
Psifr

Release v0.8.2

Neal Morton

Sep 15, 2022

CONTENTS

1	Installation	3
2	User guide	5
2.1	Importing data	5
2.2	Scoring data	6
2.3	Recall performance	8
2.4	Recall order	11
2.5	Comparing conditions	18
3	Tutorials	21
4	API reference	23
4.1	Free recall analysis	23
4.2	Measures	49
4.3	Transitions	51
4.4	Outputs	62
5	Development	65
5.1	Transitions	65
6	References	69
	Bibliography	71
	Index	73

In free recall, participants study a list of items and then name all of the items they can remember in any order they choose. Many sophisticated analyses have been developed to analyze data from free recall experiments, but these analyses are often complicated and difficult to implement.

Psifr leverages the Pandas data analysis package to make precise and flexible analysis of free recall data faster and easier.

If you use Psifr, please cite the paper [\[Mor20\]](#).

See the code repository for version [release notes](#).

INSTALLATION

You can install the latest stable version of Psifr using pip:

```
pip install psifr
```

You can also install the development version directly from the code repository on GitHub:

```
pip install git+https://github.com/mortonne/psifr
```


2.1 Importing data

In Psifr, free recall data are imported in the form of a “long” format table. Each row corresponds to one *study* or *recall* event. Study events include any time an item was presented to the participant. Recall events correspond to any recall attempt; this includes *repeats* of items there were already recalled and *intrusions* of items that were not present in the study list.

This type of information is well represented in a CSV spreadsheet, though any file format supported by pandas may be used for input. To import from a CSV, use `pandas.read_csv()`. For example:

```
import pandas as pd
data = pd.read_csv("my_data.csv")
```

2.1.1 Trial information

The basic information that must be included for each event is the following:

subject

Some code (numeric or string) indicating individual participants. Must be unique for a given experiment. For example, sub-101.

list

Numeric code indicating individual lists. Must be unique within subject.

trial_type

String indicating whether each event is a *study* event or a *recall* event.

position

Integer indicating position within a given phase of the list. For *study* events, this corresponds to *input position* (also referred to as *serial position*). For *recall* events, this corresponds to *output position*.

item

Individual thing being recalled, such as a word. May be specified with text (e.g., pumpkin, Jack Nicholson) or a numeric code (682, 121). Either way, the text or number must be unique to that item. Text is easier to read and does not require any additional information for interpretation and is therefore preferred if available.

2.1.2 Example

Table 1: Sample data

subject	list	trial_type	position	item
1	1	study	1	absence
1	1	study	2	hollow
1	1	study	3	pupil
1	1	recall	1	pupil
1	1	recall	2	absence

2.1.3 Additional information

Additional fields may be included in the data to indicate other aspects of the experiment, such as presentation time, stimulus category, experimental session, distraction length, etc. All of these fields can then be used for analysis in Psifr.

2.2 Scoring data

After *importing free recall data*, we have a DataFrame with a row for each study event and a row for each recall event. Next, we need to score the data by matching study events with recall events.

2.2.1 Scoring list recall

First, let's create a simple sample dataset with two lists. We can use the `table_from_lists()` convenience function to create a sample dataset with a given set of study lists and recalls:

```
In [1]: from psifr import fr

In [2]: list_subject = [1, 1]

In [3]: study_lists = [['absence', 'hollow', 'pupil'], ['fountain', 'piano', 'pillow']]

In [4]: recall_lists = [['pupil', 'absence', 'empty'], ['pillow', 'pupil', 'pillow']]

In [5]: data = fr.table_from_lists(list_subject, study_lists, recall_lists)

In [6]: data
Out[6]:
```

	subject	list	trial_type	position	item
0	1	1	study	1	absence
1	1	1	study	2	hollow
2	1	1	study	3	pupil
3	1	1	recall	1	pupil
4	1	1	recall	2	absence
5	1	1	recall	3	empty
6	1	2	study	1	fountain
7	1	2	study	2	piano
8	1	2	study	3	pillow
9	1	2	recall	1	pillow

(continues on next page)

(continued from previous page)

10	1	2	recall	2	pupil
11	1	2	recall	3	pillow

Next, we'll merge together the study and recall events by matching up corresponding events using `merge_free_recall()`. This scoring and merging step labels recall attempts in terms of whether they were correct recalls, repeats, or intrusions. At the same time, it also labels study events in terms of whether they were correctly recalled, and, if so, at which output position they were recalled. Free-recall analyses in Psifr are computed from data in this "merged" format.

```
In [7]: merged = fr.merge_free_recall(data)

In [8]: merged
Out[8]:
```

	subject	list	item	input	...	repeat	intrusion	prior_list	prior_input
0	1	1	absence	1.0	...	0	False	NaN	NaN
1	1	1	hollow	2.0	...	0	False	NaN	NaN
2	1	1	pupil	3.0	...	0	False	NaN	NaN
3	1	1	empty	NaN	...	0	True	NaN	NaN
4	1	2	fountain	1.0	...	0	False	NaN	NaN
5	1	2	piano	2.0	...	0	False	NaN	NaN
6	1	2	pillow	3.0	...	0	False	NaN	NaN
7	1	2	pillow	3.0	...	1	False	NaN	NaN
8	1	2	pupil	NaN	...	0	True	1.0	3.0

```
[9 rows x 11 columns]
```

For each item, there is one row for each unique combination of input and output position. For example, if an item is presented once in the list, but is recalled multiple times, there is one row for each of the recall attempts. Repeated recalls are indicated by the `repeat` column, which is greater than zero for recalls of an item after the first. Unique study events are indicated by the `study` column; this excludes intrusions and repeated recalls.

Items that were not recalled have the `recall` column set to `False`. Because they were not recalled, they have no defined output position, so `output` is set to `NaN`. Finally, intrusions have an output position but no input position because they did not appear in the list. There is an `intrusion` field for convenience to label these recall attempts. The `prior_list` and `prior_input` fields give information about prior-list intrusions (PLIs) of items from prior lists. The `prior_list` field gives the list where the item appeared and `prior_input` indicates the position in which it was presented on that list.

`merge_free_recall()` can also handle additional attributes beyond the standard ones, such as codes indicating stimulus category or list condition. See [Working with custom columns](#) for details.

2.2.2 Filtering and sorting

Now that we have a merged `DataFrame`, we can use Pandas methods to quickly get different views of the data. For some analyses, we may want to organize in terms of the study list by removing repeats and intrusions. Because our data are in a `DataFrame`, we can use the `query()` method:

```
In [9]: merged.query('study')
Out[9]:
```

	subject	list	item	input	...	repeat	intrusion	prior_list	prior_input
0	1	1	absence	1.0	...	0	False	NaN	NaN
1	1	1	hollow	2.0	...	0	False	NaN	NaN

(continues on next page)

(continued from previous page)

2	1	1	pupil	3.0	...	0	False	NaN	NaN
4	1	2	fountain	1.0	...	0	False	NaN	NaN
5	1	2	piano	2.0	...	0	False	NaN	NaN
6	1	2	pillow	3.0	...	0	False	NaN	NaN

[6 rows x 11 columns]

Alternatively, we may also want to get just the recall events, sorted by output position instead of input position:

```
In [10]: merged.query('recall').sort_values(['list', 'output'])
Out[10]:
```

	subject	list	item	input	...	repeat	intrusion	prior_list	prior_input
2	1	1	pupil	3.0	...	0	False	NaN	NaN
0	1	1	absence	1.0	...	0	False	NaN	NaN
3	1	1	empty	NaN	...	0	True	NaN	NaN
6	1	2	pillow	3.0	...	0	False	NaN	NaN
8	1	2	pupil	NaN	...	0	True	1.0	3.0
7	1	2	pillow	3.0	...	1	False	NaN	NaN

[6 rows x 11 columns]

Note that we first sort by list, then output position, to keep the lists together.

In addition to using the `query()` method directly, we can also use `filter_data()` to get subsets of data. For example, to get the first list only:

```
In [11]: fr.filter_data(merged, lists=1)
Out[11]:
```

	subject	list	item	input	...	repeat	intrusion	prior_list	prior_input
0	1	1	absence	1.0	...	0	False	NaN	NaN
1	1	1	hollow	2.0	...	0	False	NaN	NaN
2	1	1	pupil	3.0	...	0	False	NaN	NaN
3	1	1	empty	NaN	...	0	True	NaN	NaN

[4 rows x 11 columns]

2.3 Recall performance

First, load some sample data and create a merged DataFrame:

```
In [1]: from psifr import fr

In [2]: df = fr.sample_data('Morton2013')

In [3]: data = fr.merge_free_recall(df)
```

2.3.1 Raster plot

Raster plots can give you a quick overview of a whole dataset [RKT16]. We'll look at all of the first subject's recalls using `plot_raster()`. This will plot every individual recall, colored by the serial position of the recalled item in the list. Items near the end of the list are shown in yellow, and items near the beginning of the list are shown in purple. Intrusions of items not on the list are shown in red.

```
In [4]: subj = fr.filter_data(data, 1)

In [5]: g = fr.plot_raster(subj).add_legend()
```

2.3.2 Serial position curve

We can calculate average recall for each serial position [Mur62] using `spc()` and plot using `plot_spc()`.

```
In [6]: recall = fr.spc(data)

In [7]: g = fr.plot_spc(recall)
```

Using the same plotting function, we can plot the curve for each individual subject:

```
In [8]: g = fr.plot_spc(recall, col='subject', col_wrap=5)
```

2.3.3 Probability of Nth recall

We can also split up recalls, to test for example how likely participants were to initiate recall with the last item on the list, using `pnr()`.

```
In [9]: prob = fr.pnr(data)

In [10]: prob
Out[10]:
```

subject	output	input	prob	actual	possible
1	1	1	0.000000	0	48
		2	0.020833	1	48
		3	0.000000	0	48
		4	0.000000	0	48
		5	0.000000	0	48
...		
47	24	20	NaN	0	0
		21	NaN	0	0
		22	NaN	0	0
		23	NaN	0	0
		24	NaN	0	0

[23040 rows x 3 columns]

This gives us the probability of recall by output position ('output') and serial or input position ('input'). This is a lot to look at all at once, so it may be useful to plot just the first three output positions. We can plot the curves using `plot_spc()`, which takes an optional hue input to specify a variable to use to split the data into curves of different colors.

```
In [11]: pfr = prob.query('output <= 3')
In [12]: g = fr.plot_spc(pfr, hue='output').add_legend()
```

This plot shows what items tend to be recalled early in the recall sequence.

2.3.4 Prior-list intrusions

Participants will sometimes accidentally recall items from prior lists; these recalls are known as prior-list intrusions (PLIs). To better understand how prior-list intrusions are happening, you can look at how many lists back those items were originally presented using `pli_list_lag()`.

First, you need to choose a maximum list lag that you will consider. This determines which lists will be included in the analysis. For example, if you have a maximum lag of 3, then the first 3 lists will be excluded from the analysis. This ensures that each included list can potentially have intrusions of each possible list lag.

```
In [13]: pli = fr.pli_list_lag(data, max_lag=3)
In [14]: pli
Out[14]:
```

		count	per_list	prob
subject	list_lag			
1	1	7	0.155556	0.259259
	2	5	0.111111	0.185185
	3	0	0.000000	0.000000
2	1	9	0.200000	0.191489
	2	2	0.044444	0.042553
...	
46	2	1	0.022222	0.100000
	3	0	0.000000	0.000000
47	1	5	0.111111	0.277778
	2	1	0.022222	0.055556
	3	0	0.000000	0.000000

[120 rows x 3 columns]

```
In [15]: pli.groupby('list_lag').agg(['mean', 'sem'])
Out[15]:
```

	count		per_list		prob	
	mean	sem	mean	sem	mean	sem
list_lag						
1	5.55	0.547664	0.123333	0.012170	0.210631	0.014726
2	1.35	0.230801	0.030000	0.005129	0.043458	0.007032
3	0.75	0.174496	0.016667	0.003878	0.023385	0.005602

The analysis returns a raw count of intrusions at each lag (`count`), the count divided by the number of included lists (`per_list`), and the probability of a given intrusion coming from a given lag (`prob`). In the sample dataset, recently presented items (i.e., with lower list lag) are more likely to be intruded.

2.4 Recall order

A key advantage of free recall is that it provides information not only about what items are recalled, but also the order in which they are recalled. A number of analyses have been developed to characterize different influences on recall order, such as the temporal order in which the items were presented at study, the category of the items themselves, or the semantic similarity between pairs of items.

Each conditional response probability (CRP) analysis involves calculating the probability of some type of transition event. For the lag-CRP analysis, transition events of interest are the different lags between serial positions of items recalled adjacent to one another. Similar analyses focus not on the serial position in which items are presented, but the properties of the items themselves. A semantic-CRP analysis calculates the probability of transitions between items in different semantic relatedness bins. A special case of this analysis is when item pairs are placed into one of two bins, depending on whether they are in the same stimulus category or not. In Psifr, this is referred to as a category-CRP analysis.

