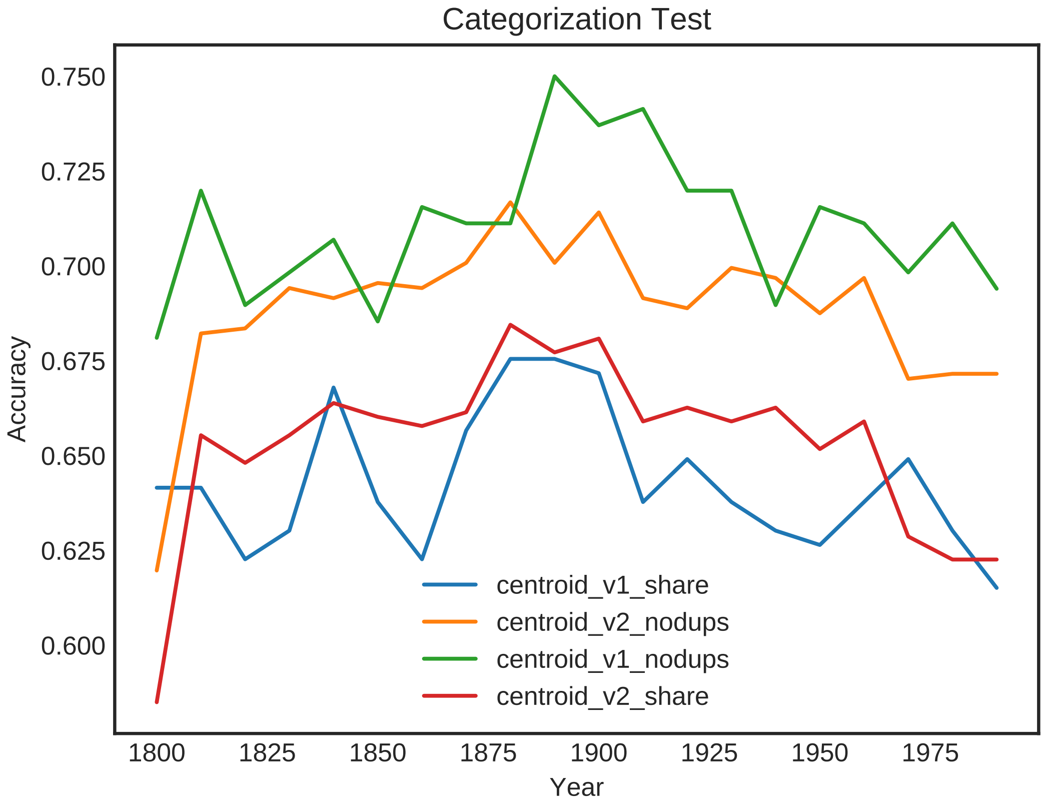
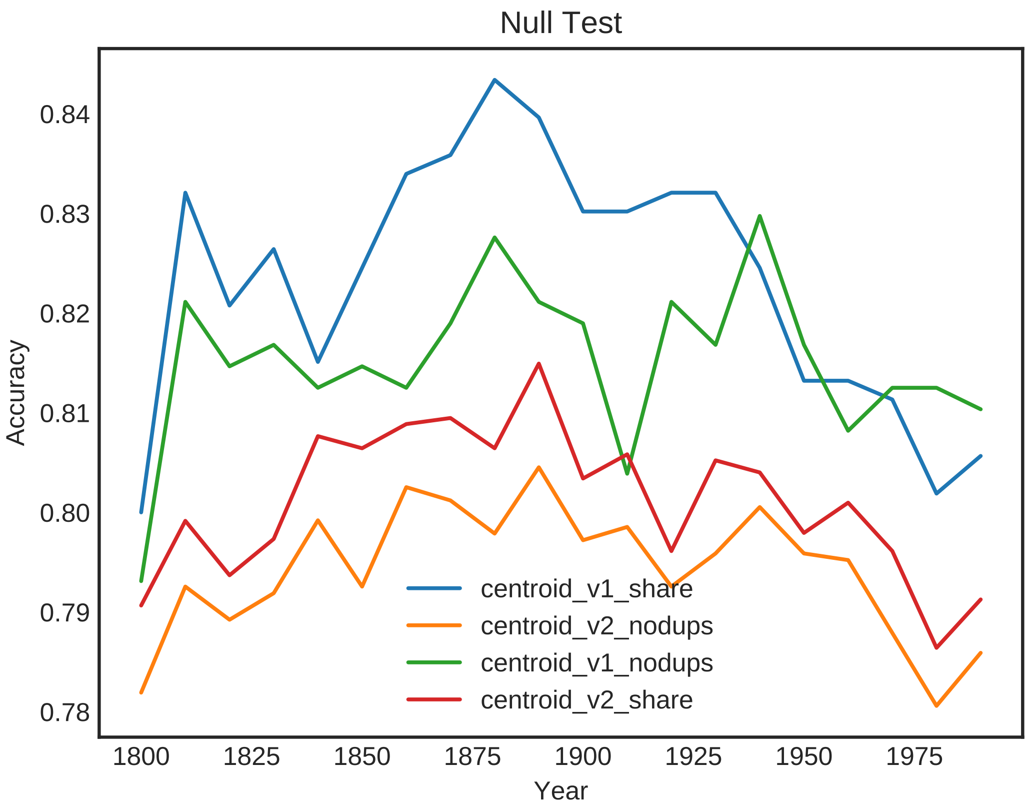
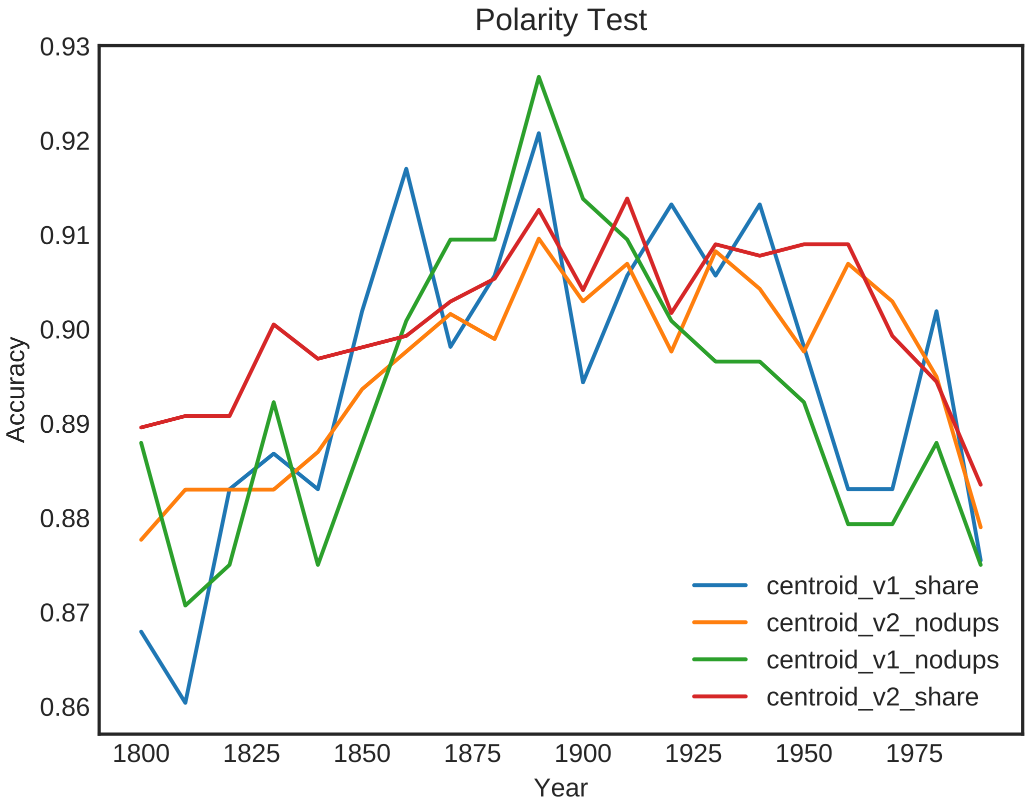
Table 1: Average accuracy of all models and data schemes.

