In [1]:
```
import json
import warnings
warnings.simplefilter('ignore')
from core.analysis import *
```

```
<Figure size 432x288 with 0 Axes>
```

Data Prep

In [2]:
```
db = access_db()
```

In [3]:
```
trials = get_trial_data(db)
```

In [4]:
```
participants = get_participant_data(db)
```

In [5]:
```
wids = [27046826551,67313721753,14715705020,65643517184,86979612220,494
#Worker IDs of subjects who completed expt
wids = [str(i) for i in wids]
```

Getting data for users who completed the experiment

In [6]:
```
user_df = participants[participants['workerID'].isin(wids) & participan
user_df
```

Out[6]:

	userID	workerID	userIP	completedTask	timeTaken
22	-M3pVTw9ili6muvdZKg4	27046826551	-188934845	1	546.814
23	-M3qqejnlf4v_ja8eE7n	67313721753	727219182	1	526.296
29	-M3rSxyqKh87n8IxCNFz	14715705020	1146968680	1	597.604
31	-M3tou9y2NuJSM2wEped	65643517184	-15487531	1	712.426
33	-M3wWKZuvM6607THGsj3	86979612220	-1107186255	1	308.230
34	-M3xbEdg3XIZca4MA7SZ	26312944991	-1470029051	1	924.305
36	-M40VLLla6hlhfPLwyWk	4949336941	-1012892879	1	921.908
55	-M4RJU1sPJwCqxkN93qZ	46875911951	1225707233	1	706.181
58	-M4WPYQVe_g2Cz7JGkvf	43605166255	565906706	1	380.887

In [7]:
```
users = participants[participants['workerID'].isin(wids) & participants
results = trials[trials['userID'].isin(users)]
results
```

Out[7]:

	trialID	userID	trialIndex	trialType	prevChanged	lemma	
206	-M3pVhmrS4LVzC6_glcJ	-M3pVTw9ili6muvdZKg4	1	training	1	bank_n	
207	-M3pVhmrS4LVzC6_glcJ	-M3pVTw9ili6muvdZKg4	1	training	1	bank_n	
208	-M3pVhmrS4LVzC6_glcJ	-M3pVTw9ili6muvdZKg4	1	training	1	bank_n	dep
209	-M3pVmug0DJj2iajKInK	-M3pVTw9ili6muvdZKg4	2	training	0	bass_n	
210	-M3pVmug0DJj2iajKInK	-M3pVTw9ili6muvdZKg4	2	training	0	bass_n	
...	...	...	...	...	...	...	
775	-M4WQil33JL5IUOVqtkX	-M4WPYQVe_g2Cz7JGkvf	17	repeat	1	book_n	
776	-M4WQil33JL5IUOVqtkX	-M4WPYQVe_g2Cz7JGkvf	17	repeat	1	book_n	
777	-M4WQktKprTLtiCVjEuZ	-M4WPYQVe_g2Cz7JGkvf	18	repeat	0	glass_n	
778	-M4WQktKprTLtiCVjEuZ	-M4WPYQVe_g2Cz7JGkvf	18	repeat	0	glass_n	
779	-M4WQktKprTLtiCVjEuZ	-M4WPYQVe_g2Cz7JGkvf	18	repeat	0	glass_n	

562 rows × 9 columns

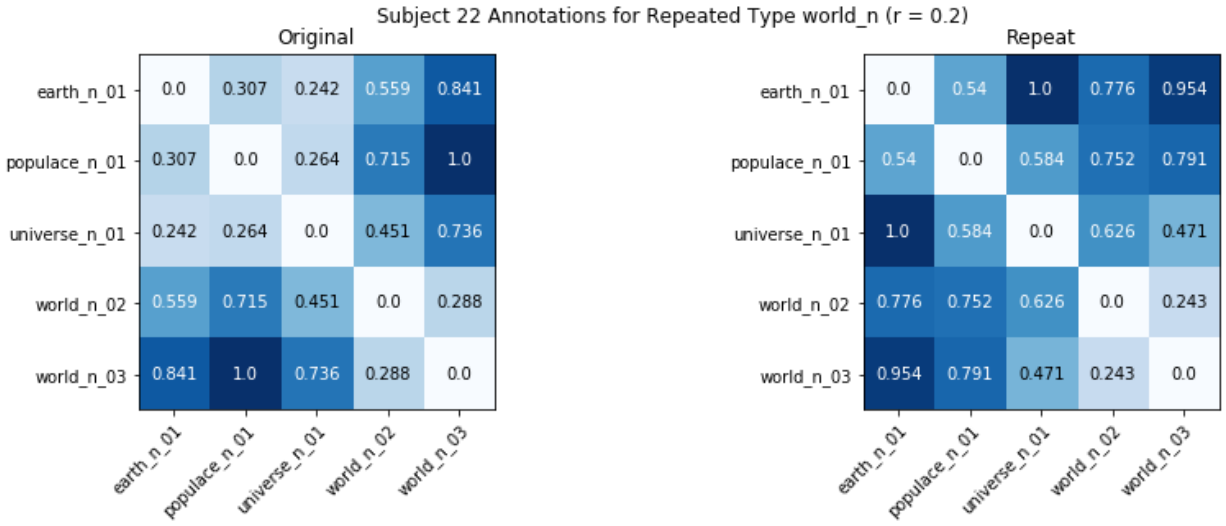
Repeat Trial Results

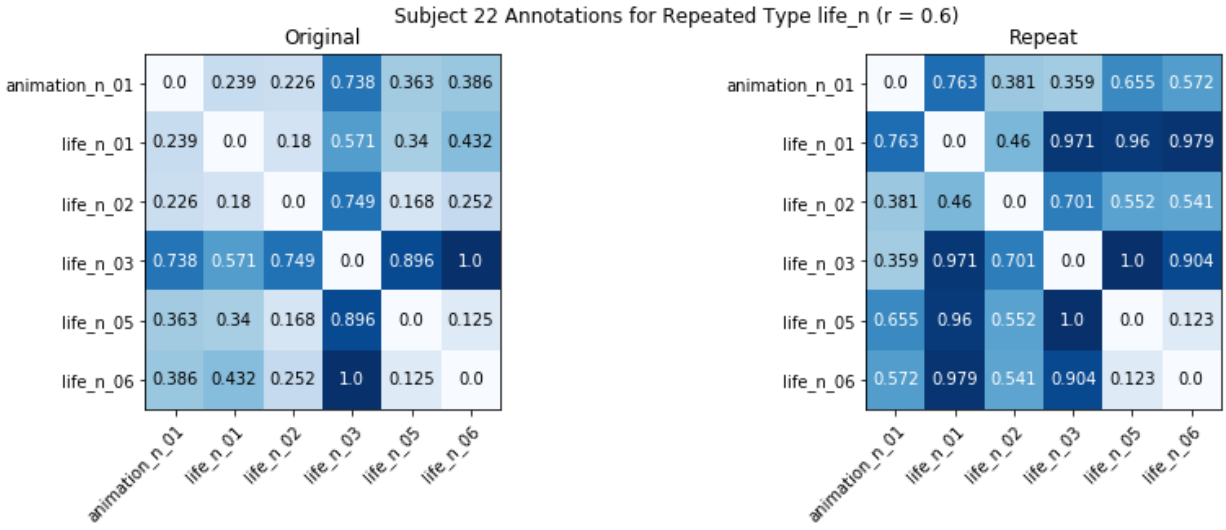
Original WN sense definitions(make sure to click "Show Sense Numbers" under "Display Options"): http://wordnetweb.princeton.edu/perl/webwn (http://wordnetweb.princeton.edu/perl/webwn)

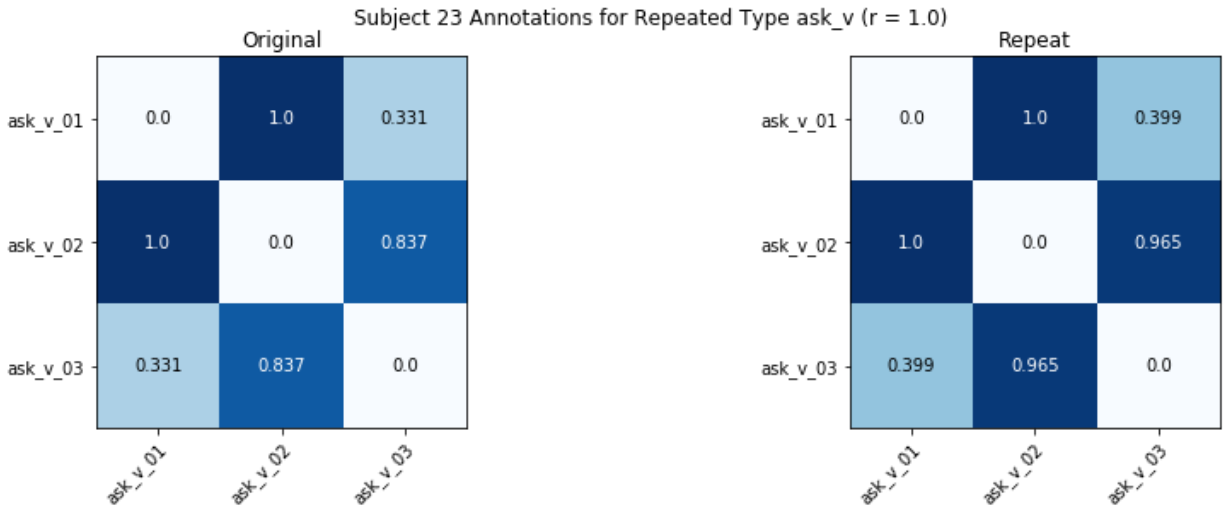
Distance matrices for when a subject got the same word (2x per trial). Spearman rank correlations between the subject's original and repeat trials are also recorded.