2.4.1 Lag-CRP

In all CRP analyses, transition probabilities are calculated conditional on a given transition being available [Kah96]. For example, in a six-item list, if the items 6, 1, and 4 have been recalled, then possible items that could have been recalled next are 2, 3, or 5; therefore, possible lags at that point in the recall sequence are -2, -1, or +1. The number of actual transitions observed for each lag is divided by the number of times that lag was possible, to obtain the CRP for each lag.

First, load some sample data and create a merged DataFrame:

```
In [1]: from psifr import fr
In [2]: df = fr.sample_data('Morton2013')
In [3]: data = fr.merge_free_recall(df, study_keys=['category'])
```

Next, call `lag_crp()` to calculate conditional response probability as a function of lag.

```
In [4]: crp = fr.lag_crp(data)
In [5]: crp
Out[5]:
```

		prob	actual	possible
subject	lag			
1	-23.0	0.020833	1	48
	-22.0	0.035714	3	84
	-21.0	0.026316	3	114
	-20.0	0.024000	3	125
	-19.0	0.014388	2	139
...
47	19.0	0.061224	3	49
	20.0	0.055556	2	36
	21.0	0.045455	1	22
	22.0	0.071429	1	14
	23.0	0.000000	0	6

[1880 rows x 3 columns]

The results show the count of times a given transition actually happened in the observed recall sequences (**actual**) and the number of times a transition could have occurred (**possible**). Finally, the **prob** column gives the estimated probability of a given transition occurring, calculated by dividing the actual count by the possible count.

Use `plot_lag_crp()` to display the results:

```
In [6]: g = fr.plot_lag_crp(crp)
```

The peaks at small lags (e.g., +1 and -1) indicate that the recall sequences show evidence of a temporal contiguity effect; that is, items presented near to one another in the list are more likely to be recalled successively than items that are distant from one another in the list.

2.4.2 Compound lag-CRP

The compound lag-CRP was developed to measure how temporal clustering changes as a result of prior clustering during recall [LK14]. They found evidence that temporal clustering is greater immediately after transitions with short lags compared to long lags. The `lag_crp_compound()` analysis calculates conditional response probability by lag, but with the additional condition of the lag of the previous transition.

```
In [7]: crp = fr.lag_crp_compound(data)
```

```
In [8]: crp
```

```
Out[8]:
```

			prob	actual	possible
subject	previous	current			
1	-23.0	-23.0	NaN	0	0
		-22.0	NaN	0	0
		-21.0	NaN	0	0
		-20.0	NaN	0	0
		-19.0	NaN	0	0
...		
47	23.0	19.0	NaN	0	0
		20.0	NaN	0	0
		21.0	NaN	0	0
		22.0	NaN	0	0
		23.0	NaN	0	0

```
[88360 rows x 3 columns]
```

The results show conditional response probabilities as in the standard lag-CRP analysis, but with two lag columns: **previous** (the lag of the prior transition) and **current** (the lag of the current transition).

This is a lot of information, and the sample size for many bins is very small. Following [LK14], we can apply bins to the lag of the previous transition to increase the sample size in each bin. We first sum the actual and possible transition counts, and then calculate the probability of each of the new bins.

```
In [9]: binned = crp.reset_index()
```

```
In [10]: binned.loc[binned['previous'].abs() > 3, 'Previous'] = '|Lag|>3'
```

```
In [11]: binned.loc[binned['previous'] == 1, 'Previous'] = 'Lag=+1'
```

(continues on next page)

(continued from previous page)

```

In [12]: binned.loc[binned['previous'] == -1, 'Previous'] = 'Lag=-1'

In [13]: summed = binned.groupby(['subject', 'Previous', 'current'])[['actual', 'possible'
↳ ]].sum()

In [14]: summed['prob'] = summed['actual'] / summed['possible']

In [15]: summed
Out[15]:

```

			actual	possible	prob
subject	Previous	current			
1	Lag=+1	-23.0	0	2	0.000000
		-22.0	0	2	0.000000
		-21.0	0	4	0.000000
		-20.0	0	6	0.000000
		-19.0	1	7	0.142857
...		
47	Lag >3	19.0	1	30	0.033333
		20.0	2	19	0.105263
		21.0	1	14	0.071429
		22.0	0	7	0.000000
		23.0	0	2	0.000000

[5640 rows x 3 columns]

We can then plot the compound lag-CRP using the standard `plot_lag_crp()` plotting function.

```

In [16]: g = fr.plot_lag_crp(summed, lag_key='current', hue='Previous').add_legend()

```

Note that some lags are considered impossible as they would require a repeat of a previously recalled item (e.g., a +1 lag followed by a -1 lag is not possible). For both of the adjacent conditions (+1 and -1), the lag-CRP is sharper compared to the long-lag condition ($|\text{lag}| > 3$). This suggests that there is compound temporal clustering.

2.4.3 Lag rank

We can summarize the tendency to group together nearby items by running a lag rank analysis [PNK09] using `lag_rank()`. For each recall, this determines the absolute lag of all remaining items available for recall and then calculates their percentile rank. Then the rank of the actual transition made is taken, scaled to vary between 0 (furthest item chosen) and 1 (nearest item chosen). Chance clustering will be 0.5; clustering above that value is evidence of a temporal contiguity effect.

```

In [17]: ranks = fr.lag_rank(data)

In [18]: ranks
Out[18]:

```

	rank
subject	
1	0.610953
2	0.635676
3	0.612607

(continues on next page)

(continued from previous page)

```

4      0.667090
5      0.643923
...
43     0.554024
44     0.561005
45     0.598151
46     0.652748
47     0.621245

```

```
[40 rows x 1 columns]
```

```
In [19]: ranks.agg(['mean', 'sem'])
```

```
Out[19]:
```

```

      rank
mean  0.624699
sem   0.006732

```

2.4.4 Category CRP

If there are multiple categories or conditions of trials in a list, we can test whether participants tend to successively recall items from the same category. The category-CRP, calculated using `category_crp()`, estimates the probability of successively recalling two items from the same category [PNK09].

```
In [20]: cat_crp = fr.category_crp(data, category_key='category')
```

```
In [21]: cat_crp
```

```
Out[21]:
```

```

      prob  actual  possible
subject
1      0.801147    419      523
2      0.733456    399      544
3      0.763158    377      494
4      0.814882    449      551
5      0.877273    579      660
...
43     0.809187    458      566
44     0.744376    364      489
45     0.763780    388      508
46     0.763573    436      571
47     0.806907    514      637

```

```
[40 rows x 3 columns]
```

```
In [22]: cat_crp[['prob']].agg(['mean', 'sem'])
```

```
Out[22]:
```

```

      prob
mean  0.782693
sem   0.006262

```

The expected probability due to chance depends on the number of categories in the list. In this case, there are three categories, so a category CRP of 0.33 would be predicted if recalls were sampled randomly from the list.

2.4.5 Category clustering

A number of measures have been developed to measure category clustering relative to that expected due to chance, under certain assumptions. Two such measures are list-based clustering (LBC) [SBW+02] and adjusted ratio of clustering (ARC) [RTB71].

These measures can be calculated using the `category_clustering()` function.

```
In [23]: clust = fr.category_clustering(data, category_key='category')
```

```
In [24]: clust.agg(['mean', 'sem'])
```

```
Out[24]:
```

	lbc	arc
mean	2.409398	0.608763
sem	0.127651	0.016809

Both measures are defined such that positive values indicate above-chance clustering. ARC scores have a maximum of 1, while the upper bound of LBC scores depends on the number of categories and the number of items per category in the study list.

2.4.6 Distance CRP

While the category CRP examines clustering based on semantic similarity at a coarse level (i.e., whether two items are in the same category or not), recall may also depend on more nuanced semantic relationships.

Models of semantic knowledge allow the semantic distance between pairs of items to be quantified. If you have such a model defined for your stimulus pool, you can use the distance CRP analysis to examine how semantic distance affects recall transitions [HK02, MP16].

You must first define distances between pairs of items. Here, we use correlation distances based on the wiki2USE model.

```
In [25]: items, distances = fr.sample_distances('Morton2013')
```

We also need a column indicating the index of each item in the distances matrix. We use `pool_index()` to create a new column called `item_index` with the index of each item in the pool corresponding to the distances matrix.

```
In [26]: data['item_index'] = fr.pool_index(data['item'], items)
```

Finally, we must define distance bins. Here, we use 10 bins with equally spaced distance percentiles. Note that, when calculating distance percentiles, we use the `squareform()` function to get only the non-diagonal entries.

```
In [27]: from scipy.spatial.distance import squareform
```

```
In [28]: edges = np.percentile(squareform(distances), np.linspace(1, 99, 10))
```

We can now calculate conditional response probability as a function of distance bin using `distance_crp()`, to examine how response probability varies with semantic distance.

```
In [29]: dist_crp = fr.distance_crp(data, 'item_index', distances, edges)
```

```
In [30]: dist_crp
```

```
Out[30]:
```

	bin	prob	actual	possible
subject center				

(continues on next page)

(continued from previous page)

1	0.467532	(0.352, 0.583]	0.085456	151	1767
	0.617748	(0.583, 0.653]	0.067916	87	1281
	0.673656	(0.653, 0.695]	0.062500	65	1040
	0.711075	(0.695, 0.727]	0.051836	48	926
	0.742069	(0.727, 0.757]	0.050633	44	869
...					
47	0.742069	(0.727, 0.757]	0.062822	61	971
	0.770867	(0.757, 0.785]	0.030682	27	880
	0.800404	(0.785, 0.816]	0.040749	37	908
	0.834473	(0.816, 0.853]	0.046651	39	836
	0.897275	(0.853, 0.941]	0.028868	25	866

[360 rows x 4 columns]

Use `plot_distance_crp()` to display the results:

```
In [31]: g = fr.plot_distance_crp(dist_crp).set(ylim=(0, 0.1))
```

Conditional response probability decreases with increasing semantic distance, suggesting that recall order was influenced by the semantic similarity between items. Of course, a complete analysis should address potential confounds such as the category structure of the list. See the [Restricting analysis to specific items](#) section for an example of restricting analysis based on category.

2.4.7 Distance rank

Similarly to the lag rank analysis of temporal clustering, we can summarize distance-based clustering (such as semantic clustering) with a single rank measure [PNK09]. The distance rank varies from 0 (the most-distant item is always recalled) to 1 (the closest item is always recalled), with chance clustering corresponding to 0.5. Given a matrix of item distances, we can calculate distance rank using `distance_rank()`.

```
In [32]: dist_rank = fr.distance_rank(data, 'item_index', distances)
```

```
In [33]: dist_rank.agg(['mean', 'sem'])
```

```
Out[33]:
      rank
mean  0.625932
sem   0.003466
```

2.4.8 Distance rank shifted

Like with the compound lag-CRP, we can also examine how recalls before the just-previous one may predict subsequent recalls. To examine whether distances relative to earlier items are predictive of the next recall, we can use a shifted distance rank analysis [MP16] using `distance_rank_shifted()`.

Here, to account for the category structure of the list, we will only include within-category transitions (see the [Restricting analysis to specific items](#) section for details).

```
In [34]: ranks = fr.distance_rank_shifted(
.....:     data, 'item_index', distances, 4, test_key='category', test=lambda x, y: x_
```

(continues on next page)

(continued from previous page)

```

→ == y
      ....: )
      ....:
In [35]: ranks
Out[35]:
           rank
subject shift
1         -4    0.518617
          -3    0.492103
          -2    0.516063
          -1    0.579198
2         -4    0.463931
...
46        -1    0.581420
47        -4    0.504383
          -3    0.526840
          -2    0.504953
          -1    0.586689

[160 rows x 1 columns]

```

The distance rank is returned for each shift. The -1 shift is the same as the standard distance rank analysis. We can visualize how distance rank changes with shift using `seaborn.relplot()`.

```

In [36]: g = sns.relplot(
      ....:     data=ranks.reset_index(), x='shift', y='rank', kind='line', height=3
      ....: ).set(xlabel='Output lag', ylabel='Distance rank', xticks=[-4, -3, -2, -1])
      ....:

```

2.4.9 Restricting analysis to specific items

Sometimes you may want to focus an analysis on a subset of recalls. For example, in order to exclude the period of high clustering commonly observed at the start of recall, lag-CRP analyses are sometimes restricted to transitions after the first three output positions.

You can restrict the recalls included in a transition analysis using the optional `item_query` argument. This is built on the Pandas query/eval system, which makes it possible to select rows of a `DataFrame` using a query string. This string can refer to any column in the data. Any items for which the expression evaluates to `True` will be included in the analysis.