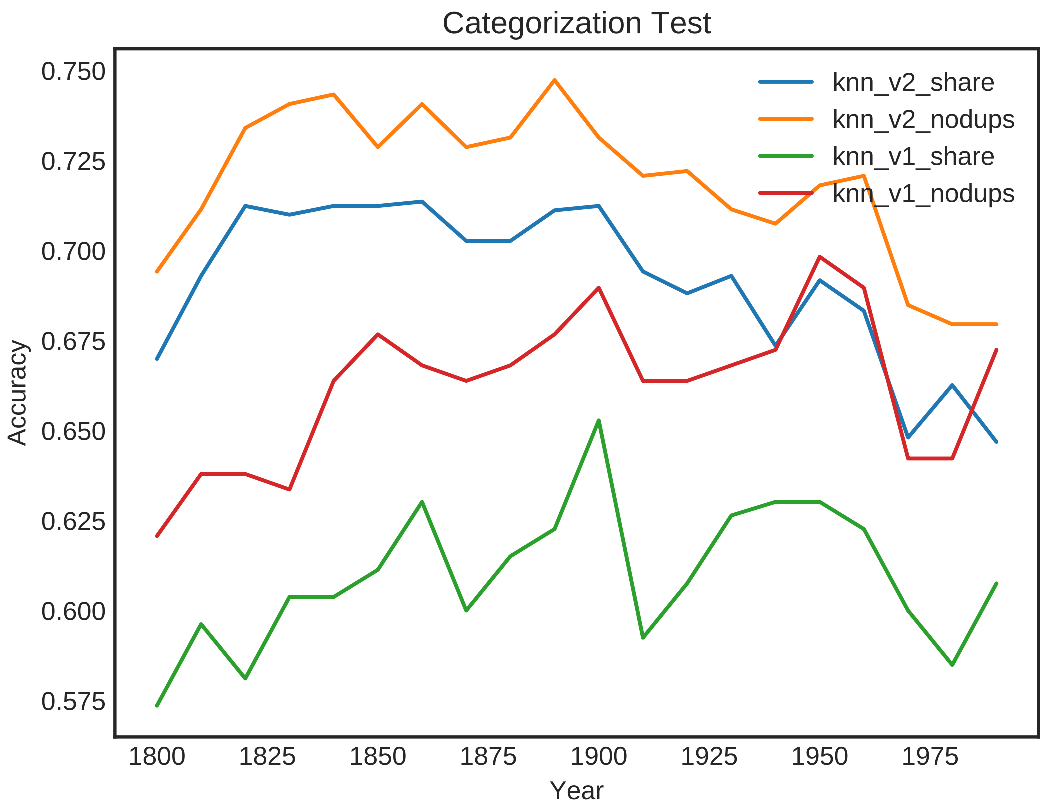
Model Name	**Categorization Test**	**Null Test**	**Polarity Test**
centroid˙v1˙share	0.64	0.82	0.89
centroid˙v1˙nodups	0.71	0.82	0.89
centroid˙v2˙share	0.65	0.80	0.90
centroid˙v2˙nodups	0.69	0.79	0.90
knn˙v1˙share	0.61	0.83	0.91
knn˙v1˙nodups	0.66	0.83	0.91
knn˙v2˙share	0.69	0.83	0.91
knn˙v2˙nodups	0.72	0.82	0.91