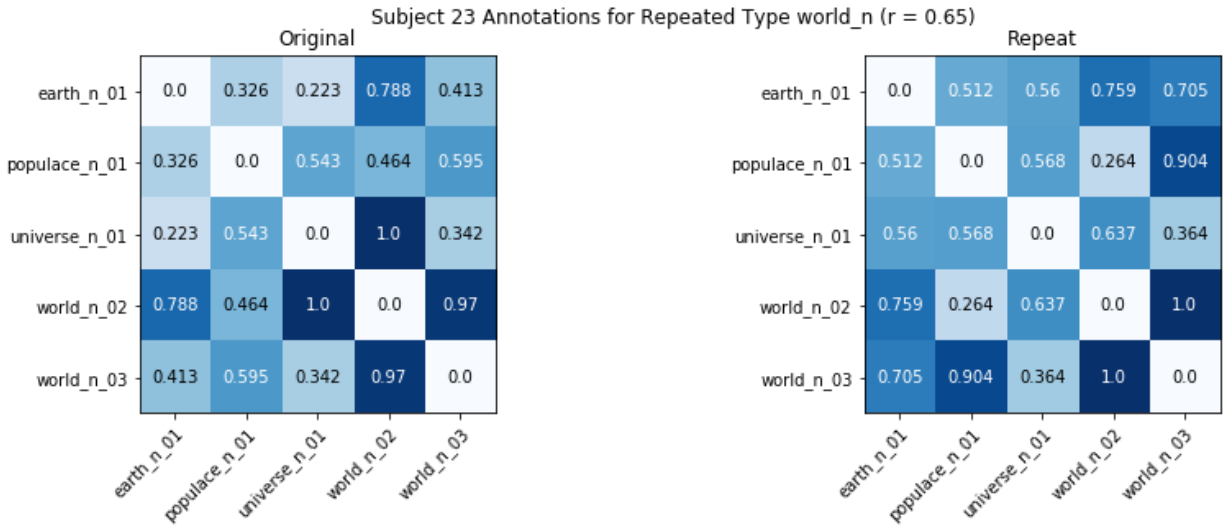
In [11]:
```
1 repeat_corr = plot_all_repeats(results, users)
```

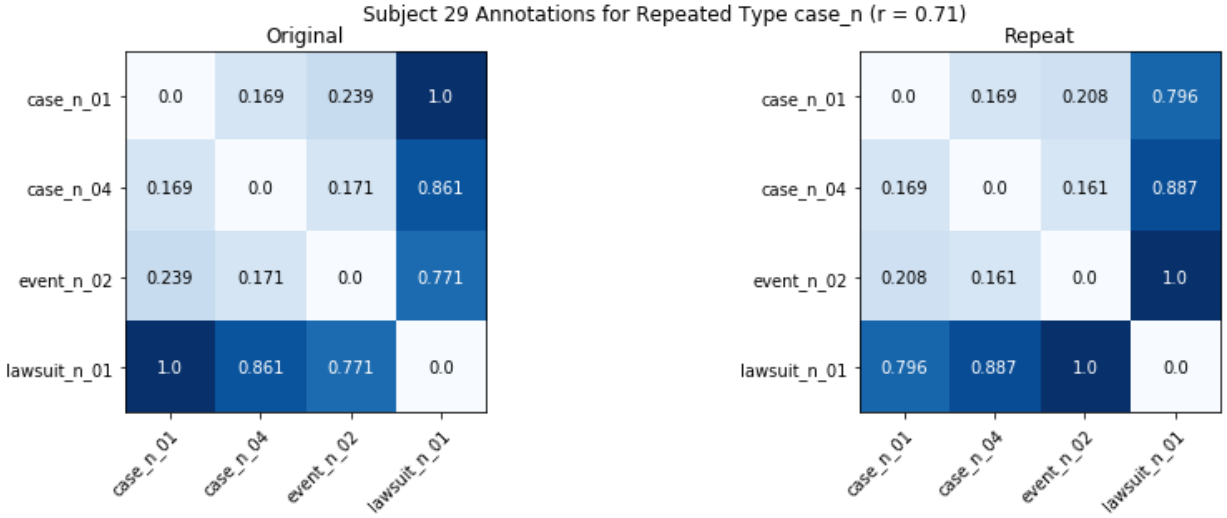
```
User 22  Correlation  0.45940746440938823
User 23  Correlation  0.7341597796143251
User 29  Correlation  0.6554621848739496
User 31  Correlation  0.8508410434878082
User 33  Correlation  -0.2100840336134454
User 34  Correlation  0.34210526315789475
User 36  Correlation  0.6470588235294118
User 55  Correlation  0.2162534435261708
User 58  Correlation  0.3676470588235294
Correlation of all original vs. repeat trials 0.4429744199162231
```