For example, we can use the `item_query` argument to exclude any items recalled in the first three output positions from analysis. Note that, because non-recalled items have no output position, we need to include them explicitly using `output > 3` or `not recall`.

```

In [37]: crp_op3 = fr.lag_crp(data, item_query='output > 3 or not recall')

In [38]: g = fr.plot_lag_crp(crp_op3)

```

2.4.10 Restricting analysis to specific transitions

In other cases, you may want to focus an analysis on a subset of transitions based on some criteria. For example, if a list contains items from different categories, it is a good idea to take this into account when measuring temporal clustering using a lag-CRP analysis [MP17, PEK11]. One approach is to separately analyze within- and across-category transitions.

Transitions can be selected for inclusion using the optional `test_key` and `test` inputs. The `test_key` indicates a column of the data to use for testing transitions; for example, here we will use the `category` column. The `test` input should be a function that takes in the test value of the previous recall and the current recall and returns True or False to indicate whether that transition should be included. Here, we will use a lambda (anonymous) function to define the test.

```
In [39]: crp_within = fr.lag_crp(data, test_key='category', test=lambda x, y: x == y)
In [40]: crp_across = fr.lag_crp(data, test_key='category', test=lambda x, y: x != y)
In [41]: crp_combined = pd.concat([crp_within, crp_across], keys=['within', 'across'],
↳axis=0)
In [42]: crp_combined.index.set_names('transition', level=0, inplace=True)
In [43]: g = fr.plot_lag_crp(crp_combined, hue='transition').add_legend()
```

The `within` curve shows the lag-CRP for transitions between items of the same category, while the `across` curve shows transitions between items of different categories.

2.5 Comparing conditions

When analyzing a dataset, it's often important to compare different experimental conditions. Psifr is built on the Pandas `DataFrame`, which has powerful ways of splitting data and applying operations to it. This makes it possible to analyze and plot different conditions using very little code.

2.5.1 Working with custom columns

First, load some sample data and create a merged `DataFrame`:

```
In [1]: from psifr import fr
In [2]: df = fr.sample_data('Morton2013')
In [3]: data = fr.merge_free_recall(
...     df, study_keys=['category'], list_keys=['list_type']
... )
...
In [4]: data.head()
Out[4]:
```

	subject	list	item	input	...	list_type	category	prior_list	prior_input
0	1	1	TOWEL	1.0	...	pure	obj	NaN	NaN

(continues on next page)

(continued from previous page)

1	1	1	LADLE	2.0	...	pure	obj	NaN	NaN
2	1	1	THERMOS	3.0	...	pure	obj	NaN	NaN
3	1	1	LEGO	4.0	...	pure	obj	NaN	NaN
4	1	1	BACKPACK	5.0	...	pure	obj	NaN	NaN

[5 rows x 13 columns]

The `merge_free_recall()` function only includes columns from the raw data if they are one of the standard columns or if they've explicitly been included using `study_keys`, `recall_keys`, or `list_keys`. `list_keys` apply to all events in a list, while `study_keys` and `recall_keys` are relevant only for study and recall events, respectively.

We've included a list key here, to indicate that the `list_type` field should be included for all study and recall events in each list, even intrusions. The `category` field will be included for all study events and all valid recalls. Intrusions will have an undefined category.

2.5.2 Analysis by condition

Now we can run any analysis separately for the different conditions. We'll use the serial position curve analysis as an example.

```
In [5]: spc = data.groupby('list_type').apply(fr.spc)
```

```
In [6]: spc.head()
```

```
Out[6]:
```

```

              recall
list_type subject input
mixed      1      1.0  0.500000
           2      2.0  0.466667
           3      3.0  0.600000
           4      4.0  0.300000
           5      5.0  0.333333

```

The `spc` DataFrame has separate groups with the results for each `list_type`.

Warning: When using `groupby` with order-based analyses like `lag_crp()`, make sure all recalls in all recall sequences for a given list have the same label. Otherwise, you will be breaking up recall sequences, which could result in an invalid analysis.

2.5.3 Plotting by condition

We can then plot a separate curve for each condition. All plotting functions take optional `hue`, `col`, `col_wrap`, and `row` inputs that can be used to divide up data when plotting. Most inputs to `seaborn.relplot()` are supported.

For example, we can plot two curves for the different list types:

```
In [7]: g = fr.plot_spc(spc, hue='list_type').add_legend()
```

We can also plot the curves in different axes using the `col` option:

```
In [8]: g = fr.plot_spc(spc, col='list_type')
```

We can also plot all combinations of two conditions:

```
In [9]: spc_split = data.groupby(['list_type', 'category']).apply(fr.spc)
```

```
In [10]: g = fr.plot_spc(spc_split, col='list_type', row='category')
```

2.5.4 Plotting by subject

All analyses can be plotted separately by subject. A nice way to do this is using the `col` and `col_wrap` optional inputs, to make a grid of plots with 6 columns per row:

```
In [11]: g = fr.plot_spc(  
.....:     spc, hue='list_type', col='subject', col_wrap=6, height=2  
.....: ).add_legend()  
.....:
```


TUTORIALS

See the [psifr-notebooks](#) project for a set of Jupyter notebooks with sample code. These examples go more in depth into the options available for each analysis and how they can be used for advanced analyses such as conditionalizing CRP analysis on specific transitions.

API REFERENCE

4.1 Free recall analysis

4.1.1 Managing data

<code>table_from_lists(subjects, study, recall[, ...])</code>	Create table format data from list format data.
<code>check_data(df)</code>	Run checks on free recall data.
<code>merge_free_recall(data, **kwargs)</code>	Score free recall data by matching up study and recall events.
<code>merge_lists(study, recall[, merge_keys, ...])</code>	Merge study and recall events together for each list.
<code>filter_data(data[, subjects, lists, ...])</code>	Filter data to get a subset of trials.
<code>reset_list(df)</code>	Reset list index in a DataFrame.
<code>split_lists(frame, phase[, keys, names, ...])</code>	Convert free recall data from one phase to split format.
<code>pool_index(trial_items, pool_items_list)</code>	Get the index of each item in the full pool.
<code>block_index(list_labels)</code>	Get index of each block in a list.

psifr.fr.table_from_lists

`psifr.fr.table_from_lists(subjects, study, recall, lists=None, **kwargs)`

Create table format data from list format data.

Parameters

- **subjects** (*list of hashable*) – Subject identifier for each list.
- **study** (*list of list of hashable*) – List of items for each study list.
- **recall** (*list of list of hashable*) – List of recalled items for each study list.
- **lists** (*list of hashable, optional*) – List of list numbers. If not specified, lists for each subject will be numbered sequentially starting from one.

Returns

data – Data in table format.

Return type

`pandas.DataFrame`

See also:

`split_lists`

Split a table into list format.

Examples

```
>>> from psifr import fr
>>> subjects_list = [1, 1, 2, 2]
>>> study_lists = [['a', 'b'], ['c', 'd'], ['e', 'f'], ['g', 'h']]
>>> recall_lists = [['b'], ['d', 'c'], ['f', 'e'], []]
>>> fr.table_from_lists(subjects_list, study_lists, recall_lists)
  subject  list trial_type  position item
0         1         1      study         1  a
1         1         1      study         2  b
2         1         1      recall         1  b
3         1         2      study         1  c
4         1         2      study         2  d
5         1         2      recall         1  d
6         1         2      recall         2  c
7         2         1      study         1  e
8         2         1      study         2  f
9         2         1      recall         1  f
10        2         1      recall         2  e
11        2         2      study         1  g
12        2         2      study         2  h
```

```
>>> subjects_list = [1, 1]
>>> study_lists = [['a', 'b'], ['c', 'd']]
>>> recall_lists = [['b'], ['d', 'c']]
>>> col1 = ([[1, 2], [1, 2]], [[2], [2, 1]])
>>> col2 = ([[1, 1], [2, 2]], None)
>>> fr.table_from_lists(subjects_list, study_lists, recall_lists, col1=col1,
↳ col2=col2)
  subject  list trial_type  position item  col1  col2
0         1         1      study         1  a      1  1.0
1         1         1      study         2  b      2  1.0
2         1         1      recall         1  b      2  NaN
3         1         2      study         1  c      1  2.0
4         1         2      study         2  d      2  2.0
5         1         2      recall         1  d      2  NaN
6         1         2      recall         2  c      1  NaN
```

psifr.fr.check_data

psifr.fr.check_data(df)

Run checks on free recall data.

Parameters

df (*pandas.DataFrame*) –

Contains one row for each trial (study and recall). Must have fields:

subject

[number or str] Subject identifier.

list

[number] List identifier. This applies to both study and recall trials.

trial_type

[str] Type of trial; may be 'study' or 'recall'.

position

[number] Position within the study list or recall sequence.

item

[str] Item that was either presented or recalled on this trial.

Examples

```
>>> from psifr import fr
>>> import pandas as pd
>>> raw = pd.DataFrame(
...     {'subject': [1, 1], 'list': [1, 1], 'position': [1, 2], 'item': ['a', 'b']}
... )
>>> fr.check_data(raw)
Traceback (most recent call last):
  File "psifr/fr.py", line 253, in check_data
    assert col in df.columns, f'Required column {col} is missing.'
AssertionError: Required column trial_type is missing.
```

psifr.fr.merge_free_recall

`psifr.fr.merge_free_recall(data, **kwargs)`

Score free recall data by matching up study and recall events.

Parameters

- **data** (*pandas.DataFrame*) – Free recall data in Psifr format. Must have subject, list, trial_type, position, and item columns.
- **merge_keys** (*list, optional*) – Columns to use to designate events to merge. Default is ['subject', 'list', 'item'], which will merge events related to the same item, but only within list.
- **list_keys** (*list, optional*) – Columns that apply to both study and recall events.
- **study_keys** (*list, optional*) – Columns that only apply to study events.
- **recall_keys** (*list, optional*) – Columns that only apply to recall events.
- **position_key** (*str, optional*) – Column indicating the position of each item in either the study list or the recall sequence.

Returns

merged – Merged information about study and recall events. Each row corresponds to one unique input/output pair.

The following columns will be added:

input

[int] Position of each item in the input list (i.e., serial position).

output

[int] Position of each item in the recall sequence.

study

[bool] True for rows corresponding to a unique study event.

recall

[bool] True for rows corresponding to a unique recall event.

repeat

[int] Number of times this recall event has been repeated (0 for the first recall of an item).

intrusion

[bool] True for recalls that do not correspond to any study event.

prior_list

[int] For prior-list intrusions, the list the item was presented.

prior_position

[int] For prior-list intrusions, the position the item was presented.

Return type

`pandas.DataFrame`

See also:

`merge_lists`

Flexibly merge study events with recall events. Useful for recall phases that don't match the typical free recall setup, like final free recall of all lists.

Examples

```
>>> import numpy as np
>>> from psifr import fr
>>> study = [['absence', 'hollow'], ['fountain', 'piano']]
>>> recall = [['absence'], ['piano', 'hollow']]
>>> raw = fr.table_from_lists([1, 1], study, recall)
>>> raw
```

	subject	list	trial_type	position	item
0	1	1	study	1	absence
1	1	1	study	2	hollow
2	1	1	recall	1	absence
3	1	2	study	1	fountain
4	1	2	study	2	piano
5	1	2	recall	1	piano
6	1	2	recall	2	hollow

Score the data to create a table with matched study and recall events.

```
>>> data = fr.merge_free_recall(raw)
>>> data
```

	subject	list	item	input	output	study	recall	repeat	intrusion	prior_
↪ list		prior_input								
0	1	1	absence	1.0	1.0	True	True	0	False	↪
↪ NaN		NaN								
1	1	1	hollow	2.0	NaN	True	False	0	False	↪
↪ NaN		NaN								
2	1	2	fountain	1.0	NaN	True	False	0	False	↪
↪ NaN		NaN								

(continues on next page)

(continued from previous page)

3	1	2	piano	2.0	1.0	True	True	0	False	↵
↵ NaN		NaN								
4	1	2	hollow	NaN	2.0	False	True	0	True	↵
↵ 1.0		2.0								

You can also include non-standard columns. Information that only applies to study events (here, the encoding task used) can be indicated using the `study_keys` input.

```
>>> raw['task'] = np.array([1, 2, np.nan, 2, 1, np.nan, np.nan])
>>> fr.merge_free_recall(raw, study_keys=['task'])
```

	subject	list	item	input	output	study	recall	repeat	intrusion	task	↵
↵	prior_list	prior_input									
0	1	1	absence	1.0	1.0	True	True	0	False	1.0	↵
↵	NaN		NaN								
1	1	1	hollow	2.0	NaN	True	False	0	False	2.0	↵
↵	NaN		NaN								
2	1	2	fountain	1.0	NaN	True	False	0	False	2.0	↵
↵	NaN		NaN								
3	1	2	piano	2.0	1.0	True	True	0	False	1.0	↵
↵	NaN		NaN								
4	1	2	hollow	NaN	2.0	False	True	0	True	NaN	↵
↵	1.0		2.0								