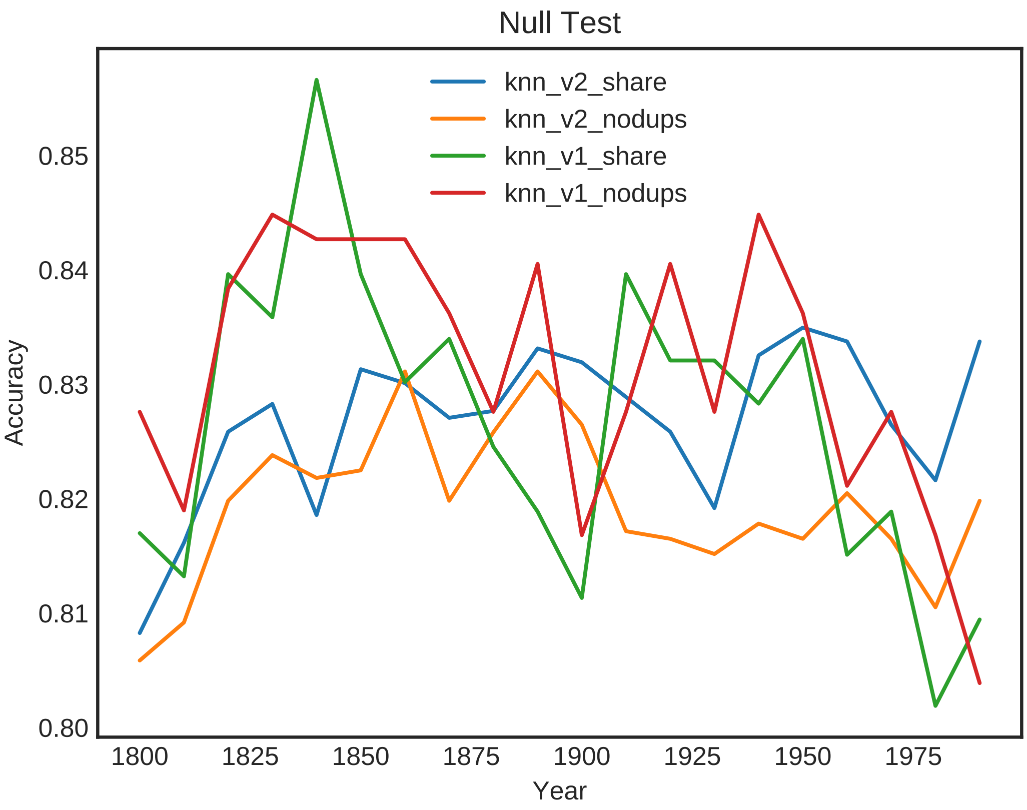
The following plots compare the centroid and kNN models under variation of the MFD version (v1 vs v2) and whether duplicate seed words are kept (”share”) or removed (”nodups”).




Null Test
Accuracy
0.84
0.83
0.82
0.81
0.80
0.79
0.78
1800
1825
1850
1875
1900
1925
1950
1975
Year
centroid_v1_share
centroid_v2_nodups
centroid_v1_nodups
centroid_v2_share


Polarity Test
Accuracy
0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
Year
1800
1825
1850
1875
1900
1925
1950
1975
centroid_v1_share
centroid_v2_nodups
centroid_v1_nodups
centroid_v2_share


Categorization Test
knn_v2_share
knn_v2_nodups
knn_v1_share
knn_v1_nodups
Accuracy
0.750
0.725
0.700
0.675
0.650
0.625
0.600
0.575
1800
1825
1850
1875
1900
1925
1950
1975
Year


Null Test
knn_v2_share
knn_v2_nodups
knn_v1_share
knn_v1_nodups
Accuracy
0.85
0.84
0.83
0.82
0.81
0.80
1800
1825
1850
1875
1900
1925
1950
1975
Year

