear combination. This effect was present both when looking at the probability scores of all languages together **(Figure 4a)** and when looking at each language individually **(Supplementary figure 1,2,3).**

When comparing the misinformation oriented vs the general subsample, we did not find any significant differences for any of the emotion combinations (anger-fear, p=0.07239; anger-sadness, p

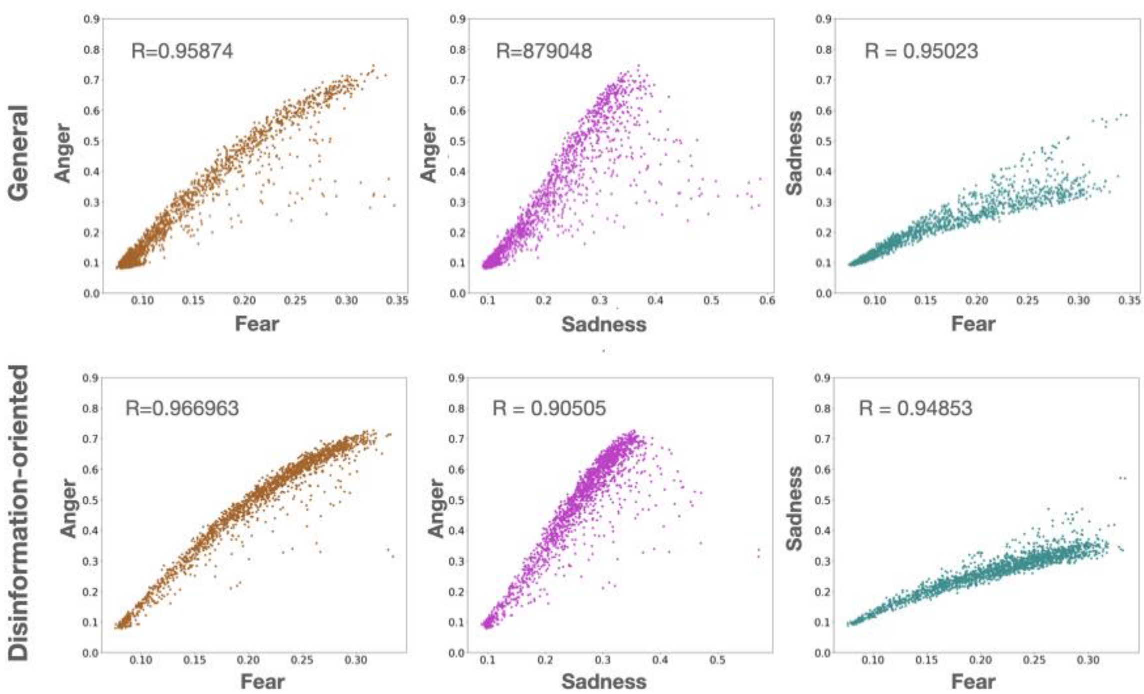
= 0.06329; sadness-fear, p = 0.68113) **(Figure 3b).** However, we observed a trend centered around anger for the combinations anger-fear and anger-sadness, with the misinformation-oriented subsample presenting a stronger correlation than the general subsample.

A)



Anger - Fear	Anger - Sadness	Sadness - Fear
p = 0.000009	p = 0.000009	p = 0.999900
rdiff=0.05709	rdiff = 0.08634	rdiff=-0.01250

B)



Anger - Fear	Anger - Sadness	Sadness - Fear
p = 0.07239	p = 0.06329	p = 0.68113
rdiff=-0.00821	rdiff = -0.02600	rdiff=0.00169

Figure 4a) Emotion interaction in crisis-oriented vs. general tweets. a) In Covid19-oriented tweets, the correlation between anger and other negative emotions (fear and sadness) is weaker than in non-crisis oriented tweets (p=0.000009 for both). This effect is absent in the relationship between other negative emotions (i.e., sadness-fear) when comparing Covid19-oriented vs general tweets. **b)** No significant differences were observed for any of the emotion combinations when comparing misinformation vs general tweets, although we observe a trend for the combinations anger-fear (p=0.07239) and anger-sadness (p=0.06329).