Information that only applies to recall onsets (here, the time in seconds after the start of the recall phase that a recall attempt was made), can be indicated using the `recall_keys` input.

```
>>> raw['onset'] = np.array([np.nan, np.nan, 1.1, np.nan, np.nan, 1.4, 3.8])
>>> fr.merge_free_recall(raw, recall_keys=['onset'])
```

	subject	list	item	input	output	study	recall	repeat	intrusion	onset	↵
↵	prior_list	prior_input									
0	1	1	absence	1.0	1.0	True	True	0	False	1.1	↵
↵	NaN		NaN								
1	1	1	hollow	2.0	NaN	True	False	0	False	NaN	↵
↵	NaN		NaN								
2	1	2	fountain	1.0	NaN	True	False	0	False	NaN	↵
↵	NaN		NaN								
3	1	2	piano	2.0	1.0	True	True	0	False	1.4	↵
↵	NaN		NaN								
4	1	2	hollow	NaN	2.0	False	True	0	True	3.8	↵
↵	1.0		2.0								

Use `list_keys` to indicate columns that apply to both study and recall events. If `list_keys` do not match for a pair of study and recall events, they will not be matched in the output.

```
>>> raw['condition'] = np.array([1, 1, 1, 2, 2, 2, 2])
>>> fr.merge_free_recall(raw, list_keys=['condition'])
```

	subject	list	item	input	output	study	recall	repeat	intrusion	↵
↵	condition	prior_list	prior_input							
0	1	1	absence	1.0	1.0	True	True	0	False	↵
↵ 1		NaN	NaN							
1	1	1	hollow	2.0	NaN	True	False	0	False	↵
↵ 1		NaN	NaN							
2	1	2	fountain	1.0	NaN	True	False	0	False	↵

(continues on next page)

(continued from previous page)

↪ 2		NaN		NaN							
3	1	2	piano	2.0	1.0	True	True	0	False		↪
↪ 2		NaN		NaN							
4	1	2	hollow	NaN	2.0	False	True	0	True		↪
↪ 2		1.0		2.0							

psifr.fr.merge_lists

`psifr.fr.merge_lists(study, recall, merge_keys=None, list_keys=None, study_keys=None, recall_keys=None, position_key='position')`

Merge study and recall events together for each list.

Parameters

- **study** (*pandas.DataFrame*) – Information about all study events. Should have one row for each study event.
- **recall** (*pandas.DataFrame*) – Information about all recall events. Should have one row for each recall attempt.
- **merge_keys** (*list, optional*) – Columns to use to designate events to merge. Default is ['subject', 'list', 'item'], which will merge events related to the same item, but only within list.
- **list_keys** (*list, optional*) – Columns that apply to both study and recall events.
- **study_keys** (*list, optional*) – Columns that only apply to study events.
- **recall_keys** (*list, optional*) – Columns that only apply to recall events.
- **position_key** (*str, optional*) – Column indicating the position of each item in either the study list or the recall sequence.

Returns

merged – Merged information about study and recall events. Each row corresponds to one unique input/output pair.

The following columns will be added:

input

[int] Position of each item in the input list (i.e., serial position).

output

[int] Position of each item in the recall sequence.

study

[bool] True for rows corresponding to a unique study event.

recall

[bool] True for rows corresponding to a unique recall event.

repeat

[int] Number of times this recall event has been repeated (0 for the first recall of an item).

intrusion

[bool] True for recalls that do not correspond to any study event.

Return type

pandas.DataFrame

See also:

[`merge_free_recall`](#)

Score standard free recall data.

Examples

```
>>> import pandas as pd
>>> from psifr import fr
>>> study = pd.DataFrame(
...     {'subject': [1, 1], 'list': [1, 1], 'position': [1, 2], 'item': ['a', 'b']}
... )
>>> recall = pd.DataFrame(
...     {'subject': [1], 'list': [1], 'position': [1], 'item': ['b']}
... )
>>> fr.merge_lists(study, recall)
```

	subject	list	item	input	output	study	recall	repeat	intrusion
0	1	1	a	1	NaN	True	False	0	False
1	1	1	b	2	1.0	True	True	0	False

psifr.fr.filter_data

`psifr.fr.filter_data(data, subjects=None, lists=None, trial_type=None, positions=None, inputs=None, outputs=None)`

Filter data to get a subset of trials.

Parameters

- **data** (*pandas.DataFrame*) – Raw or merged data to filter.
- **subjects** (*hashable or list of hashable*) – Subject or subjects to include.
- **lists** (*hashable or list of hashable*) – List or lists to include.
- **trial_type** (*{'study', 'recall'}*) – Trial type to include.
- **positions** (*int or list of int*) – Position or positions to include.
- **inputs** (*int or list of int*) – Input position or positions to include.
- **outputs** (*int or list of int*) – Output position or positions to include.

Returns

filtered – The filtered subset of data.

Return type

pandas.DataFrame

Examples

```
>>> from psifr import fr
>>> subjects_list = [1, 1, 2, 2]
>>> study_lists = [['a', 'b'], ['c', 'd'], ['e', 'f'], ['g', 'h']]
>>> recall_lists = [['b'], ['d', 'c'], ['f', 'e'], []]
>>> raw = fr.table_from_lists(subjects_list, study_lists, recall_lists)
>>> fr.filter_data(raw, subjects=1, trial_type='study')
  subject  list trial_type  position item
0         1         1      study         1  a
1         1         1      study         2  b
3         1         2      study         1  c
4         1         2      study         2  d
```

```
>>> data = fr.merge_free_recall(raw)
>>> fr.filter_data(data, subjects=2)
  subject  list item  input  output  study  recall  repeat  intrusion  prior_list
prior_input
4         2         1    e         1      2.0    True    True         0      False      NaN
5         2         1    f         2      1.0    True    True         0      False      NaN
6         2         2    g         1      NaN    True   False         0      False      NaN
7         2         2    h         2      NaN    True   False         0      False      NaN
```

psifr.fr.reset_list

`psifr.fr.reset_list(df)`

Reset list index in a DataFrame.

Parameters

df (*pandas.DataFrame*) – Raw or merged data. Must have subject and list fields.

Returns

Data with a renumbered list field, starting from 1.

Return type

pandas.DataFrame

Examples

```
>>> from psifr import fr
>>> subjects_list = [1, 1]
>>> study_lists = [['a', 'b'], ['c', 'd']]
>>> recall_lists = [['b'], ['c', 'd']]
>>> list_nos = [3, 4]
>>> raw = fr.table_from_lists(subjects_list, study_lists, recall_lists, lists=list_
prior_input)
>>> raw
  subject  list trial_type  position item
```

(continues on next page)

(continued from previous page)

0	1	3	study	1	a
1	1	3	study	2	b
2	1	3	recall	1	b
3	1	4	study	1	c
4	1	4	study	2	d
5	1	4	recall	1	c
6	1	4	recall	2	d

```
>>> fr.reset_list(raw)
```

	subject	list	trial_type	position	item
0	1	1	study	1	a
1	1	1	study	2	b
2	1	1	recall	1	b
3	1	2	study	1	c
4	1	2	study	2	d
5	1	2	recall	1	c
6	1	2	recall	2	d

psifr.fr.split_lists

`psifr.fr.split_lists(frame, phase, keys=None, names=None, item_query=None, as_list=False)`

Convert free recall data from one phase to split format.

Parameters

- **frame** (*pandas.DataFrame*) – Free recall data with separate study and recall events.
- **phase** (*{'study', 'recall', 'raw'}*) – Phase of recall to split. If 'raw', all trials will be included.
- **keys** (*list of str, optional*) – Data columns to include in the split data. If not specified, all columns will be included.
- **names** (*list of str, optional*) – Name for each column in the returned split data. Default is to use the same names as the input columns.
- **item_query** (*str, optional*) – Query string to select study trials to include. See *pandas.DataFrame.query* for allowed format.
- **as_list** (*bool, optional*) – If true, each column will be output as a list; otherwise, outputs will be *numpy.ndarray*.

Returns

split – Data in split format. Each included column will be a key in the dictionary, with a list of either *numpy.ndarray* (default) or lists, containing the values for that column.

Return type

dict of str: list

See also:

[table_from_lists](#)

Convert list-format data to a table.

Examples

```
>>> from psifr import fr
>>> study = [['absence', 'hollow'], ['fountain', 'piano']]
>>> recall = [['absence'], ['piano', 'fountain']]
>>> raw = fr.table_from_lists([1, 1], study, recall)
>>> data = fr.merge_free_recall(raw)
>>> data
```

	subject	list	item	input	output	study	recall	repeat	intrusion	prior_
↪list		prior_input								
0	1	1	absence	1	1.0	True	True	0	False	↪
↪NaN		NaN								
1	1	1	hollow	2	NaN	True	False	0	False	↪
↪NaN		NaN								
2	1	2	fountain	1	2.0	True	True	0	False	↪
↪NaN		NaN								
3	1	2	piano	2	1.0	True	True	0	False	↪
↪NaN		NaN								

Get study events split by list, just including the list and item fields.

```
>>> fr.split_lists(data, 'study', keys=['list', 'item'], as_list=True)
{'list': [[1, 1], [2, 2]], 'item': [['absence', 'hollow'], ['fountain', 'piano']]}
```

Export recall events, split by list.

```
>>> fr.split_lists(data, 'recall', keys=['item'], as_list=True)
{'item': [['absence'], ['piano', 'fountain']]}
```

Raw events (i.e., events that haven't been scored) can also be exported to list format.

```
>>> fr.split_lists(raw, 'raw', keys=['position'])
{'position': [array([1, 2, 1]), array([1, 2, 1, 2])]}
```

psifr.fr.pool_index

psifr.fr.pool_index(*trial_items*, *pool_items_list*)

Get the index of each item in the full pool.

Parameters

- **trial_items** (*pandas.Series*) – The item presented on each trial.
- **pool_items_list** (*list* or *numpy.ndarray*) – List of items in the full pool.

Returns

item_index – Index of each item in the pool. Trials with items not in the pool will be <NA>.

Return type

pandas.Series

Examples

```
>>> import pandas as pd
>>> from psifr import fr
>>> trial_items = pd.Series(['b', 'a', 'z', 'c', 'd'])
>>> pool_items_list = ['a', 'b', 'c', 'd', 'e', 'f']
>>> fr.pool_index(trial_items, pool_items_list)
0      1
1      0
2    <NA>
3      2
4      3
dtype: Int64
```

psifr.fr.block_index

`psifr.fr.block_index(list_labels)`

Get index of each block in a list.

Parameters

list_labels (*list* or *numpy.ndarray*) – Position labels that define the blocks.

Returns

block – Block index of each position.

Return type

numpy.ndarray

Examples

```
>>> from psifr import fr
>>> list_labels = [2, 2, 3, 3, 3, 1, 1]
>>> fr.block_index(list_labels)
array([1, 1, 2, 2, 2, 3, 3])
```

4.1.2 Recall probability

<code>spc(df)</code>	Serial position curve.
<code>pnr(df[, item_query, test_key, test])</code>	Probability of recall by serial position and output position.

psifr.fr.spc

`psifr.fr.spc(df)`

Serial position curve.

Parameters

df (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`.

Returns

recall – Index includes:

subject

[hashable] Subject identifier.

input

[int] Serial position in the list.

Values are:

recall

[float] Recall probability for each serial position.

Return type

pandas.Series

See also:

[plot_spc](#)

Plot serial position curve results.

[pnr](#)

Probability of nth recall.