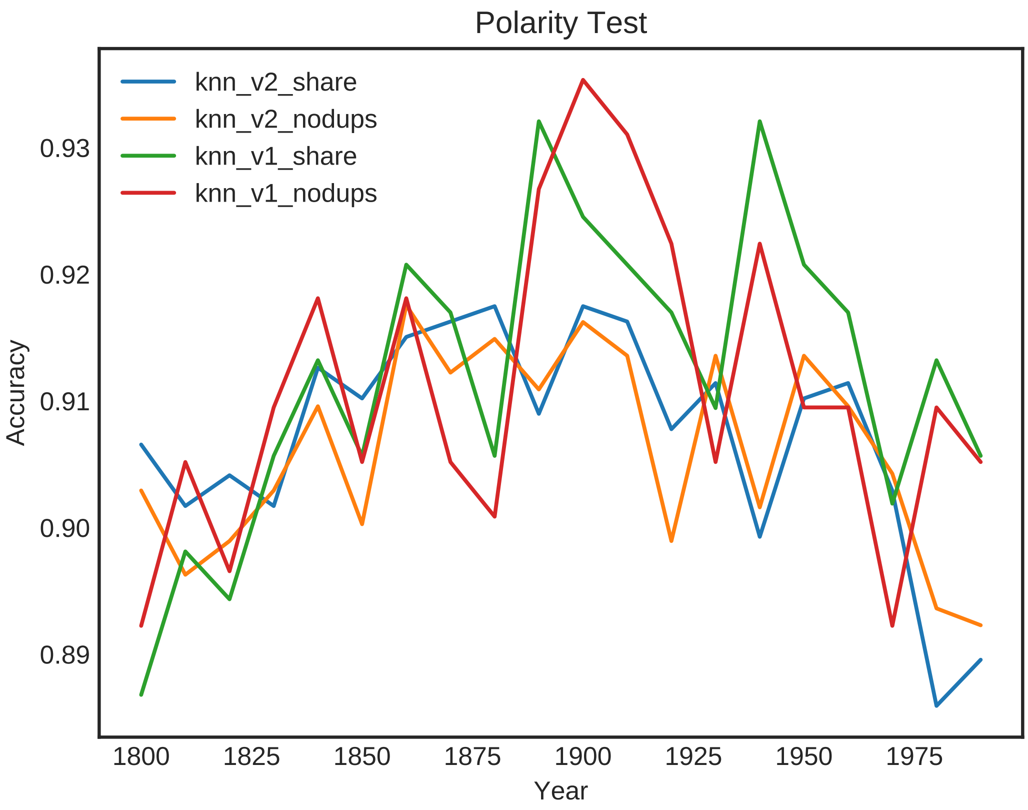
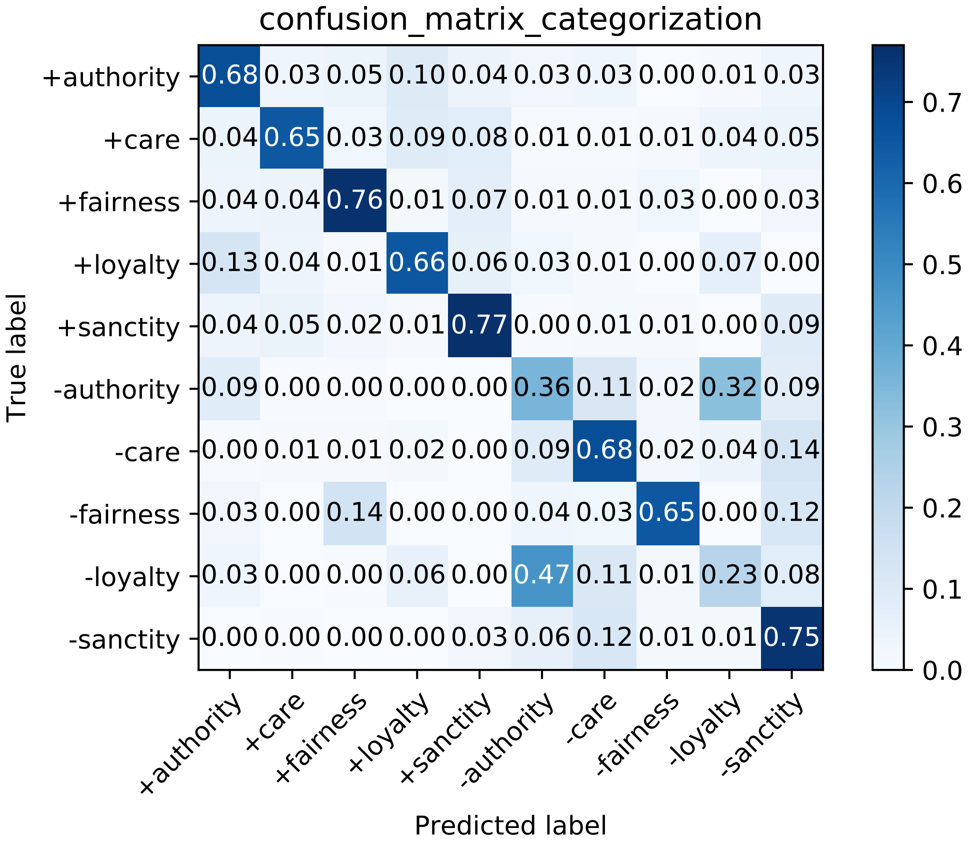
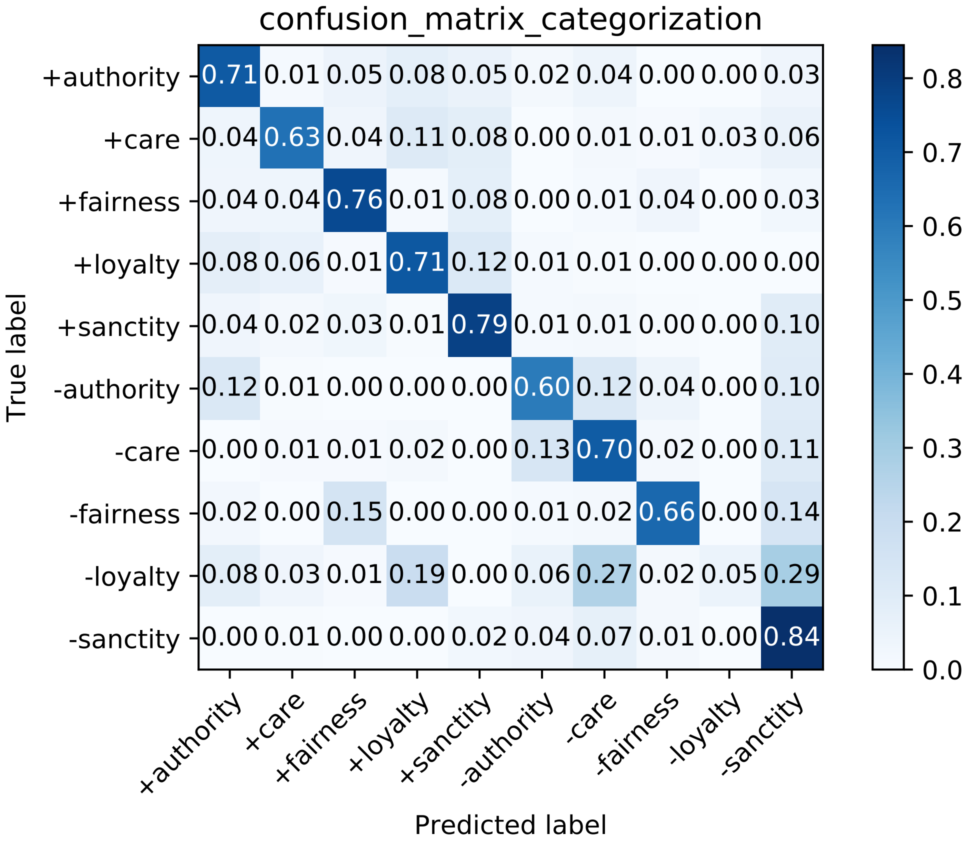