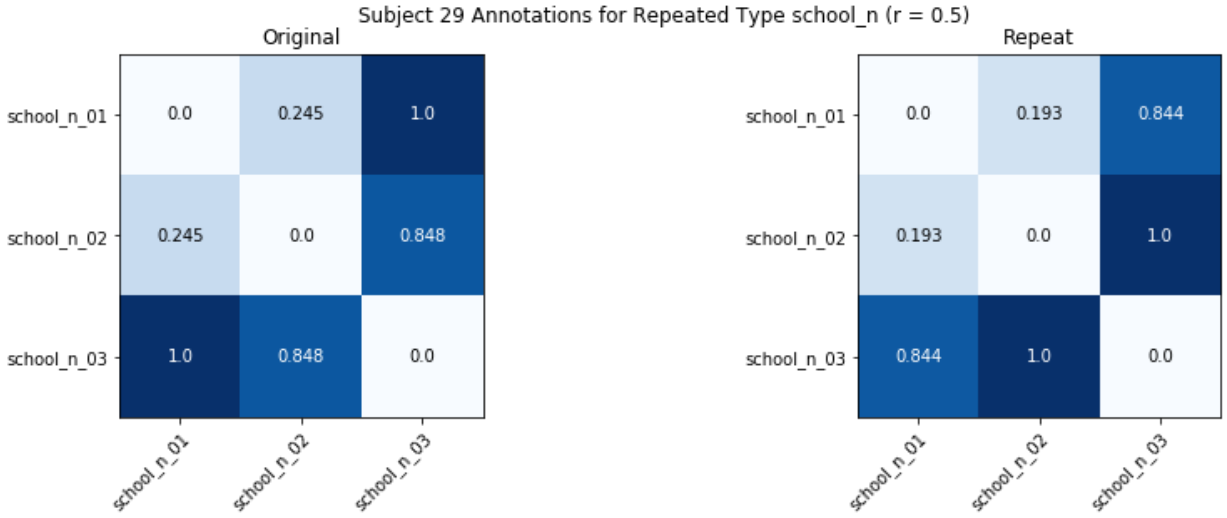


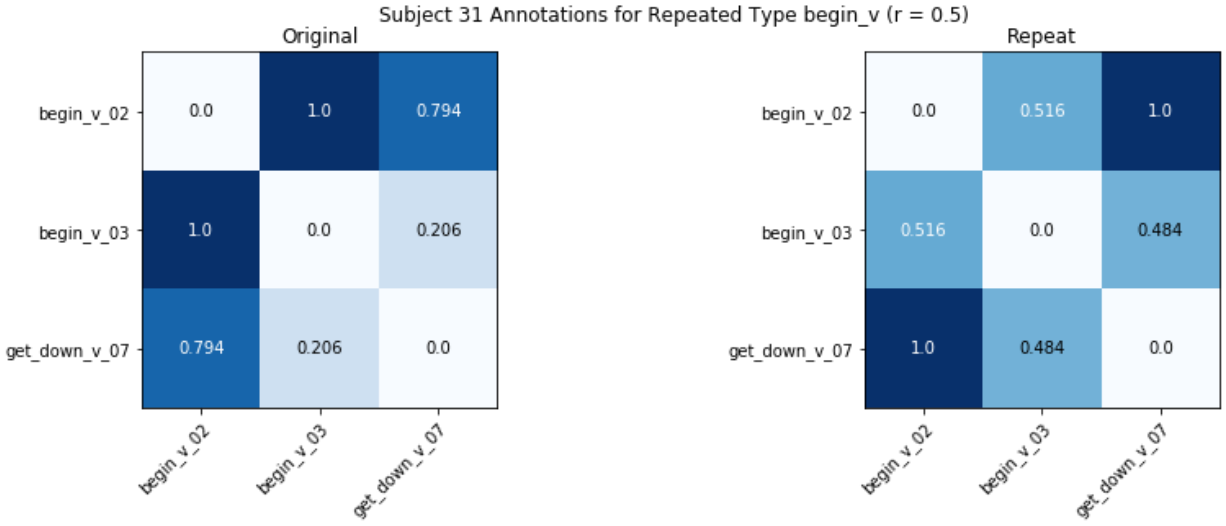


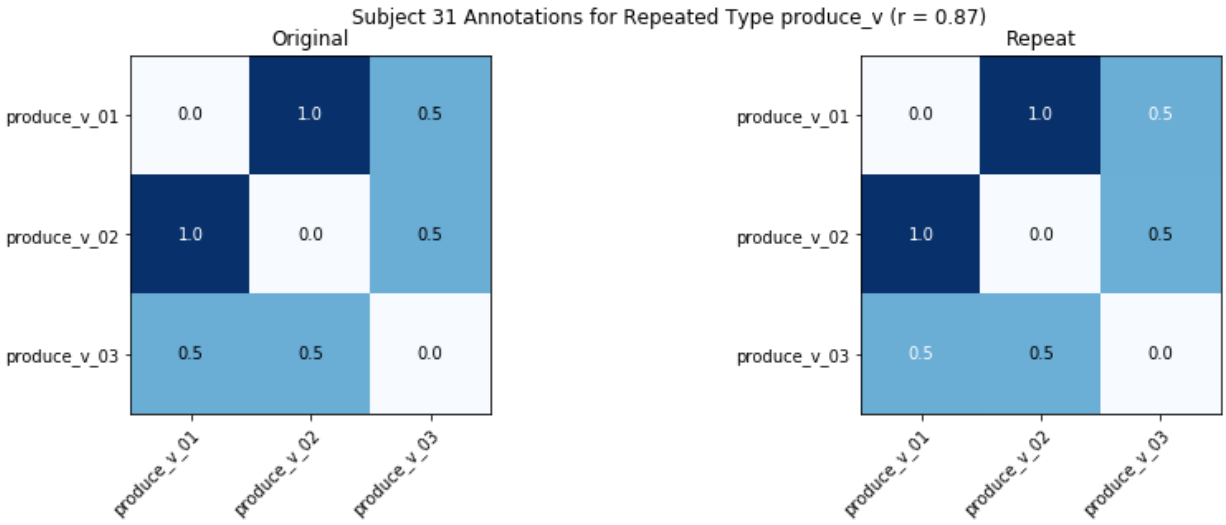

Subject 23 Annotations for Repeated Type ask_v (r = 1.0)
Original
ask_v_01
ask_v_02
ask_v_03
0.0
1.0
0.331
1.0
0.0
0.837
0.331
0.837
0.0
Repeat
ask_v_01
ask_v_02
ask_v_03
0.0
1.0
0.399
1.0
0.0
0.965
0.399
0.965
0.0

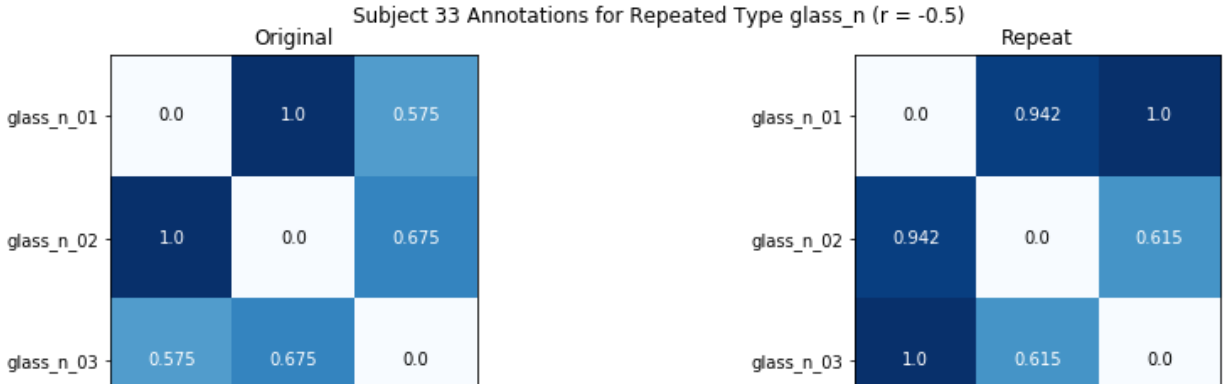

Subject 23 Annotations for Repeated Type world_n (r = 0.65)
Original
earth_n_01
populace_n_01
universe_n_01
world_n_02
world_n_03
0.0 0.326 0.223 0.788 0.413
0.326 0.0 0.543 0.464 0.595
0.223 0.543 0.0 1.0 0.342
0.788 0.464 1.0 0.0 0.97
0.413 0.595 0.342 0.97 0.0
Repeat
earth_n_01
populace_n_01
universe_n_01
world_n_02
world_n_03
0.0 0.512 0.56 0.759 0.705
0.512 0.0 0.568 0.264 0.904
0.56 0.568 0.0 0.637 0.364
0.759 0.264 0.637 0.0 1.0
0.705 0.904 0.364 1.0 0.0

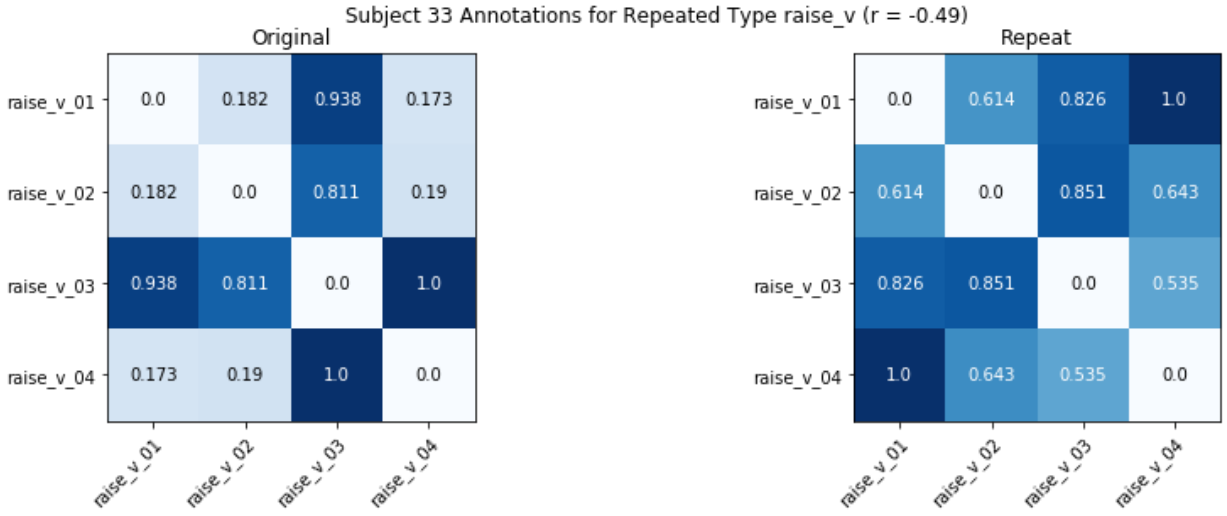

Subject 29 Annotations for Repeated Type case_n (r = 0.71)
Original
case_n_01
case_n_04
event_n_02
lawsuit_n_01
0.0 0.169 0.239 1.0
0.169 0.0 0.171 0.861
0.239 0.171 0.0 0.771
1.0 0.861 0.771 0.0
Repeat
0.0 0.169 0.208 0.796
0.169 0.0 0.161 0.887
0.208 0.161 0.0 1.0
0.796 0.887 1.0 0.0