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> fr.spc(data)
      recall
subject input
1      1.0    0.541667
      2.0    0.458333
      3.0    0.625000
      4.0    0.333333
      5.0    0.437500
...
47     20.0   0.500000
      21.0   0.770833
      22.0   0.729167
      23.0   0.895833
      24.0   0.958333

[960 rows x 1 columns]
```

psifr.fr.pnr

`psifr.fr.pnr(df, item_query=None, test_key=None, test=None)`

Probability of recall by serial position and output position.

Calculate probability of Nth recall, where N is each output position. Invalid recalls (repeats and intrusions) are ignored and not counted toward output position.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists within each subject. Must have fields: `subject`, `list`, `input`, `output`, `study`, `recall`. Input position must be defined such that the first serial position is 1, not 0.
- **item_query** (*str, optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str, optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable, optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

prob – Analysis results. Has fields: `subject`, `output`, `input`, `prob`, `actual`, `possible`. The `prob` column for output `x` and input `y` indicates the probability of recalling input position `y` at output position `x`. The `actual` and `possible` columns give the raw tallies for how many times an event actually occurred and how many times it was possible given the recall sequence.

Return type

pandas.DataFrame

See also:

plot_spc

Plot recall probability as a function of serial position.

spc

Overall recall probability by serial position.

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> fr.pnr(data)
```

subject	output	input	prob	actual	possible
1	1	1	0.000000	0	48
		2	0.020833	1	48
		3	0.000000	0	48
		4	0.000000	0	48
		5	0.000000	0	48
...	
47	24	20	NaN	0	0
		21	NaN	0	0

(continues on next page)

(continued from previous page)

22	NaN	0	0
23	NaN	0	0
24	NaN	0	0

[23040 rows x 3 columns]

4.1.3 Intrusions

`pli_list_lag(df, max_lag)`

List lag of prior-list intrusions.

psifr.fr.pli_list_lag

`psifr.fr.pli_list_lag(df, max_lag)`

List lag of prior-list intrusions.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_free_recall`. Must have fields: `subject`, `list`, `intrusion`, `prior_list`. Lists must be numbered starting from 1 and all lists must be included.
- **max_lag** (*int*) – Maximum list lag to consider. The initial `max_lag` lists for each subject will be excluded so that all considered lags are possible for all included lists.

Returns

results – For each subject and list lag, the proportion of intrusions at that lag, in the `results['prob']` column.

Return type

pandas.DataFrame

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> fr.pli_list_lag(data, 3)
```

		count	per_list	prob
subject	list_lag			
1	1	7	0.155556	0.259259
	2	5	0.111111	0.185185
	3	0	0.000000	0.000000
2	1	9	0.200000	0.191489
	2	2	0.044444	0.042553
...
46	2	1	0.022222	0.100000
	3	0	0.000000	0.000000
47	1	5	0.111111	0.277778
	2	1	0.022222	0.055556
	3	0	0.000000	0.000000

(continues on next page)

(continued from previous page)

[120 rows x 3 columns]

4.1.4 Transition probability

<code>lag_crp(df[, lag_key, count_unique, ...])</code>	Lag-CRP for multiple subjects.
<code>lag_crp_compound(df[, lag_key, ...])</code>	Conditional response probability by lag of current and prior transitions.
<code>category_crp(df, category_key[, item_query, ...])</code>	Conditional response probability of within-category transitions.
<code>distance_crp(df, index_key, distances, edges)</code>	Conditional response probability by distance bin.

psifr.fr.lag_crp

`psifr.fr.lag_crp(df, lag_key='input', count_unique=False, item_query=None, test_key=None, test=None)`

Lag-CRP for multiple subjects.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists. Must have fields: `subject`, `list`, `input`, `output`, `recalled`. Input position must be defined such that the first serial position is 1, not 0.
- **lag_key** (*str*, *optional*) – Name of column to use when calculating lag between recalled items. Default is to calculate lag based on input position.
- **count_unique** (*bool*, *optional*) – If true, possible transitions of the same lag will only be incremented once per transition.
- **item_query** (*str*, *optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str*, *optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable*, *optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

results – Has fields:

subject

[hashable] Results are separated by each subject.

lag

[int] Lag of input position between two adjacent recalls.

prob

[float] Probability of each lag transition.

actual

[int] Total of actual made transitions at each lag.

possible

[int] Total of times each lag was possible, given the prior input position and the remaining items to be recalled.

Return type

`pandas.DataFrame`

See also:

`lag_rank`

Rank of the absolute lags in recall sequences.

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> fr.lag_crp(data)
```

		prob	actual	possible
subject	lag			
1	-23.0	0.020833	1	48
	-22.0	0.035714	3	84
	-21.0	0.026316	3	114
	-20.0	0.024000	3	125
	-19.0	0.014388	2	139
...
47	19.0	0.061224	3	49
	20.0	0.055556	2	36
	21.0	0.045455	1	22
	22.0	0.071429	1	14
	23.0	0.000000	0	6

[1880 rows x 3 columns]

`psifr.fr.lag_crp_compound`

`psifr.fr.lag_crp_compound(df, lag_key='input', count_unique=False, item_query=None, test_key=None, test=None)`

Conditional response probability by lag of current and prior transitions.

Parameters

- **df** (`pandas.DataFrame`) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists. Must have fields: `subject`, `list`, `input`, `output`, `recalled`. Input position must be defined such that the first serial position is 1, not 0.
- **lag_key** (`str`, *optional*) – Name of column to use when calculating lag between recalled items. Default is to calculate lag based on input position.
- **count_unique** (`bool`, *optional*) – If true, possible transitions of the same lag will only be incremented once per transition.
- **item_query** (`str`, *optional*) – Query string to select items to include in the pool of possible recalls to be examined. See `pandas.DataFrame.query` for allowed format.

- **test_key** (*str*, *optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable*, *optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

results – Has fields:

subject

[hashable] Results are separated by each subject.

previous

[int] Lag of the previous transition.

current

[int] Lag of the current transition.

prob

[float] Probability of each lag transition.

actual

[int] Total of actual made transitions at each lag.

possible

[int] Total of times each lag was possible, given the prior input position and the remaining items to be recalled.

Return type

`pandas.DataFrame`

See also:

`lag_crp`

Conditional response probability by lag.

Examples

```
>>> from psifr import fr
>>> subjects = [1]
>>> study = [['absence', 'hollow', 'pupil', 'fountain']]
>>> recall = [['fountain', 'hollow', 'absence']]
>>> raw = fr.table_from_lists(subjects, study, recall)
>>> data = fr.merge_free_recall(raw)
>>> crp = fr.lag_crp_compound(data)
>>> crp.head(14)
```

			prob	actual	possible
subject	previous	current			
1	-3	-3	NaN	0	0
		-2	NaN	0	0
		-1	NaN	0	0
		0	NaN	0	0
		1	NaN	0	0
		2	NaN	0	0
		3	NaN	0	0
	-2	-3	NaN	0	0
		-2	NaN	0	0

(continues on next page)

(continued from previous page)

-1	1.0	1	1
0	NaN	0	0
1	0.0	0	1
2	NaN	0	0
3	NaN	0	0

psifr.fr.category_crp

`psifr.fr.category_crp(df, category_key, item_query=None, test_key=None, test=None)`

Conditional response probability of within-category transitions.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists within each subject. Must have fields: subject, list, input, output, recalled.
- **category_key** (*str*) – Name of column with category labels.
- **item_query** (*str, optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str, optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable, optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

results – Has fields:

subject

[hashable] Results are separated by each subject.

prob

[float] Probability of each lag transition.

actual

[int] Total of actual made transitions at each lag.

possible

[int] Total of times each lag was possible, given the prior input position and the remaining items to be recalled.

Return type

pandas.DataFrame

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw, study_keys=['category'])
>>> cat_crp = fr.category_crp(data, 'category')
>>> cat_crp.head()
```

	prob	actual	possible
subject			
1	0.801147	419	523
2	0.733456	399	544
3	0.763158	377	494
4	0.814882	449	551
5	0.877273	579	660

psifr.fr.distance_crp

`psifr.fr.distance_crp(df, index_key, distances, edges, centers=None, count_unique=False, item_query=None, test_key=None, test=None)`

Conditional response probability by distance bin.

Parameters

- **df** (*pandas.DataFrame*) – Merged free recall data.
- **index_key** (*str*) – Name of column containing the index of each item in the *distances* matrix.
- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **edges** (*array-like*) – Edges of bins to apply to the distances.
- **centers** (*array-like, optional*) – Centers to label each bin with. If not specified, the center point between edges will be used.
- **count_unique** (*bool, optional*) – If true, possible transitions to a given distance bin will only count once for a given transition.
- **item_query** (*str, optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str, optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable, optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

crp – Has fields:

subject

[hashable] Results are separated by each subject.

bin

[int] Distance bin.

prob

[float] Probability of each distance bin.

actual

[int] Total of actual transitions for each distance bin.

possible

[int] Total of times each distance bin was possible, given the prior input position and the remaining items to be recalled.

Return type

`pandas.DataFrame`

See also:

`pool_index`

Given a list of presented items and an item pool, look up the pool index of each item.

`distance_rank`

Calculate rank of transition distances.

Examples

```
>>> import numpy as np
>>> from scipy.spatial.distance import squareform
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> items, distances = fr.sample_distances('Morton2013')
>>> data['item_index'] = fr.pool_index(data['item'], items)
>>> edges = np.percentile(squareform(distances), np.linspace(1, 99, 10))
>>> fr.distance_crp(data, 'item_index', distances, edges)
```

		bin	prob	actual	possible
subject	center				
1	0.467532	(0.352, 0.583]	0.085456	151	1767
	0.617748	(0.583, 0.653]	0.067916	87	1281
	0.673656	(0.653, 0.695]	0.062500	65	1040
	0.711075	(0.695, 0.727]	0.051836	48	926
	0.742069	(0.727, 0.757]	0.050633	44	869
...	
47	0.742069	(0.727, 0.757]	0.062822	61	971
	0.770867	(0.757, 0.785]	0.030682	27	880
	0.800404	(0.785, 0.816]	0.040749	37	908
	0.834473	(0.816, 0.853]	0.046651	39	836
	0.897275	(0.853, 0.941]	0.028868	25	866

[360 rows x 4 columns]

4.1.5 Transition rank

<code>lag_rank(df[, item_query, test_key, test])</code>	Calculate rank of the absolute lags in free recall lists.
<code>distance_rank(df, index_key, distances[, ...])</code>	Calculate rank of transition distances in free recall lists.
<code>distance_rank_shifted(df, index_key, ...[, ...])</code>	Rank of transition distances relative to earlier items.

psifr.fr.lag_rank

`psifr.fr.lag_rank(df, item_query=None, test_key=None, test=None)`

Calculate rank of the absolute lags in free recall lists.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists within each subject. Must have fields: `subject`, `list`, `input`, `output`, `recalled`. Input position must be defined such that the first serial position is 1, not 0.
- **item_query** (*str*, *optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str*, *optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable*, *optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

stat – Has fields ‘subject’ and ‘rank’.

Return type

pandas.DataFrame

See also:

lag_crp

Conditional response probability by input lag.

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> lag_rank = fr.lag_rank(data)
>>> lag_rank.head()
      rank
subject
1      0.610953
2      0.635676
3      0.612607
4      0.667090
5      0.643923
```

psifr.fr.distance_rank

`psifr.fr.distance_rank(df, index_key, distances, item_query=None, test_key=None, test=None)`

Calculate rank of transition distances in free recall lists.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists within each subject. Must have fields: `subject`, `list`, `input`, `output`, `recalled`. Input position must be defined such that the first serial position is 1, not 0.
- **index_key** (*str*) – Name of column containing the index of each item in the *distances* matrix.
- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **item_query** (*str, optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str, optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable, optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

stat – Has fields ‘subject’ and ‘rank’.

Return type

pandas.DataFrame

See also:

[pool_index](#)

Given a list of presented items and an item pool, look up the pool index of each item.

[distance_crp](#)

Conditional response probability by distance bin.

Examples

```
>>> from scipy.spatial.distance import squareform
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> items, distances = fr.sample_distances('Morton2013')
>>> data['item_index'] = fr.pool_index(data['item'], items)
>>> dist_rank = fr.distance_rank(data, 'item_index', distances)
>>> dist_rank.head()
```

	rank
subject	
1	0.635571
2	0.571457
3	0.627282
4	0.637596
5	0.646181

psifr.fr.distance_rank_shifted

`psifr.fr.distance_rank_shifted(df, index_key, distances, max_shift, item_query=None, test_key=None, test=None)`

Rank of transition distances relative to earlier items.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_lists`. List length is assumed to be the same for all lists within each subject. Must have fields: `subject`, `list`, `input`, `output`, `recalled`. Input position must be defined such that the first serial position is 1, not 0.
- **index_key** (*str*) – Name of column containing the index of each item in the *distances* matrix.
- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **max_shift** (*int*) – Maximum number of items back for which to rank distances.
- **item_query** (*str, optional*) – Query string to select items to include in the pool of possible recalls to be examined. See *pandas.DataFrame.query* for allowed format.
- **test_key** (*str, optional*) – Name of column with labels to use when testing transitions for inclusion.
- **test** (*callable, optional*) – Callable that takes in previous and current item values and returns True for transitions that should be included.

Returns

stat – Has fields ‘subject’ and ‘rank’.

Return type

pandas.DataFrame

See also:

pool_index

Given a list of presented items and an item pool, look up the pool index of each item.

distance_rank

Rank of transition distances relative to the just-previous item.