Polarity Test
knn_v2_share
knn_v2_nodups
knn_v1_share
knn_v1_nodups
Accuracy
0.93
0.92
0.91
0.90
0.89
1800
1825
1850
1875
1900
1925
1950
1975
Year

Figure 1: Categorization confusion matrix: Centroid, V1, 'share'



	+authority	+care	+fairness	+loyalty	+sanctity	-authority	-care	-fairness	-loyalty	-sanctity
+authority	0.68	0.03	0.05	0.10	0.04	0.03	0.03	0.00	0.01	0.03
+care	0.04	0.65	0.03	0.09	0.08	0.01	0.01	0.01	0.04	0.05
+fairness	0.04	0.04	0.76	0.01	0.07	0.01	0.01	0.03	0.00	0.03
+loyalty	0.13	0.04	0.01	0.66	0.06	0.03	0.01	0.00	0.07	0.00
+sanctity	0.04	0.05	0.02	0.01	0.77	0.00	0.01	0.01	0.00	0.09
-authority	0.09	0.00	0.00	0.00	0.00	0.36	0.11	0.02	0.32	0.09
-care	0.00	0.01	0.01	0.02	0.00	0.09	0.68	0.02	0.04	0.14
-fairness	0.03	0.00	0.14	0.00	0.00	0.04	0.03	0.65	0.00	0.12
-loyalty	0.03	0.00	0.00	0.06	0.00	0.47	0.11	0.01	0.23	0.08
-sanctity	0.00	0.00	0.00	0.00	0.03	0.06	0.12	0.01	0.01	0.75

Figure 2: Categorization confusion matrix: Centroid, V1, 'nodups'



	+authority	+care	+fairness	+loyalty	+sanctity	-authority	-care	-fairness	-loyalty	-sanctity
+authority	0.71	0.01	0.05	0.08	0.05	0.02	0.04	0.00	0.00	0.03
+care	0.04	0.63	0.04	0.11	0.08	0.00	0.01	0.01	0.03	0.06
+fairness	0.04	0.04	0.76	0.01	0.08	0.00	0.01	0.04	0.00	0.03
+loyalty	0.08	0.06	0.01	0.71	0.12	0.01	0.01	0.00	0.00	0.00
+sanctity	0.04	0.02	0.03	0.01	0.79	0.01	0.01	0.00	0.00	0.10
-authority	0.12	0.01	0.00	0.00	0.00	0.60	0.12	0.04	0.00	0.10
-care	0.00	0.01	0.01	0.02	0.00	0.13	0.70	0.02	0.00	0.11
-fairness	0.02	0.00	0.15	0.00	0.00	0.01	0.02	0.66	0.00	0.14
-loyalty	0.08	0.03	0.01	0.19	0.00	0.06	0.27	0.02	0.05	0.29
-sanctity	0.00	0.01	0.00	0.00	0.02	0.04	0.07	0.01	0.00	0.84

Figure 3: Categorization confusion matrix: Centroid, V2, 'share'

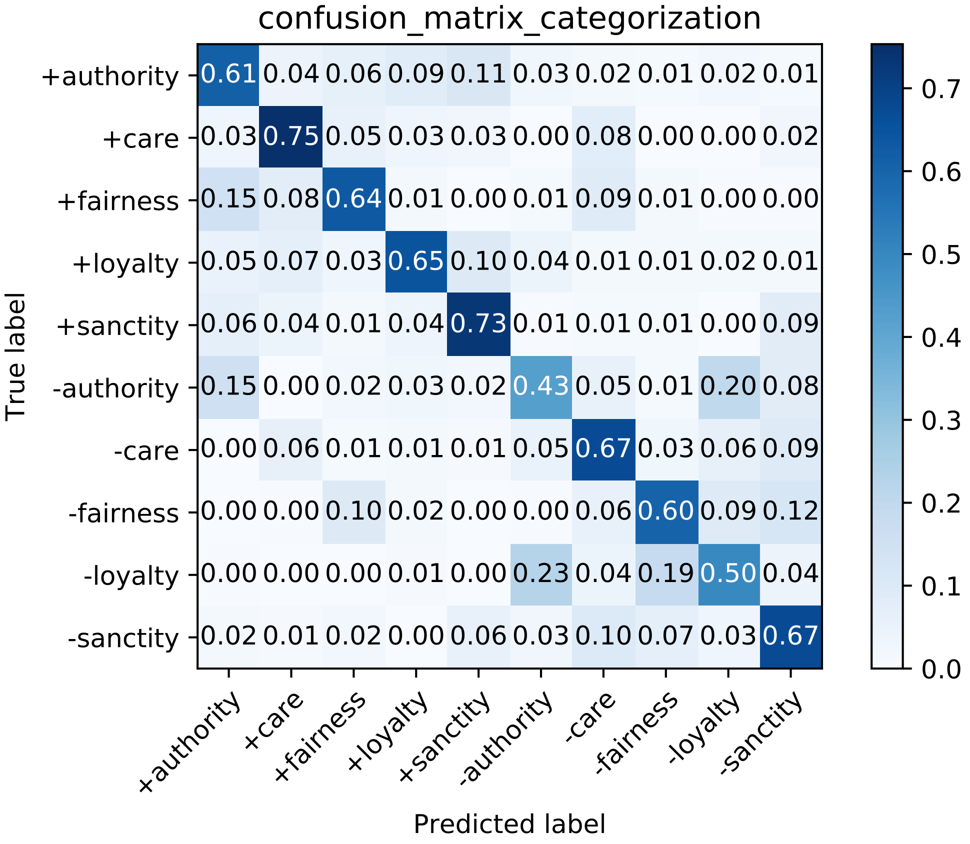


Figure 4: Categorization confusion matrix: Centroid, V2, 'nodups'