5.0 Discussion

In this study, we investigate the emotional landscapes of crisis-oriented tweets in a large Nordic Twitter dataset.

Overall, our results show significant differences between both crisis-oriented samples and the general subsample. However, the differences between the misinformation oriented and the general oriented sample had larger effect sizes than the differences between the Covid-10 and the general subsample.

This seems to indicate that appraisals regarding the Covid-19 crisis do not contain much more emotional expression that general appraisal tweets and, as such, they are less likely to lead to collective action based on emotional expression as one of the predictive factors. Given the context of the Nordic countries, with their resilient democracies and trusting societies, the factor of emotional expression is considered particularly important when assessing potential for collective action. On the other hand, the differences in emotional expression between the misinformation oriented and the general appraisals were significant and showing large effect sizes, possibly indicating a mature context for collective action. Further analyses are needed in order to better understand to what degree does the potential for collective action materialise in both scenarios.

Our analyses on the relationship between emotions revealed a weaker linear relationship between anger and both fear and sadness in the Covid19-oriented subsample in contrast to the non crisis-oriented subsample. This effect was not present when looking at the relationship between other negative emotions, in particular, sadness and fear, suggesting that it is an effect specific to anger. While the same effect was not present in the misinformation vs. non crisis oriented sample, it is worth noting that the sample size was much smaller in this case. Nevertheless, we observed a trend in the anger-fear and anger-sadness combinations, suggesting that the relationship between those two emotions differ when appraising a crisis.

While we aim for a comprehensive account of emotion, we would like to highlight the behavioural consequences of specific negative emotions that often appear in response to threat (i.e., crises). In particular experiencing anger often translates behaviourally to action/approach, diminished risk perception and less careful processing of information. On the contrary, fear is linked to avoidance, risk overestimation and increased attention to threat. In theory, collective action can be facilitated by high levels of anger Mackie et al. 2000 and low levels of fear and sadness (Miller et al., 2009). As such, a change in the ratio between anger and the other two emotions, where fear and sadness become more prominent, might indicate decreased chances for collective action. This is what we found in the Covid-19 vs. general tweet appraisals. On the other hand, a change in the ratio in which fear and sadness become less prominent, leaving anger dominate, might lead to increased chances for collective action. We found a trend towards this effect in the misinformation vs. general appraisals. Further research is needed to investigate how this quantifies in online collective action and beyond the social media context into everyday life.

It is important to mention that this study also has some limitations, some of which could be addressed in future studies. In particular, the availability of training data in the Nordic languages is very scarce, and as such we had to automatically translate an English dataset into the languages of the study, with the consequent loss of information that this involves. Nevertheless, this approach was less detrimental and time costly than translating the entire Twitter dataset into English. Furthermore, our training dataset was unbalanced, containing different amounts of tweets expressing each of the emotions. As such, the model was not able to detect emotions such as anticipation and trust and less able to detect e.g., fear than anger. In addition, some phenomena like the use of irony and sarcasm, very present on Twitter during earlier stages of the pandemic at least in other countries (Vicari & Murru, 2020), will have gone unnoticed and might have resulted in some tweets being erroneously labelled as joy, when they should have been labelled as anger.

Overall, we found indicators for differences in emotional expression when comparing two different types of crisis oriented tweet appraisals in the Nordic Twittersphere. Taken into the context of Nordic countries, with their resilient democracies and high trust societies, emotions have been suggested to be particularly important in organising for collective action (Groenedyck, 2011). This study suggests that the misinformation crisis would be more likely to present a fertile environment for collective action in the Nordic countries than the co-occurring crisis around Covid-19. However, more research is needed to investigate the degree to which this translates into collective action both in online and offline behaviour.