Examples

```
>>> from scipy.spatial.distance import squareform
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> data = fr.merge_free_recall(raw)
>>> items, distances = fr.sample_distances('Morton2013')
>>> data['item_index'] = fr.pool_index(data['item'], items)
>>> dist_rank = fr.distance_rank_shifted(data, 'item_index', distances, 3)
>>> dist_rank
```

		rank
subject	shift	
1	-3	0.523426
	-2	0.559199
	-1	0.634392
2	-3	0.475931

(continues on next page)

(continued from previous page)

	-2	0.507574
...		...
46	-2	0.515332
	-1	0.603304
47	-3	0.542951
	-2	0.565001
	-1	0.635415
[120 rows x 1 columns]		

4.1.6 Clustering

<code>category_clustering(df, category_key)</code>	Category clustering of recall sequences.
----------------------------------------------------	------------------------------------------

psifr.fr.category_clustering

`psifr.fr.category_clustering(df, category_key)`

Category clustering of recall sequences.

Calculates ARC (adjusted ratio of clustering) and LBC (list-based clustering) statistics indexing recall clustering by category.

The papers introducing these measures do not describe how to handle repeats and intrusions. Here, to maintain the assumptions of the measures, they are removed from the recall sequences.

Note that ARC is undefined when only one category is recalled. Lists with undefined statistics will be excluded from calculation of mean subject-level statistics. To calculate for each list separately, group by list before calling the function. For example: `df.groupby('list').apply(fr.category_clustering, 'category')`.

Parameters

- **df** (*pandas.DataFrame*) – Merged study and recall data. See `merge_free_recall`. Must have a field indicating the category of each study and recall event.
- **category_key** (*str*) – Column with category labels. Labels may be any hashable (e.g., a str or int).

Returns

stats – For each subject, includes columns with the mean ARC and LBC statistics.

Return type

pandas.DataFrame

Examples

```
>>> from psifr import fr
>>> raw = fr.sample_data('Morton2013')
>>> mixed = raw.query('list_type == "mixed"')
>>> data = fr.merge_free_recall(mixed, list_keys=['category'])
>>> stats = fr.category_clustering(data, 'category')
>>> stats.head()
      lbc      arc
subject
1    3.657971  0.614545
2    2.953623  0.407839
3    3.363768  0.627371
4    4.444928  0.688761
5    7.530435  0.873755
```

4.1.7 Plotting

<code>plot_raster(df[, hue, palette, marker, ...])</code>	Plot recalls in a raster plot.
<code>plot_spc(recall, **facet_kws)</code>	Plot a serial position curve.
<code>plot_lag_crp(recall[, max_lag, lag_key, split])</code>	Plot conditional response probability by lag.
<code>plot_distance_crp(crp[, min_samples])</code>	Plot response probability by distance bin.
<code>plot_swarm_error(data[, x, y, swarm_color, ...])</code>	Plot points as a swarm plus mean with error bars.

psifr.fr.plot_raster

`psifr.fr.plot_raster(df, hue='input', palette=None, marker='s', intrusion_color=None, orientation='horizontal', length=6, aspect=None, legend='auto', **facet_kws)`

Plot recalls in a raster plot.

Parameters

- **df** (*pandas.DataFrame*) – Scored free recall data.
- **hue** (*str or None, optional*) – Column to use to set marker color.
- **palette** (*optional*) – Palette specification supported by Seaborn.
- **marker** (*str, optional*) – Marker code supported by Seaborn.
- **intrusion_color** (*optional*) – Color of intrusions.
- **orientation** (*{'horizontal', 'vertical'}, optional*) – Whether lists should be stacked horizontally or vertically in the plot.
- **length** (*float, optional*) – Size of the plot dimension along which list varies.
- **aspect** (*float, optional*) – Aspect ratio of plot for lists over items.
- **legend** (*str, optional*) – Legend setting. See `seaborn.scatterplot` for details.
- **facet_kws** (*optional*) – Additional key words to pass to `seaborn.FacetGrid`.

psifr.fr.plot_spc

`psifr.fr.plot_spc(recall, **facet_kws)`

Plot a serial position curve.

Additional arguments are passed to `seaborn.relplot`.

Parameters

recall (*pandas.DataFrame*) – Results from calling *spc*.

psifr.fr.plot_lag_crp

`psifr.fr.plot_lag_crp(recall, max_lag=5, lag_key='lag', split=True, **facet_kws)`

Plot conditional response probability by lag.

Additional arguments are passed to `seaborn.FacetGrid`.

Parameters

- **recall** (*pandas.DataFrame*) – Results from calling *lag_crp*.
- **max_lag** (*int*, *optional*) – Maximum absolute lag to plot.
- **lag_key** (*str*, *optional*) – Name of the column indicating lag.
- **split** (*bool*, *optional*) – If true, will plot as two separate lines with a gap at lag 0.

psifr.fr.plot_distance_crp

`psifr.fr.plot_distance_crp(crp, min_samples=None, **facet_kws)`

Plot response probability by distance bin.

Parameters

- **crp** (*pandas.DataFrame*) – Results from *fr.distance_crp*.
- **min_samples** (*int*) – Minimum number of samples a bin must have per subject to include in the plot.
- ****facet_kws** – Additional inputs to pass to *seaborn.relplot*.

psifr.fr.plot_swarm_error

`psifr.fr.plot_swarm_error(data, x=None, y=None, swarm_color=None, swarm_size=5, point_color='k', **facet_kws)`

Plot points as a swarm plus mean with error bars.

Parameters

- **data** (*pandas.DataFrame*) – DataFrame with statistics to plot.
- **x** (*str*) – Name of variable to plot on x-axis.
- **y** (*str*) – Name of variable to plot on y-axis.
- **swarm_color** – Color for swarm plot points. May use any specification supported by *seaborn*.
- **swarm_size** (*float*) – Size of swarm plot points.

- **point_color** – Color for the point plot (error bars).
- **facet_kws** – Additional keywords for the FacetGrid.

4.2 Measures

4.2.1 Transition measure base class

<code>TransitionMeasure(items_key, label_key[, ...])</code>	Measure of free recall dataset with multiple subjects.
<code>TransitionMeasure.split_lists(data, phase[, ...])</code>	Get relevant fields and split by list.
<code>TransitionMeasure.analyze(data)</code>	Analyze a free recall dataset with multiple subjects.
<code>TransitionMeasure.analyze_subject(subject, ...)</code>	Analyze a single subject.

psifr.measures.TransitionMeasure

class psifr.measures.TransitionMeasure(*items_key*, *label_key*, *item_query=None*, *test_key=None*, *test=None*)

Measure of free recall dataset with multiple subjects.

Parameters

- **items_key** (*str*) – Data column with item identifiers.
- **label_key** (*str*) – Data column with trial labels to use for the measure.
- **item_query** (*str*) – Query string to indicate trials to include in the measure.
- **test_key** (*str*) – Data column with labels to use when testing for trial inclusion.
- **test** (*callable*) – Test of trial inclusion. Takes the previous and current test values and return True if the transition should be included.

keys

List of columns to use for the measure.

Type

dict of {str: str}

item_query

Query string to indicate trials to include in the measure.

Type

str

test

Test of trial inclusion.

Type

callable

__init__(*items_key*, *label_key*, *item_query=None*, *test_key=None*, *test=None*)

Methods

<code>__init__(items_key, label_key[, item_query, ...])</code>	
<code>analyze(data)</code>	Analyze a free recall dataset with multiple subjects.
<code>analyze_subject(subject, pool_lists, ...)</code>	Analyze a single subject.
<code>split_lists(data, phase[, item_query])</code>	Get relevant fields and split by list.

psifr.measures.TransitionMeasure.split_lists

`TransitionMeasure.split_lists(data, phase, item_query=None)`

Get relevant fields and split by list.

Parameters

- **data** (*pandas.DataFrame*) – Raw free recall data.
- **phase** (*str*) – Phase to split ('study' or 'recall').
- **item_query** (*str, optional*) – Query string to determine included trials.

psifr.measures.TransitionMeasure.analyze

`TransitionMeasure.analyze(data)`

Analyze a free recall dataset with multiple subjects.

Parameters

data (*pandas.DataFrame*) – Raw (not merged) free recall data.

Returns

stat – Statistics calculated for each subject.

Return type

pandas.DataFrame

psifr.measures.TransitionMeasure.analyze_subject

abstract `TransitionMeasure.analyze_subject(subject, pool_lists, recall_lists)`

Analyze a single subject.

Parameters

- **subject** (*int or str*) – Identifier of the subject to analyze.
- **pool_lists** (*dict of lists of numpy.ndarray*) – Information about the item pool for each list, with keys for items, label, and test arrays.
- **recall_lists** (*dict of lists of numpy.ndarray*) – Information about the recall sequence for each list, with keys for items, label, and test arrays.

Returns

Results of the analysis for one subject. Should include a 'subject' column in the index.

Return type

pandas.DataFrame

4.2.2 Transition measures

<code>TransitionOutputs(list_length[, item_query, ...])</code>	Measure recall probability by input and output position.
<code>TransitionLag(list_length[, lag_key, ...])</code>	Measure conditional response probability by lag.
<code>TransitionLagRank([item_query, test_key, test])</code>	Measure lag rank of transitions.
<code>TransitionCategory(category_key[, ...])</code>	Measure conditional response probability by category transition.
<code>TransitionDistance(index_key, distances, edges)</code>	Measure conditional response probability by distance.
<code>TransitionDistanceRank(index_key, distances)</code>	Measure transition rank by distance.

4.3 Transitions

4.3.1 Counting transitions

<code>count_lags(list_length, pool_items, recall_items)</code>	Count actual and possible serial position lags.
<code>count_lags_compound(list_length, pool_items, ...)</code>	Count lags conditional on the lag of the previous transition.
<code>count_category(pool_items, recall_items, ...)</code>	Count within-category transitions.
<code>count_distance(distances, edges, pool_items, ...)</code>	Count transitions within distance bins.

psifr.transitions.count_lags

`psifr.transitions.count_lags(list_length, pool_items, recall_items, pool_label=None, recall_label=None, pool_test=None, recall_test=None, test=None, count_unique=False)`

Count actual and possible serial position lags.

Parameters

- **list_length** (*int*) – Number of items in each list.
- **pool_items** (*list*) – List of the serial positions available for recall in each list. Must match the serial position codes used in *recall_items*.
- **recall_items** (*list*) – List indicating the serial position of each recall in output order (NaN for intrusions).
- **pool_label** (*list*, *optional*) – List of the positions to use for calculating lag. Default is to use *pool_items*.
- **recall_label** (*list*, *optional*) – List of position labels in recall order. Default is to use *recall_items*.
- **pool_test** (*list*, *optional*) – List of some test value for each item in the pool.
- **recall_test** (*list*, *optional*) – List of some test value for each recall attempt by output position.
- **test** (*callable*) – Callable that evaluates each transition between items *n* and *n+1*. Must take test values for items *n* and *n+1* and return True if a given transition should be included.
- **count_unique** (*bool*, *optional*) – If true, only unique values will be counted toward the possible transitions. If multiple items are available for recall for a given transition and a given bin, that bin will only be incremented once. If false, all possible transitions will add to the count.

Returns

- **actual** (*pandas.Series*) – Count of actual lags that occurred in the recall sequence.
- **possible** (*pandas.Series*) – Count of possible lags.

See also:

rank_lags

Rank of serial position lags.

Examples

```
>>> from psifr import transitions
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 2, 3, 1]]
>>> actual, possible = transitions.count_lags(4, pool_items, recall_items)
>>> actual
lag
-3    0
-2    2
-1    0
 0    0
 1    1
 2    0
 3    0
dtype: int64
>>> possible
lag
-3    1
-2    2
-1    2
 0    0
 1    1
 2    0
 3    0
dtype: int64
```

psifr.transitions.count_lags_compound

```
psifr.transitions.count_lags_compound(list_length, pool_items, recall_items, pool_label=None,
                                     recall_label=None, pool_test=None, recall_test=None, test=None,
                                     count_unique=False)
```

Count lags conditional on the lag of the previous transition.

Parameters

- **list_length** (*int*) – Number of items in each list.
- **pool_items** (*list*) – List of the serial positions available for recall in each list. Must match the serial position codes used in *recall_items*.
- **recall_items** (*list*) – List indicating the serial position of each recall in output order (NaN for intrusions).

- **pool_label** (*list*, *optional*) – List of the positions to use for calculating lag. Default is to use *pool_items*.
- **recall_label** (*list*, *optional*) – List of position labels in recall order. Default is to use *recall_items*.
- **pool_test** (*list*, *optional*) – List of some test value for each item in the pool.
- **recall_test** (*list*, *optional*) – List of some test value for each recall attempt by output position.
- **test** (*callable*) – Callable that evaluates each transition between items *n* and *n+1*. Must take test values for items *n* and *n+1* and return True if a given transition should be included.
- **count_unique** (*bool*, *optional*) – If true, only unique values will be counted toward the possible transitions. If multiple items are available for recall for a given transition and a given bin, that bin will only be incremented once. If false, all possible transitions will add to the count.