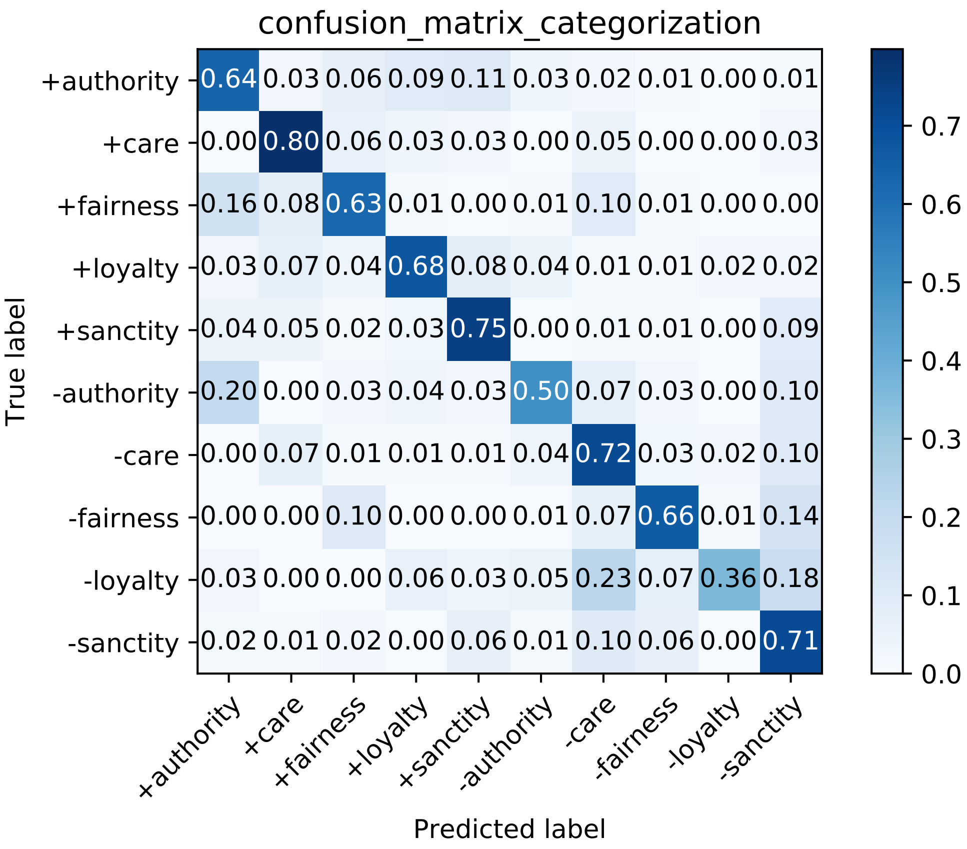
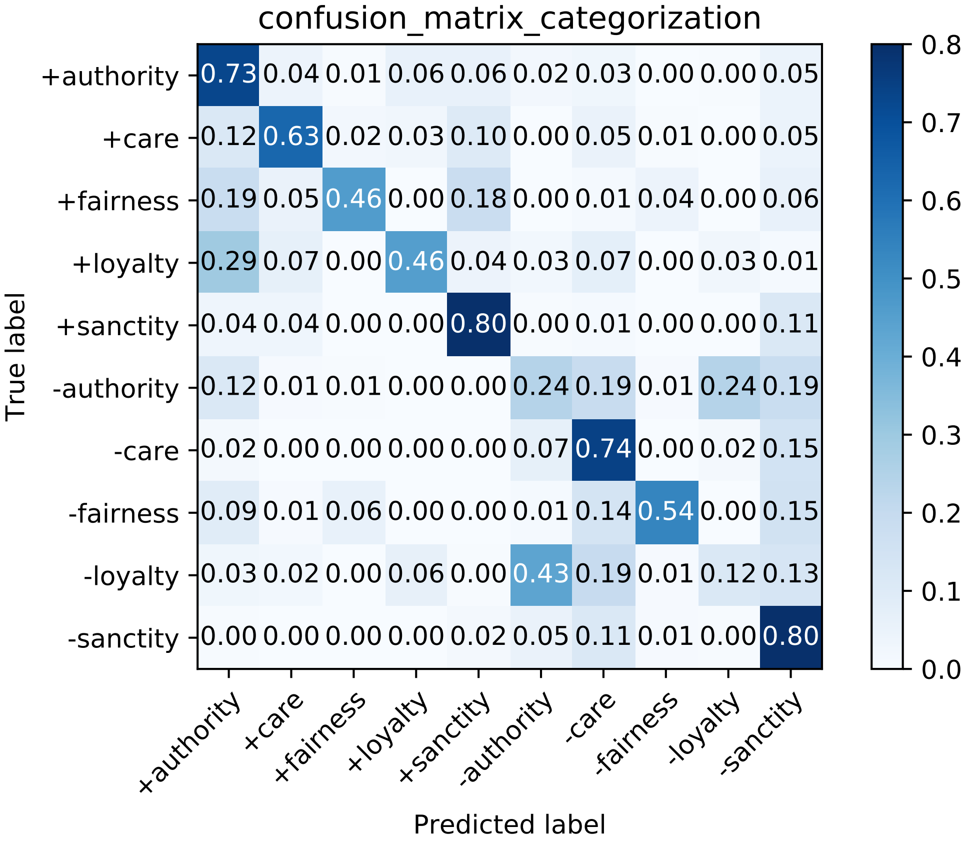