5.0 References

Aknin, L. B., Andretti, B., Goldszmidt, R., Helliwell, J. F., Petherick, A., De Neve, J.-E., Dunn, E. W., Fancourt, D., Goldberg, E., Jones, S. P., Karadag, O., Karam, E., Layard, R., Saxena, S., Thornton, E., Whillans, A., & Zaki, J. (2022). Policy stringency and mental health during the COVID-19 pandemic: A longitudinal analysis of data from 15 countries. *The Lancet Public Health*, *7*(5), e417–e426. https://doi.org/10.1016/S2468-2667(22)00060-3

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2021). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine, 51*(10), 1763-1769. doi:10.1017/S003329172000224X

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it:

Rapid review of the evidence. *The Lancet*, *395*(10227), 912–920. https://doi.org/10.1016/S0140-6736(20)30460-8

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). *Unsupervised Cross-lingual Representation Learning at Scale* (arXiv:1911.02116). arXiv. http://arxiv.org/abs/1911.02116

Delhey, J., & Newton, K. (2005). Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism? *European Sociological Review*, *21*(4), 311–327. https://doi.org/10.1093/esr/jci022

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

Dudău, D. P., & Sava, F. A. (2021). Performing Multilingual Analysis With Linguistic Inquiry and Word Count 2015 (LIWC2015). An Equivalence Study of Four Languages. *Frontiers in Psychology*, *12*, 570568. https://doi.org/10.3389/fpsyg.2021.570568

European Commission, Directorate-General for Communications Networks, Content and Technology, *A multi-dimensional approach to disinformation : report of the independent High level Group on fake news and online disinformation*, Publications Office, 2018, https://data.europa.eu/doi/10.2759/739290

Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger Is More Influential than Joy: Sentiment Correlation in Weibo. *PLoS ONE*, *9*(10), e110184. https://doi.org/10.1371/journal.pone.0110184

Gregory, K. (2018). *Anger in Academic Twitter: Sharing, Caring, and Getting Mad Online*. 18.

Groenendyk, E. (2011). Current Emotion Research in Political Science: How Emotions Help

Democracy Overcome its Collective Action Problem. Emotion Review, 3(4), 455–463. https://doi.org/10.1177/1754073911410746

Harper, C. A., Satchell, L. P., Fido, D., & Latzman, R. D. (2020). Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-020-00281-5

Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access*, *8*, 181074–181090. https://doi.org/10.1109/ACCESS.2020.3027350

Johansson, F., Côté, P., Hogg-Johnson, S., Rudman, A., Holm, L. w., Grotle, M., Jensen, I., Sundberg, T., Edlund, klara, & Skillgate, E. (2021). Depression, anxiety and stress among Swedish university students before and during six months of the COVID-19 pandemic: A cohort study. *Scandinavian Journal of Public Health*, *49*(7), 741–749. https://doi.org/10.1177/14034948211015814

Kalsnes, B. Fake News. *Oxford Research Encyclopedia of Communication.* Retrieved 27 Feb. 2023, from https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-809.

Larsen, M. E., Boonstra, T. W., Batterham, P. J., ODea, B., Paris, C., & Christensen, H. (2015). We Feel: Mapping Emotion on Twitter. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1246–1252. https://doi.org/10.1109/JBHI.2015.2403839

Lazarus, R. S. (1991). *Emotion and adaptation.* Oxford University Press.

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health and Surveillance*, *6*(2), e19447.

https://doi.org/10.2196/19447

Marroquín, B., Vine, V., & Morgan, R. (2020). Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. *Psychiatry Research*, *293*, 113419. https://doi.org/10.1016/j.psychres.2020.113419

Miller, Daniel & Cronin, Tracey & Garcia, Amber & Branscombe, Nyla. (2009). The Relative Impact of Anger and Efficacy on Collective Action is Affected by Feelings of Fear. Psychology Faculty Publications. 12. 10.1177/1368430209105046.