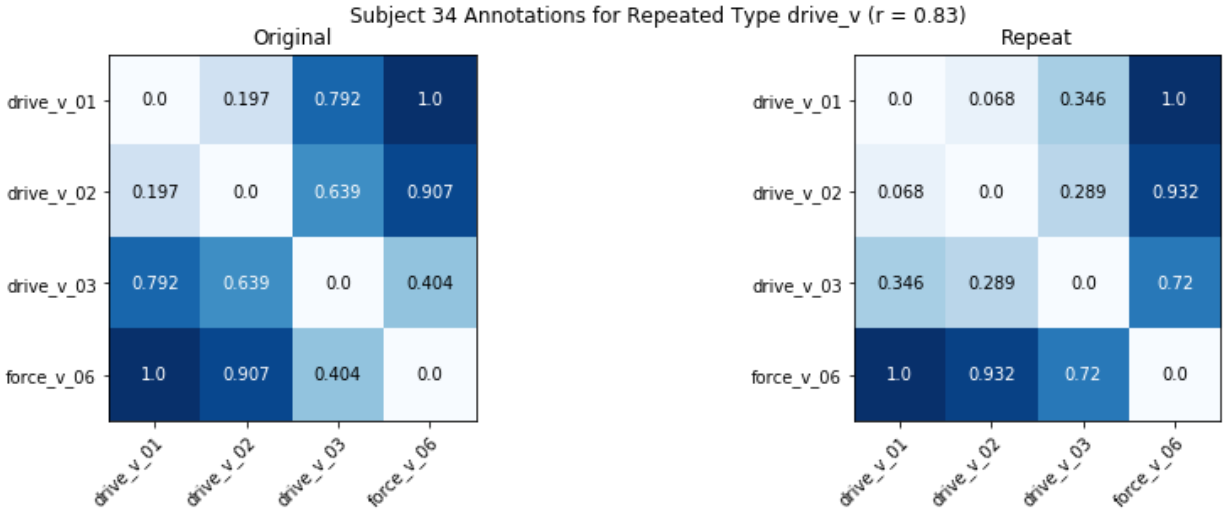

Subject 29 Annotations for Repeated Type school_n (r = 0.5)
Original
school_n_01
school_n_02
school_n_03
0.0 0.245 1.0
0.245 0.0 0.848
1.0 0.848 0.0
Repeat
0.0 0.193 0.844
0.193 0.0 1.0
0.844 1.0 0.0

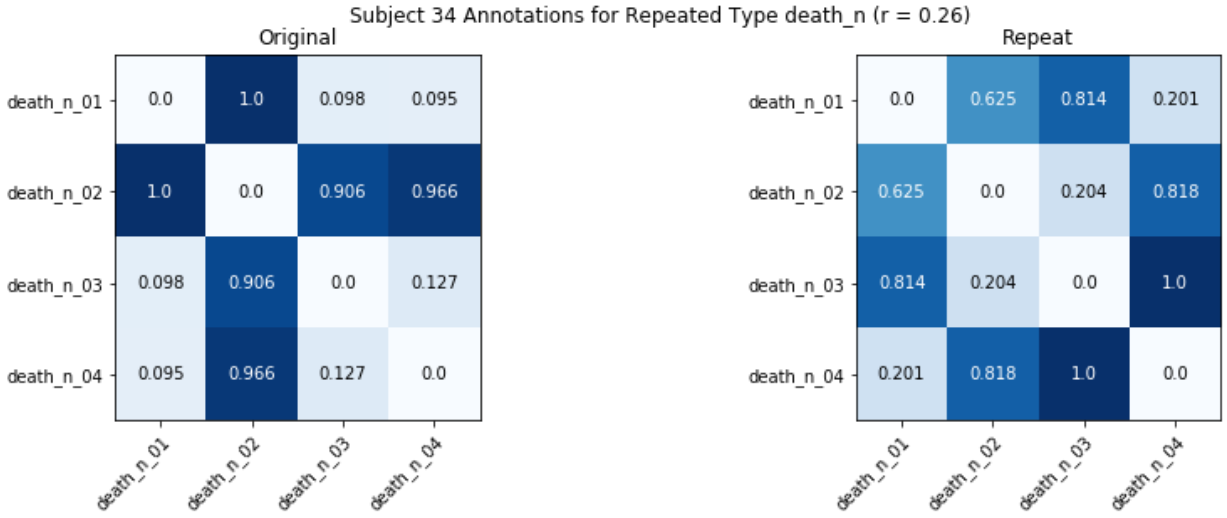

Subject 31 Annotations for Repeated Type begin_v (r = 0.5)
Original
Repeat
begin_v_02
begin_v_03
get_down_v_07
0.0
1.0
0.794
1.0
0.0
0.206
0.794
0.206
0.0
0.516
1.0
0.516
0.0
0.484
1.0
0.484
0.0

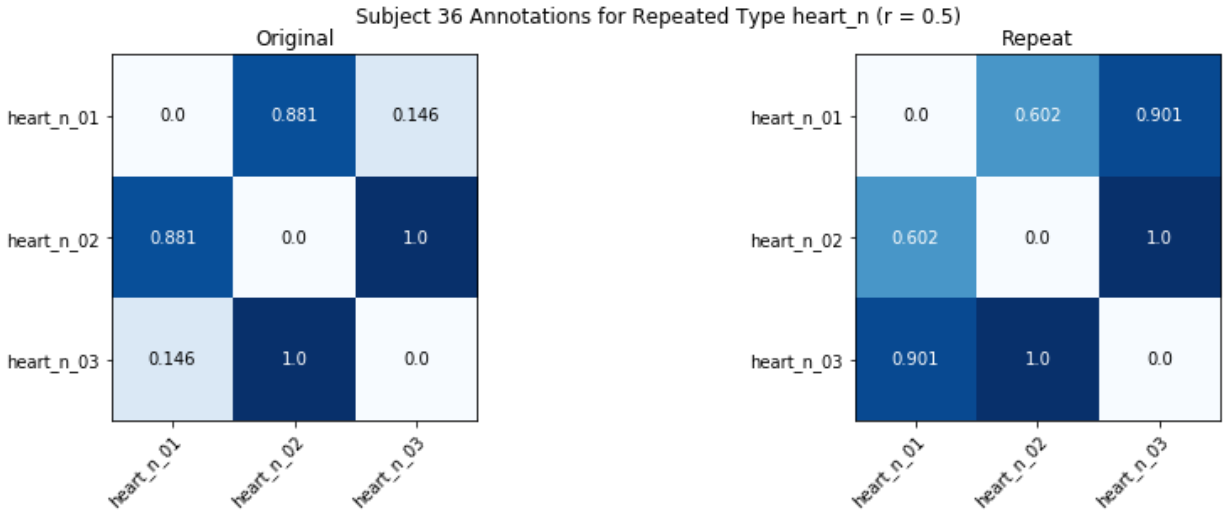

Subject 31 Annotations for Repeated Type produce_v (r = 0.87)
Original
Repeat
produce_v_01
produce_v_02
produce_v_03
0.0
1.0
0.5
1.0
0.0
0.5
0.5
0.5
0.0

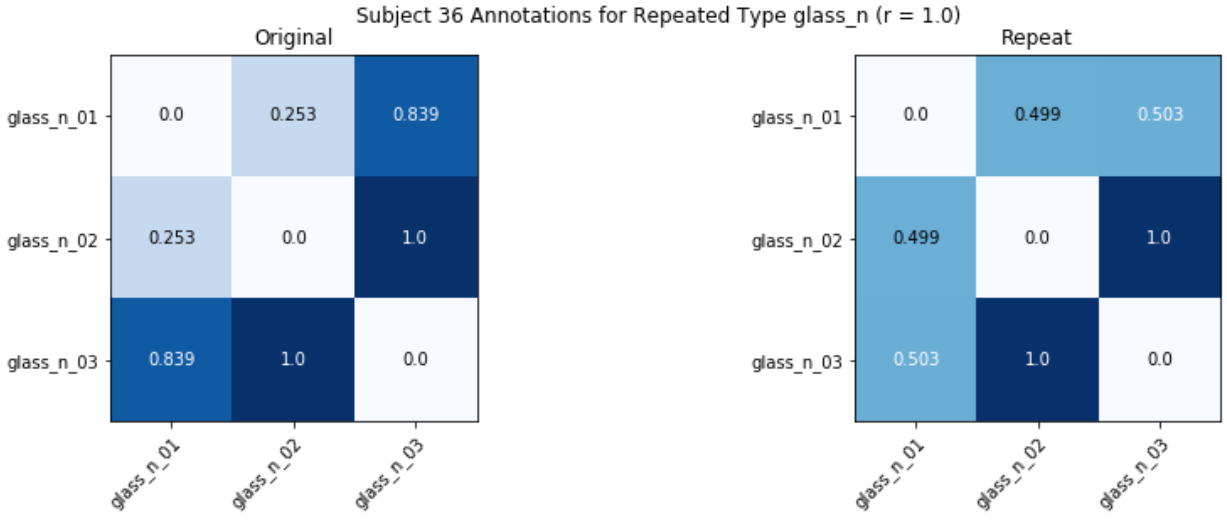

Subject 33 Annotations for Repeated Type glass_n (r = -0.5)
Original
glass_n_01
glass_n_02
glass_n_03
0.0
1.0
0.575
1.0
0.0
0.675
0.575
0.675
0.0
Repeat
glass_n_01
glass_n_02
glass_n_03
0.0
0.942
1.0
0.942
0.0
0.615
1.0
0.615
0.0