Figure 5: Categorization confusion matrix: kNN, V1, 'share'



	+authority	+care	+fairness	+loyalty	+sanctity	-authority	-care	-fairness	-loyalty	-sanctity
+authority	0.73	0.04	0.01	0.06	0.06	0.02	0.03	0.00	0.00	0.05
+care	0.12	0.63	0.02	0.03	0.10	0.00	0.05	0.01	0.00	0.05
+fairness	0.19	0.05	0.46	0.00	0.18	0.00	0.01	0.04	0.00	0.06
+loyalty	0.29	0.07	0.00	0.46	0.04	0.03	0.07	0.00	0.03	0.01
+sanctity	0.04	0.04	0.00	0.00	0.80	0.00	0.01	0.00	0.00	0.11
-authority	0.12	0.01	0.01	0.00	0.00	0.24	0.19	0.01	0.24	0.19
-care	0.02	0.00	0.00	0.00	0.00	0.07	0.74	0.00	0.02	0.15
-fairness	0.09	0.01	0.06	0.00	0.00	0.01	0.14	0.54	0.00	0.15
-loyalty	0.03	0.02	0.00	0.06	0.00	0.43	0.19	0.01	0.12	0.13
-sanctity	0.00	0.00	0.00	0.00	0.02	0.05	0.11	0.01	0.00	0.80

Figure 6: Categorization confusion matrix: kNN, V1, 'nodups'

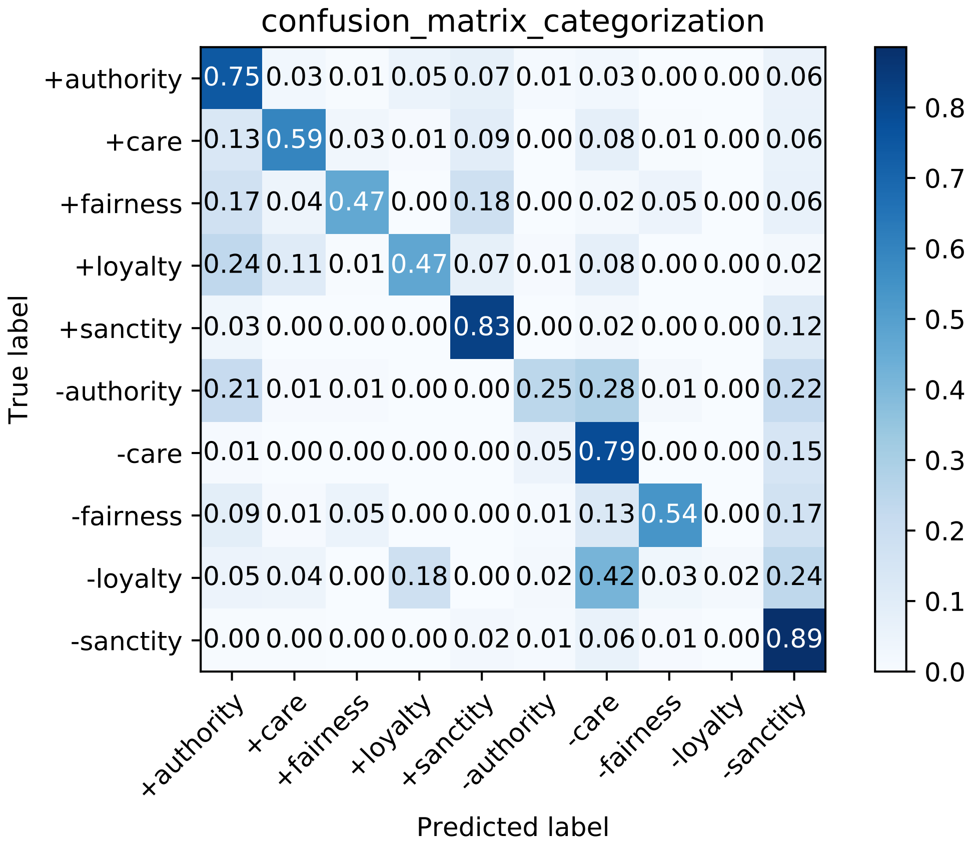


Figure 7: Categorization confusion matrix: kNN, V2, ’share’