Mohammad, S. M., & Bravo-Marquez, F. (2017). Emotion Intensities in Tweets. *ArXiv:1708.03696 [Cs]*. http://arxiv.org/abs/1708.03696

OECD. (2020). *Trust in government* [Doi:https://doi.org/10.1787/1de9675e-en]. https://www.oecd-ilibrary.org/content/data/1de9675e-en

OECD. (2021a). *Education at a Glance 2021: OECD Indicators*. OECD. https://doi.org/10.1787/b35a14e5-en

OECD. (2021b). *Health at a Glance 2021: OECD Indicators*. OECD. https://doi.org/10.1787/ae3016b9-en

OECD. (2022). *Building Trust to Reinforce Democracy*. https://doi.org/10.1787/b407f99c-en

O'Hara, L., Abdul Rahim, H. F., & Shi, Z. (2020). *Gender and trust in government modify the association between mental health and stringency of social distancing related public health measures to reduce COVID-19: A global online survey* [Preprint]. Public and Global Health. https://doi.org/10.1101/2020.07.16.20155200

Ohlsson, J., Blach-Ørsten, M., & Willig, I. (2021). *Covid-19 och de nordiska nyhetsmedierna* (p. 138). Nordicom, University of Gothenburg. https://doi.org/10.48335/9789188855473

Pennebaker, J., Booth, R., Boyd, R., & Francis, M. (2015). *Linguistic Inquiry and Word Count: LIWC2015*.

Saunes, I. S., Vrangbæk, K., Byrkjeflot, H., Jervelund, S. S., Birk, H. O., Tynkkynen, L.-K., Keskimäki, I., Sigurgeirsdóttir, S., Janlöv, N., Ramsberg, J., Hernández-Quevedo, C., Merkur, S., Sagan, A., & Karanikolos, M. (2022). Nordic responses to Covid-19: Governance and policy measures in the early phases of the pandemic. *Health Policy*, *126*(5), 418–426. https://doi.org/10.1016/j.healthpol.2021.08.011

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research.* Oxford University Press.

Sher, L. (2020). COVID-19, anxiety, sleep disturbances and suicide. *Sleep Medicine*, *70*, 124. https://doi.org/10.1016/j.sleep.2020.04.019

Syvertsen, T., Enli, G., Mjøs, O. J., & Moe, H. (2014). *The Media Welfare State: Nordic Media in the Digital Era*. University of Michigan Press. https://doi.org/10.2307/j.ctv65swsg

Edson C. Tandoc Jr., Zheng Wei Lim & Richard Ling (2018) Defining “Fake News”, Digital Journalism, 6:2, 137-153, DOI: 10.1080/21670811.2017.1360143

van Zomeren, M. (2013). Four Core Social-Psychological Motivations to Undertake Collective Action. Social and Personality Psychology Compass, 7(6), 378-388.

van Zomeren, M., Leach, C. W., & Spears, R. (2012). Protesters as “Passionate Economists”: A dynamic dual pathway model of approach coping with collective disadvantage. *Personality and Social Psychology Review, 16*(2), 180–199. https://doi.org/10.1177/1088868311430835

Vicari, S., & Murru, M. F. (2020). One Platform, a Thousand Worlds: On Twitter Irony in the Early Response to the COVID-19 Pandemic in Italy. *Social Media + Society*, *6*(3), 205630512094825. https://doi.org/10.1177/2056305120948254

Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)

Widmann, T. (2022). Fear, Hope, and COVID-19: Emotional Elite Rhetoric and Its Impact on the Public During the First Wave of the COVID-19 Pandemic. *Political Psychology*, pops.12831. https://doi.org/10.1111/pops.12831