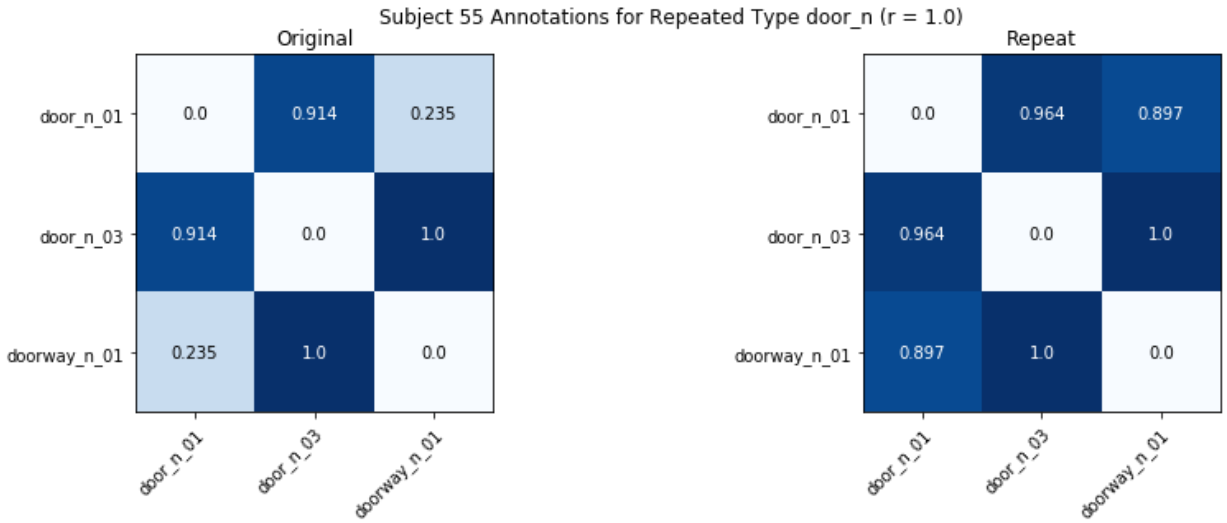

Subject 33 Annotations for Repeated Type raise_v (r = -0.49)
Original
raise_v_01
raise_v_02
raise_v_03
raise_v_04
0.0
0.182
0.938
0.173
0.182
0.0
0.811
0.19
0.938
0.811
0.0
1.0
0.173
0.19
1.0
0.0
raise_v_01
raise_v_02
raise_v_03
raise_v_04
Repeat
raise_v_01
raise_v_02
raise_v_03
raise_v_04
0.0
0.614
0.826
1.0
0.614
0.0
0.851
0.643
0.826
0.851
0.0
0.535
1.0
0.643
0.535
0.0
raise_v_01
raise_v_02
raise_v_03
raise_v_04

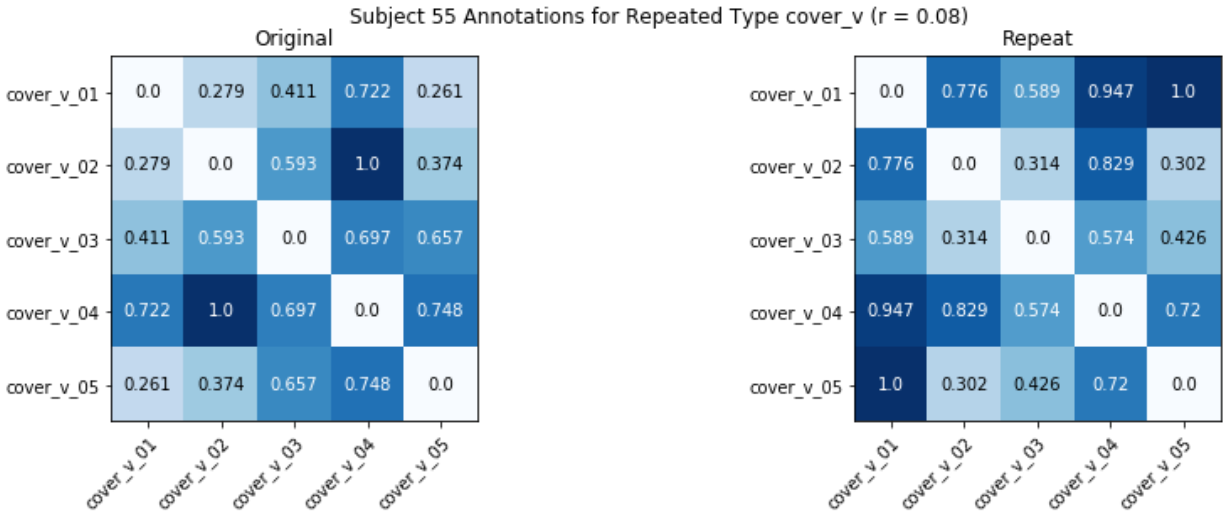

Subject 34 Annotations for Repeated Type drive_v (r = 0.83)
Original
drive_v_01
drive_v_02
drive_v_03
force_v_06
0.0
0.197
0.792
1.0
0.197
0.0
0.639
0.907
0.792
0.639
0.0
0.404
1.0
0.907
0.404
0.0
drive_v_01
drive_v_02
drive_v_03
force_v_06
Repeat
drive_v_01
drive_v_02
drive_v_03
force_v_06
0.0
0.068
0.346
1.0
0.068
0.0
0.289
0.932
0.346
0.289
0.0
0.72
1.0
0.932
0.72
0.0
drive_v_01
drive_v_02
drive_v_03
force_v_06

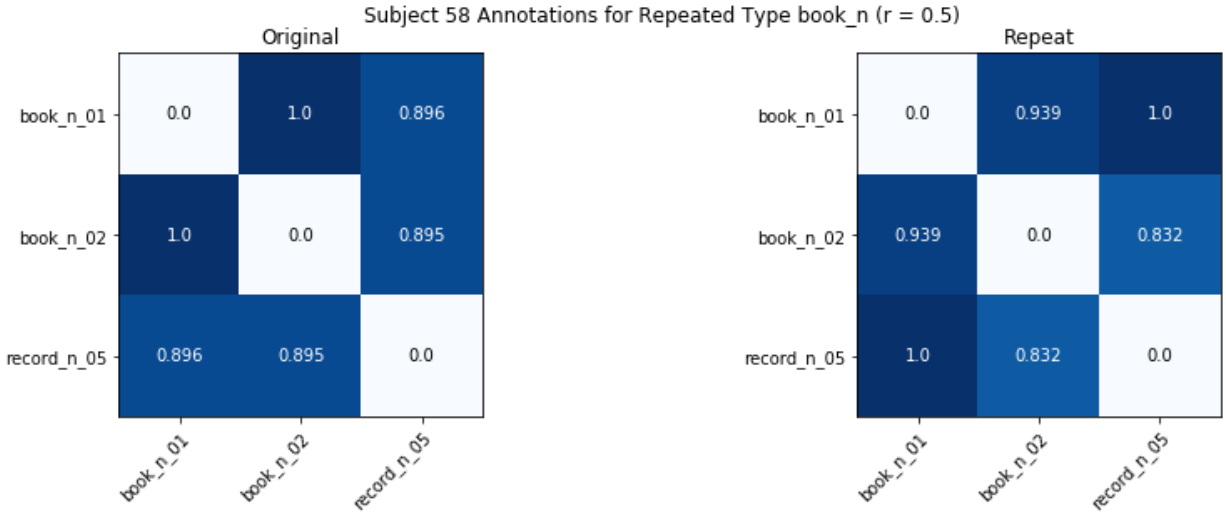

Subject 34 Annotations for Repeated Type death_n (r = 0.26)
Original
death_n_01
death_n_02
death_n_03
death_n_04
0.0 1.0 0.098 0.095
1.0 0.0 0.906 0.966
0.098 0.906 0.0 0.127
0.095 0.966 0.127 0.0
Repeat
0.0 0.625 0.814 0.201
0.625 0.0 0.204 0.818
0.814 0.204 0.0 1.0
0.201 0.818 1.0 0.0