Returns

- **actual** (*pandas.Series*) – Count of actual lags that occurred in the recall sequence.
- **possible** (*pandas.Series*) – Count of possible lags.

See also:

count_lags

Count of individual transitions.

Examples

```
>>> from psifr import transitions
>>> pool_items = [[1, 2, 3]]
>>> recall_items = [[3, 1, 2]]
>>> actual, possible = transitions.count_lags_compound(3, pool_items, recall_items)
>>> (actual == possible).all()
True
>>> actual
```

previous	current	
-2	-2	0
	-1	0
	0	0
	1	1
	2	0
-1	-2	0
	-1	0
	0	0
	1	0
	2	0
0	-2	0
	-1	0
	0	0
	1	0
	2	0
1	-2	0
	-1	0

(continues on next page)

(continued from previous page)

	0	0
	1	0
	2	0
2	-2	0
	-1	0
	0	0
	1	0
	2	0
dtype: int64		

psifr.transitions.count_category

`psifr.transitions.count_category(pool_items, recall_items, pool_category, recall_category, pool_test=None, recall_test=None, test=None)`

Count within-category transitions.

Parameters

- **pool_items** (*list*) – List of the serial positions available for recall in each list. Must match the serial position codes used in *recall_items*.
- **recall_items** (*list*) – List indicating the serial position of each recall in output order (NaN for intrusions).
- **pool_category** (*list*) – List of the category of each item in the pool for each list.
- **recall_category** (*list*) – List of item category in recall order.
- **pool_test** (*list, optional*) – List of some test value for each item in the pool.
- **recall_test** (*list, optional*) – List of some test value for each recall attempt by output position.
- **test** (*callable*) – Callable that evaluates each transition between items *n* and *n+1*. Must take test values for items *n* and *n+1* and return True if a given transition should be included.

Returns

- **actual** (*int*) – Count of actual within-category transitions.
- **possible** (*int*) – Count of possible within-category transitions.

Examples

```
>>> from psifr import transitions
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 3, 1, 2]]
>>> pool_category = [[1, 1, 2, 2]]
>>> recall_category = [[2, 2, 1, 1]]
>>> transitions.count_category(
...     pool_items, recall_items, pool_category, recall_category
... )
(2, 2)
```

psifr.transitions.count_distance

`psifr.transitions.count_distance(distances, edges, pool_items, recall_items, pool_index, recall_index, pool_test=None, recall_test=None, test=None, count_unique=False)`

Count transitions within distance bins.

Parameters

- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **edges** (*array-like*) – Edges of bins to apply to distances.
- **pool_items** (*list of list*) – Unique item codes for each item in the pool available for recall.
- **recall_items** (*list of list*) – Unique item codes of recalled items.
- **pool_index** (*list of list*) – Index of each item in the distances matrix.
- **recall_index** (*list of list*) – Index of each recalled item.
- **pool_test** (*list of list, optional*) – Test value for each item in the pool.
- **recall_test** (*list of list, optional*) – Test value for each recalled item.
- **test** (*callable*) – Called as `test(prev, curr)` or `test(prev, poss)` to screen actual and possible transitions, respectively.
- **count_unique** (*bool, optional*) – If true, only unique values will be counted toward the possible transitions. If multiple items are available for recall for a given transition and a given bin, that bin will only be incremented once. If false, all possible transitions will add to the count.

Returns

- **actual** (*pandas.Series*) – Count of actual transitions made for each bin.
- **possible** (*pandas.Series*) – Count of possible transitions for each bin.

See also:

[rank_distance](#)

Calculate percentile rank of transition distances.

Examples

```
>>> import numpy as np
>>> from psifr import transitions
>>> distances = np.array([[0, 1, 2, 2], [1, 0, 2, 2], [2, 2, 0, 3], [2, 2, 3, 0]])
>>> edges = np.array([0.5, 1.5, 2.5, 3.5])
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 2, 3, 1]]
>>> pool_index = [[0, 1, 2, 3]]
>>> recall_index = [[3, 1, 2, 0]]
>>> actual, possible = transitions.count_distance(
...     distances, edges, pool_items, recall_items, pool_index, recall_index
... )
>>> actual
(0.5, 1.5]    0
```

(continues on next page)

(continued from previous page)

```
(1.5, 2.5]    3
(2.5, 3.5]    0
dtype: int64
>>> possible
(0.5, 1.5]    1
(1.5, 2.5]    4
(2.5, 3.5]    1
dtype: int64
```

4.3.2 Ranking transitions

<code>percentile_rank(actual, possible)</code>	Get percentile rank of a score compared to possible scores.
<code>rank_lags(pool_items, recall_items[, ...])</code>	Calculate rank of absolute lag for free recall lists.
<code>rank_distance(distances, pool_items, ...[, ...])</code>	Calculate percentile rank of transition distances.
<code>rank_distance_shifted(distances, max_shift, ...)</code>	Calculate percentile rank of shifted distances.

psifr.transitions.percentile_rank

`psifr.transitions.percentile_rank(actual, possible)`

Get percentile rank of a score compared to possible scores.

Parameters

- **actual** (*float*) – Score to be ranked. Generally a distance score.
- **possible** (*numpy.ndarray* or *list*) – Possible scores to be compared to.

Returns

rank – Rank scaled to range from 0 (low score) to 1 (high score).

Return type

float

Examples

```
>>> from psifr import transitions
>>> actual = 3
>>> possible = [1, 2, 2, 2, 3]
>>> transitions.percentile_rank(actual, possible)
1.0
```

psifr.transitions.rank_lags

`psifr.transitions.rank_lags(pool_items, recall_items, pool_label=None, recall_label=None, pool_test=None, recall_test=None, test=None)`

Calculate rank of absolute lag for free recall lists.

Parameters

- **pool_items** (*list*) – List of the serial positions available for recall in each list. Must match the serial position codes used in *recall_items*.
- **recall_items** (*list*) – List indicating the serial position of each recall in output order (NaN for intrusions).
- **pool_label** (*list*, *optional*) – List of the positions to use for calculating lag. Default is to use *pool_items*.
- **recall_label** (*list*, *optional*) – List of position labels in recall order. Default is to use *recall_items*.
- **pool_test** (*list*, *optional*) – List of some test value for each item in the pool.
- **recall_test** (*list*, *optional*) – List of some test value for each recall attempt by output position.
- **test** (*callable*) – Callable that evaluates each transition between items *n* and *n+1*. Must take test values for items *n* and *n+1* and return True if a given transition should be included.

Returns

rank – Absolute lag percentile rank for each included transition. The rank is 0 if the lag was the most distant of the available transitions, and 1 if the lag was the closest. Ties are assigned to the average percentile rank.

Return type

list

See also:

[count_lags](#)

Count actual and possible serial position lags.

Examples

```
>>> from psifr import transitions
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 2, 3, 1]]
>>> transitions.rank_lags(pool_items, recall_items)
[0.5, 0.5, nan]
```

psifr.transitions.rank_distance

`psifr.transitions.rank_distance(distances, pool_items, recall_items, pool_index, recall_index, pool_test=None, recall_test=None, test=None)`

Calculate percentile rank of transition distances.

Parameters

- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **pool_items** (*list of list*) – Unique item codes for each item in the pool available for recall.
- **recall_items** (*list of list*) – Unique item codes of recalled items.
- **pool_index** (*list of list*) – Index of each item in the distances matrix.
- **recall_index** (*list of list*) – Index of each recalled item.
- **pool_test** (*list of list, optional*) – Test value for each item in the pool.
- **recall_test** (*list of list, optional*) – Test value for each recalled item.
- **test** (*callable*) – Called as `test(prev, curr)` or `test(prev, poss)` to screen actual and possible transitions, respectively.

Returns

rank – Distance percentile rank for each included transition. The rank is 0 if the distance was the largest of the available transitions, and 1 if the distance was the smallest. Ties are assigned to the average percentile rank.

Return type

list

See also:

[count_distance](#)

Count transitions within distance bins.

Examples

```
>>> import numpy as np
>>> from psifr import transitions
>>> distances = np.array([[0, 1, 2, 2], [1, 0, 2, 2], [2, 2, 0, 3], [2, 2, 3, 0]])
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 2, 3, 1]]
>>> pool_index = [[0, 1, 2, 3]]
>>> recall_index = [[3, 1, 2, 0]]
>>> transitions.rank_distance(
...     distances, pool_items, recall_items, pool_index, recall_index
... )
[0.75, 0.0, nan]
```

psifr.transitions.rank_distance_shifted

`psifr.transitions.rank_distance_shifted(distances, max_shift, pool_items, recall_items, pool_index, recall_index, pool_test=None, recall_test=None, test=None)`

Calculate percentile rank of shifted distances.

Parameters

- **distances** (*numpy.array*) – Items x items matrix of pairwise distances or similarities.
- **max_shift** (*int*) – Maximum number of items back for which to rank distances.
- **pool_items** (*list of list*) – Unique item codes for each item in the pool available for recall.
- **recall_items** (*list of list*) – Unique item codes of recalled items.
- **pool_index** (*list of list*) – Index of each item in the distances matrix.
- **recall_index** (*list of list*) – Index of each recalled item.
- **pool_test** (*list of list, optional*) – Test value for each item in the pool.
- **recall_test** (*list of list, optional*) – Test value for each recalled item.
- **test** (*callable*) – Called as `test(prev, curr)` or `test(prev, poss)` to screen actual and possible transitions, respectively.

Returns

rank – [transitions x max_shift] array with distance percentile ranks. The rank is 0 if the distance was the largest of the available transitions, and 1 if the distance was the smallest. Ties are assigned to the average percentile rank.

Return type

`numpy.ndarray`

See also:

rank_distance

to the immediately preceding item only.

Examples

```
>>> import numpy as np
>>> from psifr import transitions
>>> distances = np.array(
...     [
...         [0, 1, 2, 2, 2],
...         [1, 0, 2, 2, 2],
...         [2, 2, 0, 3, 3],
...         [2, 2, 3, 0, 2],
...         [2, 2, 3, 2, 0],
...     ]
... )
>>> pool_items = [[1, 2, 3, 4, 5]]
>>> recall_items = [[4, 2, 3, 1]]
>>> pool_index = [[0, 1, 2, 3, 4]]
>>> recall_index = [[3, 1, 2, 0]]
```

(continues on next page)

(continued from previous page)

```
>>> transitions.rank_distance_shifted(
...     distances, 2, pool_items, recall_items, pool_index, recall_index
... )
array([[0. , 0.25],
       [1. , 1.  ]])
```

4.3.3 Iterating over transitions

<code>transitions_masker(pool_items, recall_items, ...)</code>	Iterate over transitions with masking.
<code>sequences_masker(n_transitions, pool_items, ...)</code>	Yield sequences of adjacent included transitions.

psifr.transitions.transitions_masker

`psifr.transitions.transitions_masker(pool_items, recall_items, pool_output, recall_output, pool_test=None, recall_test=None, test=None)`

Iterate over transitions with masking.

Transitions are between a “previous” item and a “current” item. Non-included transitions will be skipped. A transition is yielded only if it matches the following conditions:

- (1) Each item involved in the transition is in the pool. Items are removed from the pool after they appear as the previous item.
- (2) Optionally, an additional check is run based on test values associated with the items in the transition. For example, this could be used to only include transitions where the category of the previous and current items is the same.

The masker will yield “output” values, which may be distinct from the item identifiers used to determine item repeats.

Parameters

- **pool_items** (*list*) – Items available for recall. Order does not matter. May contain repeated values. Item identifiers must be unique within pool.
- **recall_items** (*list*) – Recalled items in output position order.
- **pool_output** (*list*) – Output values for pool items. Must be the same order as pool.
- **recall_output** (*list*) – Output values in output position order.
- **pool_test** (*list, optional*) – Test values for items available for recall. Must be the same order as pool.
- **recall_test** (*list, optional*) – Test values for items in output position order.
- **test** (*callable, optional*) – Used to test whether individual transitions should be included, based on test values.

test(prev, curr) - test for included transition

test(prev, poss) - test for included possible transition

Yields

- **output** (*int*) – Output position of this transition. The first transition is 1.
- **prev** (*object*) – Output value for the “from” item on this transition.

- **curr** (*object*) – Output value for the “to” item.
- **poss** (*numpy.array*) – Output values for all possible valid “to” items.

Examples

```
>>> from psifr import transitions
>>> pool = [1, 2, 3, 4, 5, 6]
>>> recs = [6, 2, 3, 6, 1, 4]
>>> masker = transitions.transitions_masker(
...     pool_items=pool, recall_items=recs, pool_output=pool, recall_output=recs
... )
>>> for output, prev, curr, poss in masker:
...     print(output, prev, curr, poss)
1 6 2 [1 2 3 4 5]
2 2 3 [1 3 4 5]
5 1 4 [4 5]
```

psifr.transitions.sequences_masker

psifr.transitions.sequences_masker(*n_transitions*, *pool_items*, *recall_items*, *pool_output*, *recall_output*, *pool_test*=None, *recall_test*=None, *test*=None)

Yield sequences of adjacent included transitions.