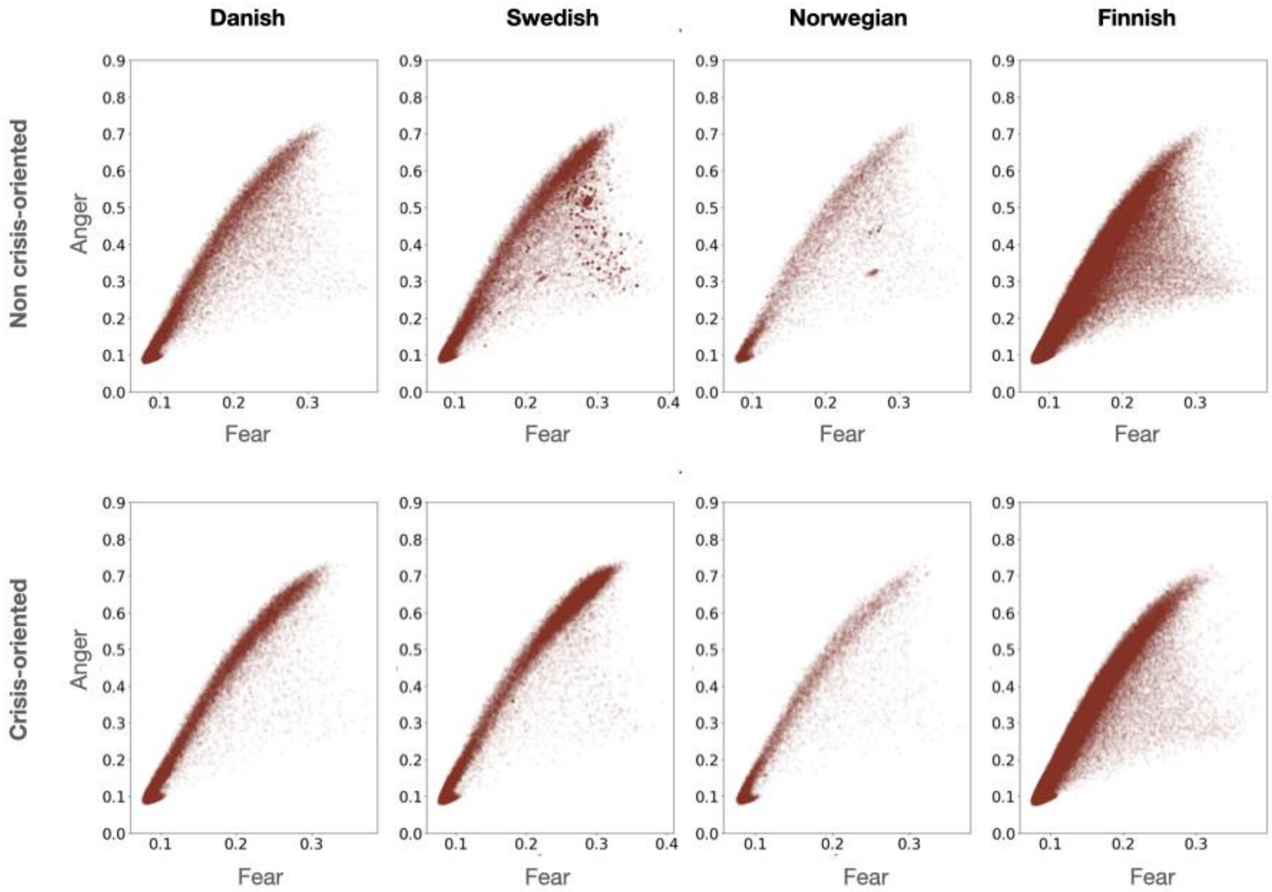
Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*, *15*(9), e0239441. https://doi.org/10.1371/journal.pone.0239441

Zhang, S., Sun, L., Zhang, D., Li, P., Liu, Y., Anand, A., Xie, Z., & Li, D. (2022). The COVID-19 Pandemic and Mental Health Concerns on Twitter in the United States. *Health Data Science*, *2022*, 1–9. https://doi.org/10.34133/2022/9758408