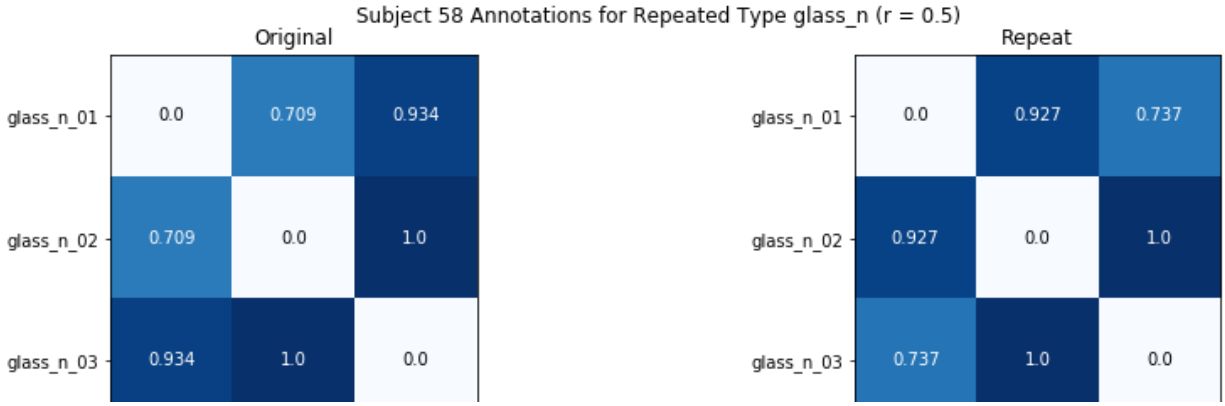

Subject 36 Annotations for Repeated Type heart_n (r = 0.5)
Original
heart_n_01
heart_n_02
heart_n_03
0.0 0.881 0.146
0.881 0.0 1.0
0.146 1.0 0.0
Repeat
0.0 0.602 0.901
0.602 0.0 1.0
0.901 1.0 0.0


Subject 36 Annotations for Repeated Type glass_n (r = 1.0)
Original
Repeat
glass_n_01
glass_n_02
glass_n_03
0.0
0.253
0.839
0.253
0.0
1.0
0.839
1.0
0.0
0.0
0.499
0.503
0.499
0.0
1.0
0.503
1.0
0.0


Subject 55 Annotations for Repeated Type door_n (r = 1.0)
Original
Repeat
door_n_01
door_n_03
doorway_n_01
0.0
0.914
0.235
0.914
0.0
1.0
0.235
1.0
0.0
0.0
0.964
0.897
0.964
0.0
1.0
0.897
1.0
0.0


Subject 55 Annotations for Repeated Type cover_v (r = 0.08)
Original
cover_v_01 0.0 0.279 0.411 0.722 0.261
cover_v_02 0.279 0.0 0.593 1.0 0.374
cover_v_03 0.411 0.593 0.0 0.697 0.657
cover_v_04 0.722 1.0 0.697 0.0 0.748
cover_v_05 0.261 0.374 0.657 0.748 0.0
cover_v_01 cover_v_02 cover_v_03 cover_v_04 cover_v_05
Repeat
cover_v_01 0.0 0.776 0.589 0.947 1.0
cover_v_02 0.776 0.0 0.314 0.829 0.302
cover_v_03 0.589 0.314 0.0 0.574 0.426
cover_v_04 0.947 0.829 0.574 0.0 0.72
cover_v_05 1.0 0.302 0.426 0.72 0.0
cover_v_01 cover_v_02 cover_v_03 cover_v_04 cover_v_05

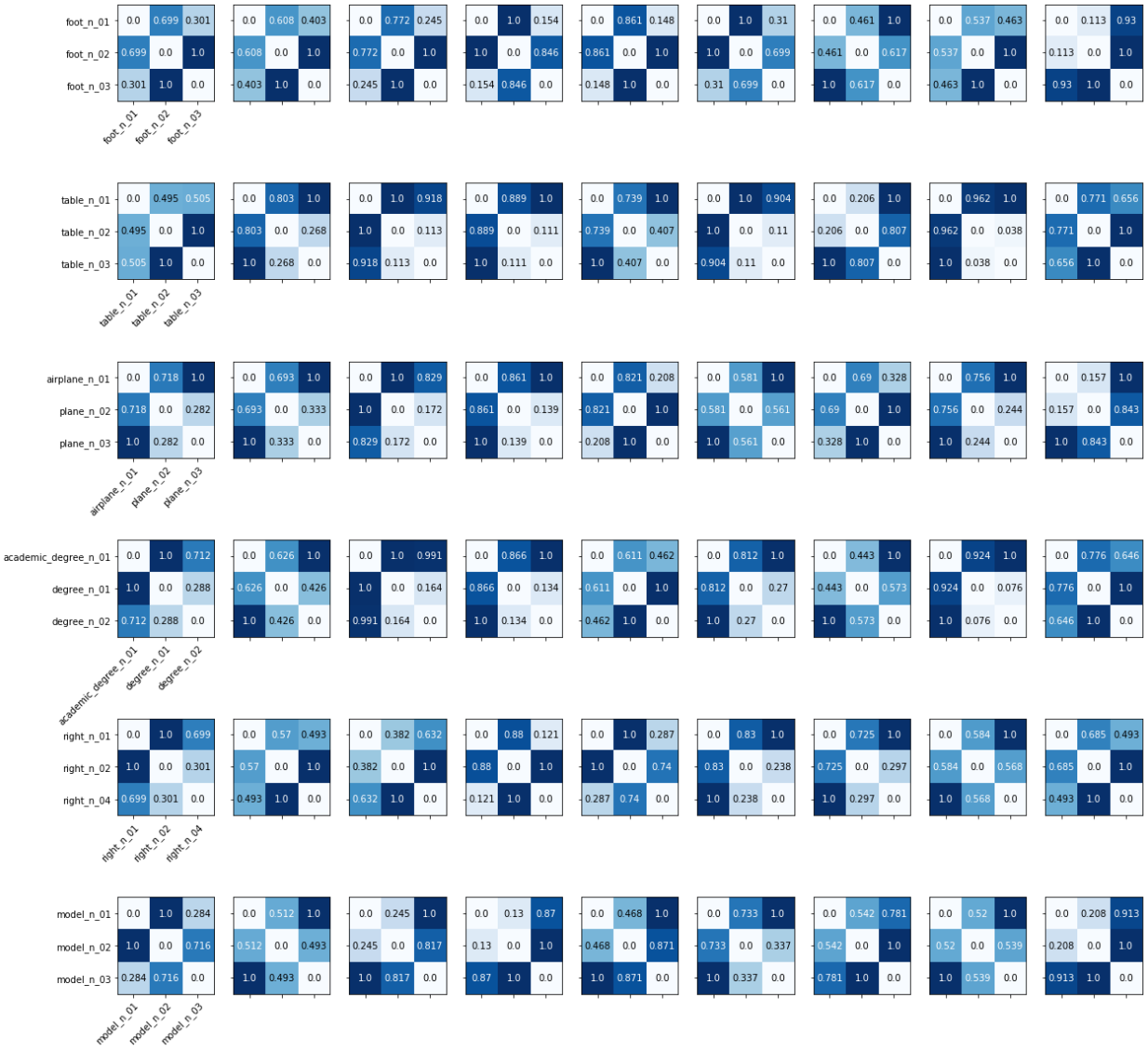

Subject 58 Annotations for Repeated Type book_n (r = 0.5)
Original
book_n_01 0.0 1.0 0.896
book_n_02 1.0 0.0 0.895
record_n_05 0.896 0.895 0.0
book_n_01 book_n_02 record_n_05
Repeat
book_n_01 0.0 0.939 1.0
book_n_02 0.939 0.0 0.832
record_n_05 1.0 0.832 0.0
book_n_01 book_n_02 record_n_05



Shared Words

All subjects received a set of homonymous words. It appears there's some consistency in the results for foot, table, degree, and plane, but model and right were harder to distinguish.