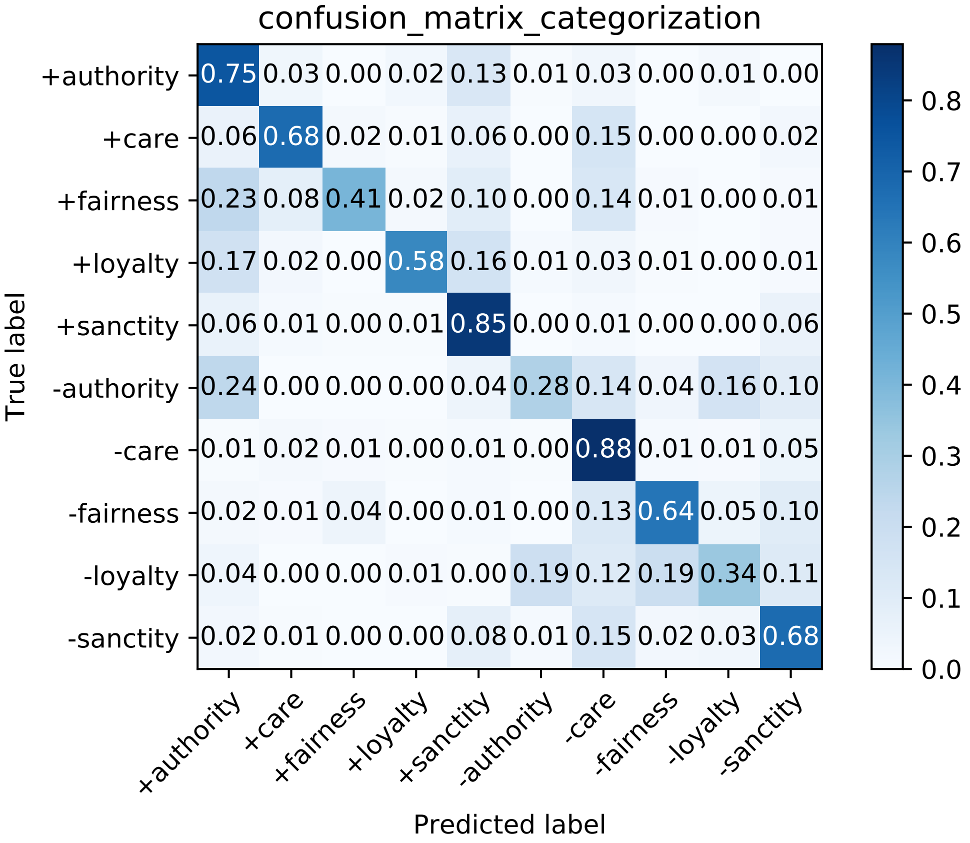
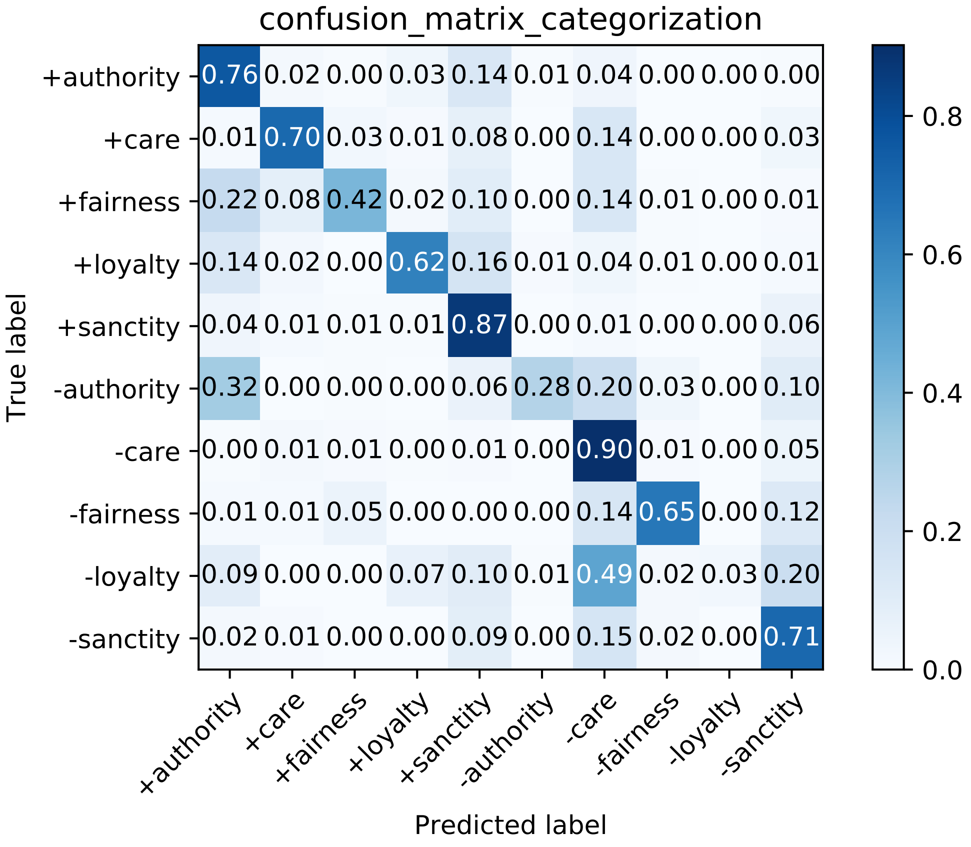


Figure 8: Categorization confusion matrix: kNN, V2, 'nodups'


confusion_matrix_categorization
True label
+authority 0.76 0.02 0.00 0.03 0.14 0.01 0.04 0.00 0.00 0.00
+care 0.01 0.70 0.03 0.01 0.08 0.00 0.14 0.00 0.00 0.03
+fairness 0.22 0.08 0.42 0.02 0.10 0.00 0.14 0.01 0.00 0.01
+loyalty 0.14 0.02 0.00 0.62 0.16 0.01 0.04 0.01 0.00 0.01
+sanctity 0.04 0.01 0.01 0.01 0.87 0.00 0.01 0.00 0.00 0.06
-authority 0.32 0.00 0.00 0.00 0.06 0.28 0.20 0.03 0.00 0.10
-care 0.00 0.01 0.01 0.00 0.01 0.00 0.90 0.01 0.00 0.05
-fairness 0.01 0.01 0.05 0.00 0.00 0.00 0.14 0.65 0.00 0.12
-loyalty 0.09 0.00 0.00 0.07 0.10 0.01 0.49 0.02 0.03 0.20
-sanctity 0.02 0.01 0.00 0.00 0.09 0.00 0.15 0.02 0.00 0.71
+authority
+care
+fairness
+loyalty
+sanctity
-authority
-care
-fairness
-loyalty
-sanctity
Predicted label