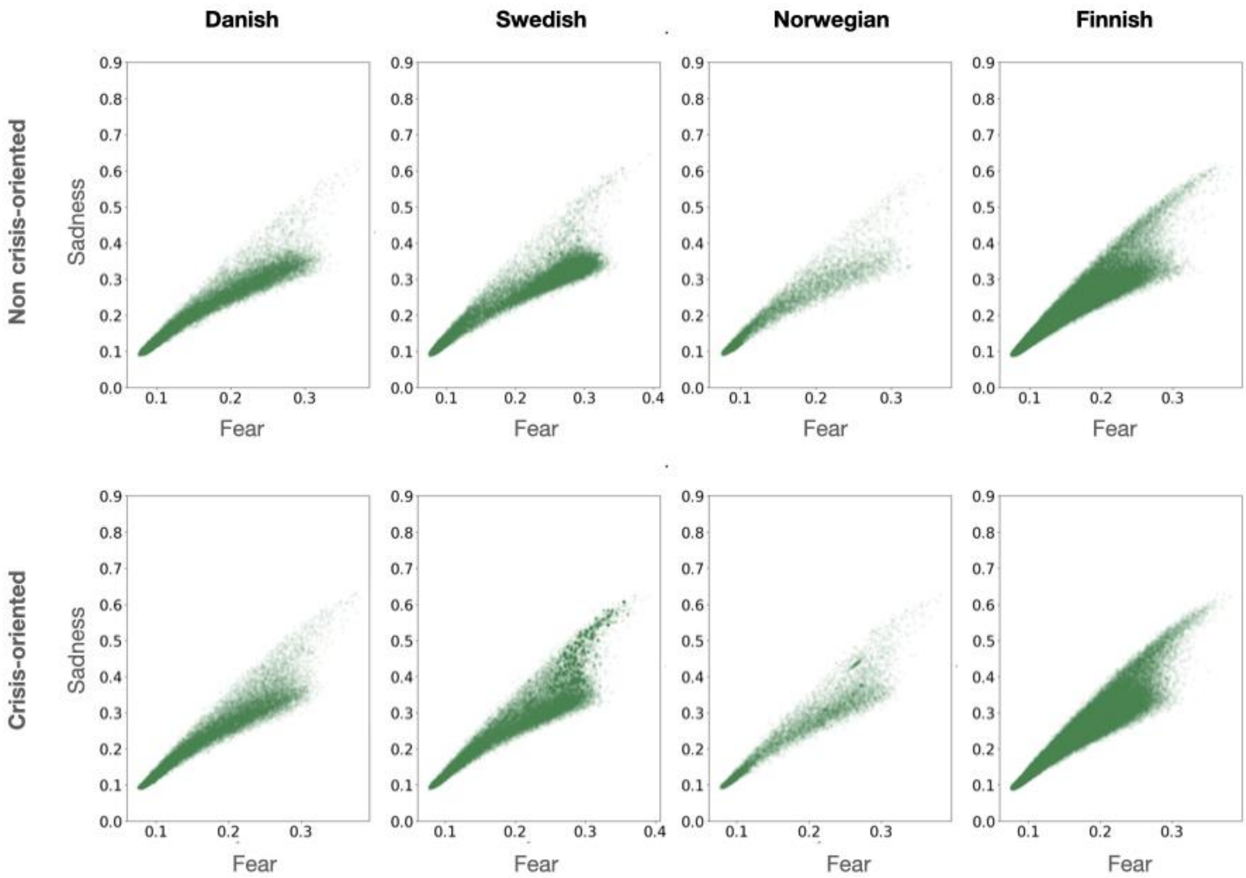
X.X Annexes

	Precision	Recall	F1-score
Joy	0.77	0.84	0.81
Anger	0.68	0.81	0.74
Disgust	0.65	0.80	0.72
Optimism	0.65	0.81	0.72
Fear	0.71	0.61	0.66
Sadness	0.64	0.64	0.64
Love	0.50	0.75	0.60
Pessimism	0.42	0.33	0.37
Anticipation	0.33	0.29	0.31
Trust	0.19	0.07	0.10
Suprise	1	0.00	0.00