In [12]:
```
1 plot_all_shared(results, users)
```



Spearman rank correlations between one subject's data and the average of the rest of the subjects

In [13]:
```
shared_results = results[results['trialType'] == 'shared']
shared_corrs = group_consistency(shared_results, users)
```

```
Hold One Out Correlation for User 22 -0.027329898660652777
Hold One Out Correlation for User 23 0.767339462395251
Hold One Out Correlation for User 29 0.6685313672375064
Hold One Out Correlation for User 31 0.5444956733160823
Hold One Out Correlation for User 33 0.06622244675465866
Hold One Out Correlation for User 34 0.3973346805279519
Hold One Out Correlation for User 36 -0.245969087945875
Hold One Out Correlation for User 55 0.8167435099741234
Hold One Out Correlation for User 58 -0.23335528856403523
```

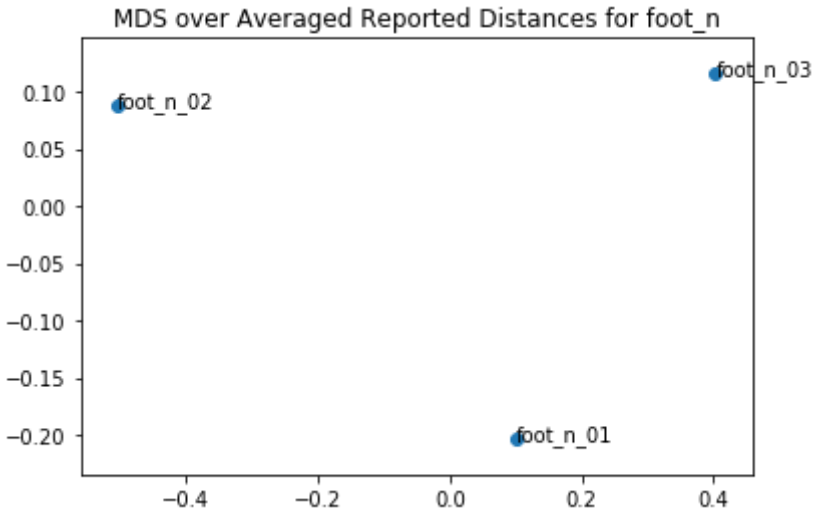
MDS for Shared Trials

In [113]:
```
all_shared_defs = display_sense_definitions(results, 'shared')
```

In [114]:
```
plot_individual_mds(results, 'foot_n', 'shared', users, db, all_shared_
```

Out[114]:

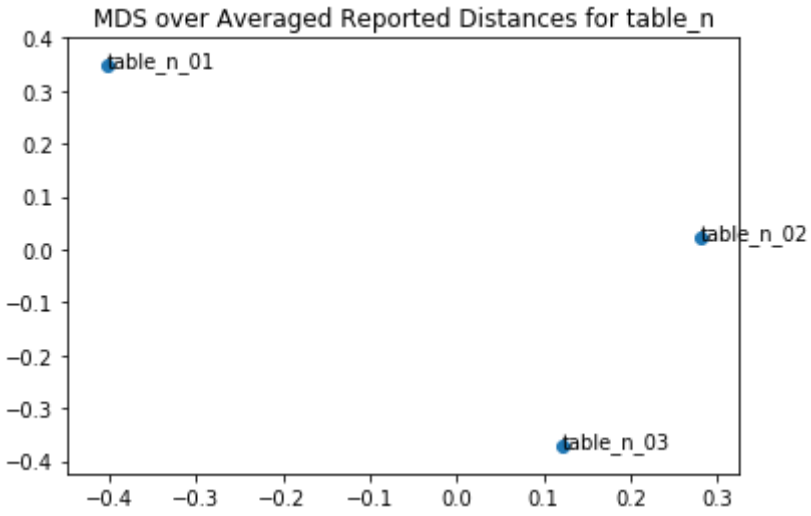
	Sense	Type	Definition
212	foot_n_01	foot_n	the part of the leg of a human being below the ankle joint
213	foot_n_02	foot_n	a linear unit of length equal to 12 inches or a third of a yard
214	foot_n_03	foot_n	the lower part of anything



In [115]:
```
1 plot_individual_mds(results, 'table_n', 'shared', users, db, all_shared
```

Out[115]:

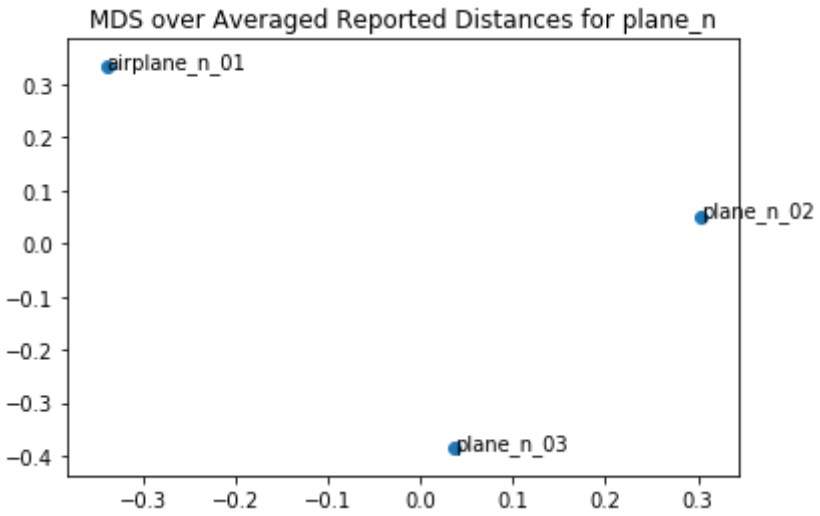
	Sense	Type	Definition
215	table_n_01	table_n	a set of data arranged in rows and columns
216	table_n_02	table_n	a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs
217	table_n_03	table_n	a piece of furniture with tableware for a meal laid out on it



In [116]:
```
1 plot_individual_mds(results, 'plane_n', 'shared', users, db, all_shared
```

Out[116]:

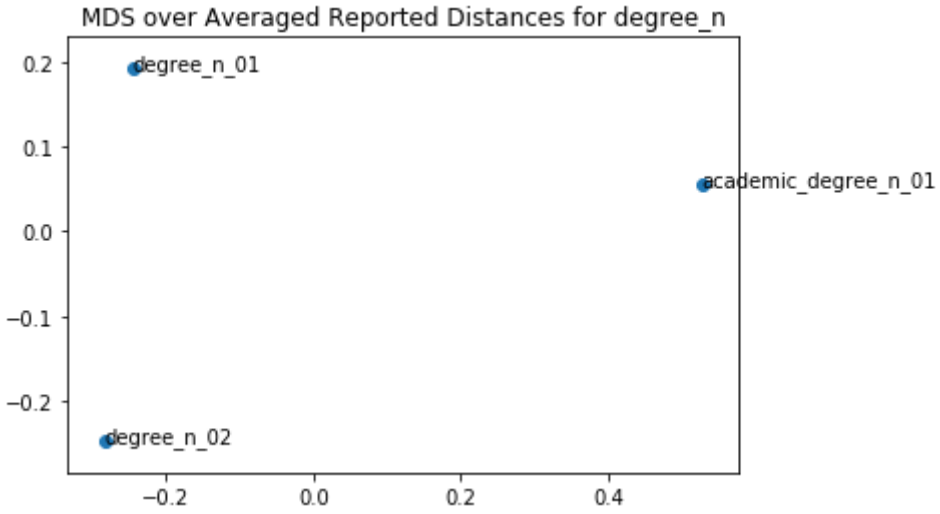
	Sense	Type	Definition
218	airplane_n_01	plane_n	an aircraft that has a fixed wing and is powered by propellers or jets
219	plane_n_02	plane_n	(mathematics) an unbounded two-dimensional shape
220	plane_n_03	plane_n	a level of existence or development



In [117]:
```
1 plot_individual_mds(results, 'degree_n', 'shared', users, db, all_share
```

Out[117]:

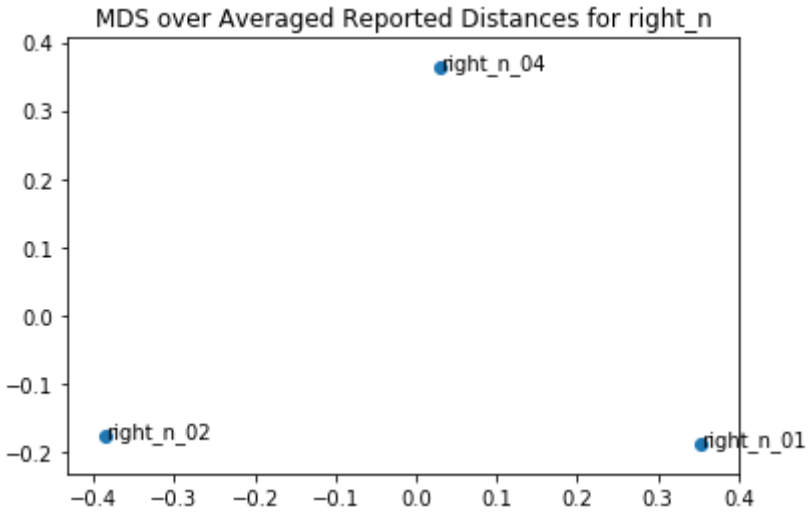
	Sense	Type	Definition
221	academic_degree_n_01	degree_n	an award conferred by a college or university signifying that the recipient has satisfactorily completed a course of study
222	degree_n_01	degree_n	a position on a scale of intensity or amount or quality
223	degree_n_02	degree_n	a specific identifiable position in a continuum or series or especially in a process



In [118]:
```
1 plot_individual_mds(results, 'right_n', 'shared', users, db, all_shared
```

Out[118]:

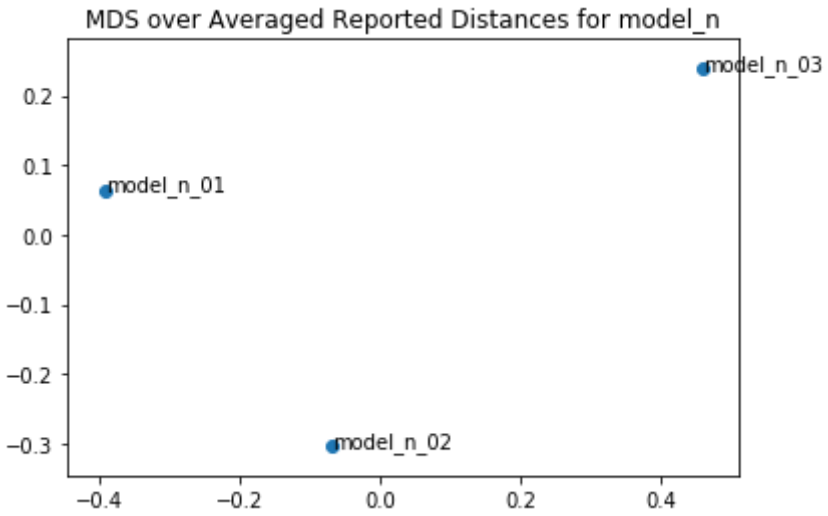
	Sense	Type	Definition
224	right_n_01	right_n	an abstract idea of that which is due to a person or governmental body by law or tradition or nature; ; - Eleanor Roosevelt
225	right_n_02	right_n	location near or direction toward the right side; i.e. the side to the south when a person or object faces east
226	right_n_04	right_n	those who support political or social or economic conservatism; those who believe that things are better left unchanged



In [119]:
```
1 plot_individual_mds(results, 'model_n', 'shared', users, db, all_shared
```

Out[119]:

	Sense	Type	Definition
227	model_n_01	model_n	a hypothetical description of a complex entity or process
228	model_n_02	model_n	a type of product
229	model_n_03	model_n	a person who poses for a photographer or painter or sculptor



Consistency Metrics

In [36]:
```
user_time_word_changes = get_time_and_changes(results, user_df)
consistency = pd.DataFrame({'Group Consistency': shared_corrs, 'Self Co
corrs = user_time_word_changes.merge(consistency, on = user_time_word_c
corrs['Correlation with SN'] = my_correlations(participants, trials, re
corrs
```

Out[36]:

	userID	timeTaken	changes	Group Consistency	Self Consistency	Correlation with SN
0	-M3pVTw9ili6muvdZKg4	546.814	5	-0.027330	0.459407	0.328158
1	-M3qqejnlf4v_ja8eE7n	526.296	3	0.767339	0.734160	0.782120
2	-M3rSxyqKh87n8IxCNFz	597.604	14	0.668531	0.655462	0.571734
3	-M3tou9y2NuJSM2wEped	712.426	10	0.544496	0.850841	0.585118
4	-M3wWKZuvM6607THGsj3	308.230	3	0.066222	-0.210084	0.230193
5	-M3xbEdg3XIZca4MA7SZ	924.305	9	0.397335	0.342105	0.731799
6	-M40VLLla6hlhfPLwyWk	921.908	0	-0.245969	0.647059	-0.115632
7	-M4RJU1sPJwCqxkN93qZ	706.181	11	0.816744	0.216253	0.790685
8	-M4WPYQVe_g2Cz7JGkvf	380.887	3	-0.233355	0.367647	-0.276231

timeTaken: Duration of experiment for each user (in seconds)

Changes: Number of times a subject changed the existing layout (benefit of grid layout over similarity judgements)

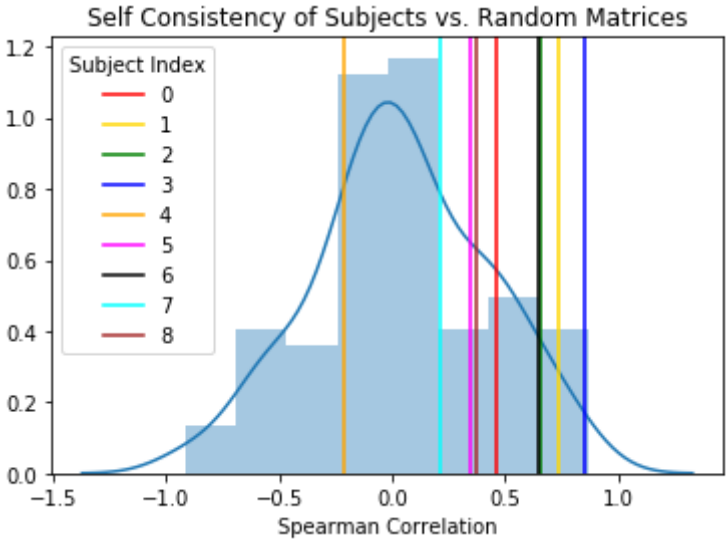
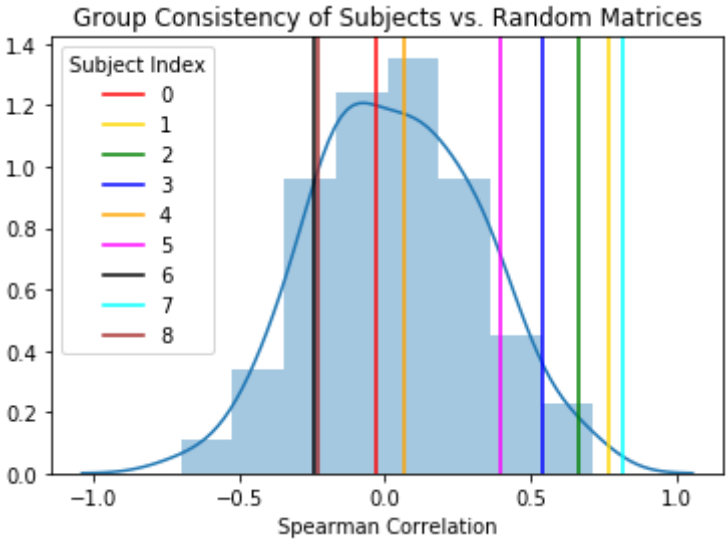
Group Consistency: Hold one out correlation on shared trials (taken from above results). Spearman rank correlation with subject's matrix and the average of all the other subjects' matrices.

Self Consistency: Spearman rank correlation over the two repeated words a participant received

Correlation with SN: The leftmost column in the dataframe is the Spearman rank correlation with my results, which would be the closet thing we have to ground truth data. For all the shared words, my correlation with the average was 1 for plane.n, degree.n, foot.n, and table.n, and 0.5 for right.n and model.n.

We replicate the self and group consistency metrics with matrices initialized for 100 random subjects.

In [54]:
```
random_self = simulate_self_correlation(100, db)
plot_consistency_hist(random_self, corrs['Self Consistency'],
                      "Self Consistency of Subjects vs. Random Matrices
```



In [50]:
```
random_group = [random_vs_all(shared_results) for _ in range(100)]
plot_consistency_hist(random_group, corrs['Group Consistency'],
                      "Group Consistency of Subjects vs. Random Matrice
```