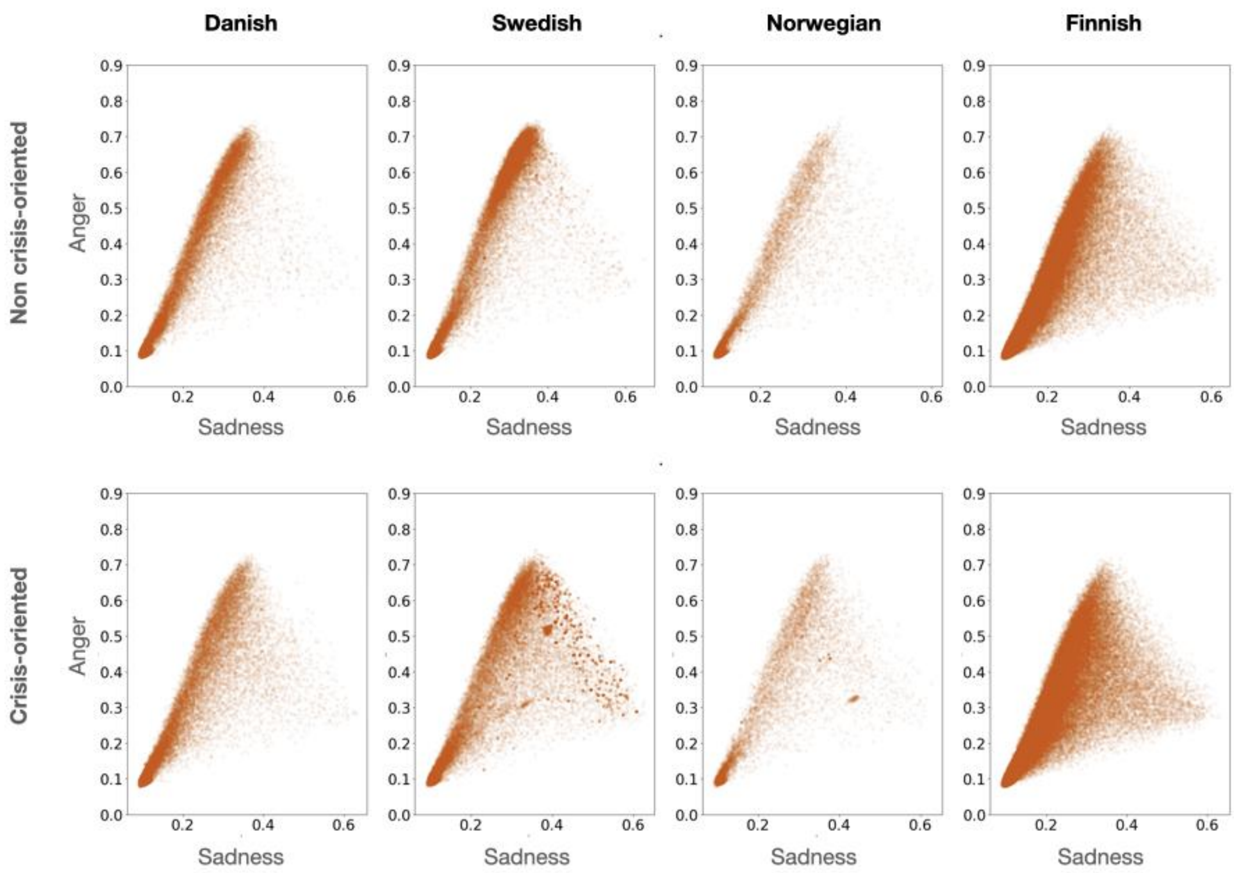
Table 1. Performance scores from the XLM-RoBERTa model for emotion detection in the Nordic languages



Supplementary Figure 1. Weakened linear relationship for the anger-fear combination in crisis-oriented vs. non crisis-oriented tweets.



Supplementary figure 2. no weakening effect present in the linear relationship for the sadness-fear combination in crisis-oriented vs. non crisis-oriented tweets.



Supplementary Figure 3. Weakened linear relationship for the anger-sadness combination in crisis-oriented vs. non crisis-oriented tweets.