Parameters

- **n_transitions** (*int*) – Number of transitions to include in yielded sequences.
- **pool_items** (*list*) – Items available for recall. Order does not matter. May contain repeated values. Item identifiers must be unique within pool.
- **recall_items** (*list*) – Recalled items in output position order.
- **pool_output** (*list*) – Output values for pool items. Must be the same order as pool.
- **recall_output** (*list*) – Output values in output position order.
- **pool_test** (*list*, *optional*) – Test values for items available for recall. Must be the same order as pool.
- **recall_test** (*list*, *optional*) – Test values for items in output position order.
- **test** (*callable*, *optional*) – Used to test whether individual transitions should be included, based on test values.
 - test(prev, curr) - test for included transition
 - test(prev, poss) - test for included possible transition

Yields

- **output** (*int*) – Output positions of included transitions. The first transition is 1.
- **prev** (*list*) – Output values for the “from” item in included transitions.
- **curr** (*list*) – Output values for the “to” item in included transitions.
- **poss** (*list of numpy.ndarray*) – Output values for all possible valid “to” items in included transitions.

See also:

transitions_masker

Yield included transitions.

Examples

```
>>> from psifr import transitions
>>> pool = [1, 2, 3, 4, 5, 6]
>>> recs = [6, 2, 3, 6, 1, 4, 5]
>>> masker = transitions.sequences_masker(
...     2, pool_items=pool, recall_items=recs, pool_output=pool, recall_output=recs
... )
>>> for output, prev, curr, poss in masker:
...     print(output, prev, curr, poss)
[1, 2] [6, 2] [2, 3] [array([1, 2, 3, 4, 5]), array([1, 3, 4, 5])]
[5, 6] [1, 4] [4, 5] [array([4, 5]), array([5])]
```

```
>>> pool = [1, 2, 3, 4]
>>> recs = [4, 3, 1, 2]
>>> masker = transitions.sequences_masker(
...     3, pool_items=pool, recall_items=recs, pool_output=pool, recall_output=recs
... )
>>> for output, prev, curr, poss in masker:
...     print(output, prev, curr, poss)
[1, 2, 3] [4, 3, 1] [3, 1, 2] [array([1, 2, 3]), array([1, 2]), array([2])]
```

4.4 Outputs

4.4.1 Counting recalls by serial position and output position

<i>count_outputs</i> (list_length, pool_items, ...)	Count actual and possible recalls for each output position.
-----------------------------------------------------	-------------------------------------------------------------

psifr.outputs.count_outputs

`psifr.outputs.count_outputs(list_length, pool_items, recall_items, pool_label, recall_label, pool_test=None, recall_test=None, test=None, count_unique=False)`

Count actual and possible recalls for each output position.

Parameters

- **list_length** (*int*) – Number of items in each list.
- **pool_items** (*list*) – List of the serial positions available for recall in each list. Must match the serial position codes used in *recall_items*.
- **recall_items** (*list*) – List indicating the serial position of each recall in output order (NaN for intrusions).

- **pool_label** (*list*) – List of the positions to use for calculating lag. Default is to use *pool_items*.
- **recall_label** (*list*) – List of position labels in recall order. Default is to use *recall_items*.
- **pool_test** (*list*, *optional*) – List of some test value for each item in the pool.
- **recall_test** (*list*, *optional*) – List of some test value for each recall attempt by output position.
- **test** (*callable*) – Callable that evaluates each transition between items *n* and *n+1*. Must take test values for items *n* and *n+1* and return True if a given transition should be included.
- **count_unique** (*bool*) – If true, possible recalls with the same label will only be counted once.

Returns

- **actual** (*numpy.ndarray*) – [outputs x inputs] array of actual recall counts.
- **possible** (*numpy.ndarray*) – [outputs x inputs] array of possible recall counts.

Examples

```
>>> from psifr import outputs
>>> pool_items = [[1, 2, 3, 4]]
>>> recall_items = [[4, 2, 3, 1]]
>>> actual, possible = outputs.count_outputs(
...     4, pool_items, recall_items, pool_items, recall_items
... )
>>> actual
array([[0, 0, 0, 1],
       [0, 1, 0, 0],
       [0, 0, 1, 0],
       [1, 0, 0, 0]])
>>> possible
array([[1, 1, 1, 1],
       [1, 1, 1, 0],
       [1, 0, 1, 0],
       [1, 0, 0, 0]])
```

4.4.2 Iterating over output positions

<code>outputs_masker(pool_items, recall_items, ...)</code>	Iterate over valid outputs.
------------------------------------------------------------	-----------------------------

psifr.outputs.outputs_masker

`psifr.outputs.outputs_masker(pool_items, recall_items, pool_output, recall_output, pool_test=None, recall_test=None, test=None)`

Iterate over valid outputs.

Parameters

- **pool_items** (*list*) – Items available for recall. Order does not matter. May contain repeated values. Item identifiers must be unique within pool.
- **recall_items** (*list*) – Recalled items in output position order.
- **pool_output** (*list*) – Output values for pool items. Must be the same order as pool.
- **recall_output** (*list*) – Output values in output position order.
- **pool_test** (*list*, *optional*) – Test values for items available for recall. Must be the same order as pool.
- **recall_test** (*list*, *optional*) – Test values for items in output position order.
- **test** (*callable*, *optional*) – Used to test whether output recalls and possible recalls should be included, based on their test values.

Yields

- **curr** (*object*) – Output value for the item at this valid output position.
- **poss** (*numpy.array*) – Output values for all possible items that could be recalled at this output position.
- **output** (*int*) – Current output position.

Examples

```
>>> from psifr import outputs
>>> pool_items = [1, 2, 3, 4]
>>> recall_items = [4, 2, 3, 1]
>>> masker = outputs.outputs_masker(
...     pool_items, recall_items, pool_items, recall_items
... )
>>> for curr, poss, output in masker:
...     print(curr, poss, output)
4 [1 2 3 4] 1
2 [1 2 3] 2
3 [1 3] 3
1 [1] 4
```

DEVELOPMENT

5.1 Transitions

Psifr has a core set of tools for analyzing transitions in free recall data. These tools focus on measuring what transitions actually occurred, and which transitions were possible given the order in which participants recalled items.

5.1.1 Actual and possible transitions

Calculating a conditional response probability involves two parts: the frequency at which a given event actually occurred in the data and frequency at which a given event could have occurred. The frequency of possible events is calculated conditional on the recalls that have been made leading up to each transition. For example, a transition between item i and item j is not considered “possible” in a CRP analysis if item i was never recalled. The transition is also not considered “possible” if, when item i is recalled, item j has already been recalled previously.

Repeated recall events are typically excluded from the counts of both actual and possible transition events. That is, the transition event frequencies are conditional on the transition not being either to or from a repeated item.

Calculating a CRP measure involves tallying how many transitions of a given type were made during a free recall test. For example, one common measure is the serial position lag between items. For a list of length N , possible lags are in the range $[-N + 1, N - 1]$. Because repeats are excluded, a lag of zero is never possible. The count of actual and possible transitions for each lag is calculated first, and then the CRP for each lag is calculated as the actual count divided by the possible count.

5.1.2 The transitions masker

The `transitions_masker()` is a generator that makes it simple to iterate over transitions while “masking” out events such as intrusions of items not on the list and repeats of items that have already been recalled.

On each step of the iterator, the output position and the previous, current, and possible items are yielded. The *output position* is the position of the transition in the recall sequence, starting from one, including all repeats and intrusions. The *previous* item is the item being transitioned from. The *current* item is the item being transitioned to. The *possible* items includes an array of all items that were valid to be recalled next, given the recall sequence up to that point (not including the current item).

```
In [1]: from psifr.transitions import transitions_masker
```

```
In [2]: pool = [1, 2, 3, 4, 5, 6]
```

```
In [3]: recs = [6, 2, 3, 6, 1, 4]
```

(continues on next page)

(continued from previous page)

```

In [4]: masker = transitions_masker(
...:     pool_items=pool, recall_items=recs, pool_output=pool, recall_output=recs
...: )
...:

In [5]: for op, prev, curr, poss in masker:
...:     print(op, prev, curr, poss)
...:
1 6 2 [1 2 3 4 5]
2 2 3 [1 3 4 5]
5 1 4 [4 5]

```

Only valid transitions are yielded, so the code for a specific analysis only needs to calculate the transition measure of interest and count the number of actual and possible transitions in each bin of interest.

Four inputs are required:

pool_items

List of identifiers for all items available for recall. Identifiers can be anything that is unique to each item in the list (e.g., serial position, a string representation of the item, an index in the stimulus pool).

recall_items

List of identifiers for the sequence of recalls, in order. Valid recalls must match an item in *pool_items*. Other items are considered intrusions.

pool_output

Output codes for each item in the pool. This should be whatever you need to calculate your transition measure.

recall_output

Output codes for each recall in the sequence of recalls.

By using different values for these four inputs and defining different transition measures, a wide range of analyses can be implemented. All conditional response probability and rank analyses are implemented using the same core of the transitions masker.

5.1.3 The sequences masker

Some analyses, such as *lag_crp_compound()* and *distance_rank_shifted()*, require examining longer sequences rather than individual one-step transitions. These analyses are implemented using the *sequences_masker()*, which allows analysis code to iterate over sequences of a specified length. Only contiguous sequences are yielded; repeats or intrusions will interrupt the sequence and start it over at the next recall.

Any other conditions, applied using the *test* input, must also apply to every transition in the sequence for that sequence to be yielded by the masker. This makes it simple to, for example, run an analysis that only examines sequences of within-category transitions. See the *user guide* for an example of conditionalizing a sequence analysis.

Similarly to the *transitions_masker()*, the sequences masker will yield the output position, previous item, current item, and possible items for each transition. Here, however, there is a list of values corresponding to positions within the sequence. For example, *curr[-1]* is the “current” item for the most recent transition, and *prev[-2]* is the “previous” item for the prior transition.

```

In [6]: from psifr.transitions import sequences_masker

In [7]: pool = [1, 2, 3, 4, 5, 6]

```

(continues on next page)

(continued from previous page)

```
In [8]: recs = [6, 2, 3, 6, 1, 4, 5]

In [9]: masker = sequences_masker(
....:     2, pool_items=pool, recall_items=recs, pool_output=pool, recall_output=recs
....: )
....:

In [10]: for output, prev, curr, poss in masker:
....:     print(output, prev, curr, poss)
....:
[1, 2] [6, 2] [2, 3] [array([1, 2, 3, 4, 5]), array([1, 3, 4, 5])]
[5, 6] [1, 4] [4, 5] [array([4, 5]), array([5])]
```

From these outputs, it is then relatively simple to do things like calculate response probabilities conditionalized on prior transitions (like in the compound lag-CRP analysis) or measure distances to recalls before the just-recalled item (like in the shifted distance rank analysis).

REFERENCES

BIBLIOGRAPHY

- [HK02] Marc W Howard and Michael J Kahana. When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, 46(1):85–98, 2002. doi:10.1006/jmla.2001.2798.
- [Kah96] Michael Jacob Kahana. Associative retrieval processes in free recall. *Memory & Cognition*, 24(1):103–109, 1996. doi:10.3758/bf03197276.
- [LK14] Lynn J. Lohman and Michael J. Kahana. Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1):12, 2014. doi:10.1037/a0033698.
- [Mor20] Neal Morton. Psifr: Analysis and visualization of free recall data. *Journal of Open Source Software*, 5(54):2669, 2020. doi:10.21105/joss.02669.
- [MP16] Neal W Morton and Sean M Polyn. A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, 86:119–140, 2016. doi:10.1016/j.jml.2015.10.002.
- [MP17] Neal W Morton and Sean M Polyn. Beta-band activity represents the recent past during episodic encoding. *NeuroImage*, 147:692–702, 2017. doi:10.1016/j.neuroimage.2016.12.049.
- [Mur62] Bennet B Murdock. The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482–488, 1962. doi:10.1037/h0045106.
- [PNK09] Sean M Polyn, Kenneth A Norman, and Michael Jacob Kahana. A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1):129–156, 2009. doi:10.1037/a0014420.
- [PEK11] Sean M. Polyn, Gennady Erlikhman, and Michael J. Kahana. Semantic cuing and the scale insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3):766, 2011. doi:10.1037/a0022475.
- [RTB71] Daniel L Roenker, Charles P Thompson, and Sam C Brown. Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 76(1):45–48, 01 1971. doi:10.1037/h0031355.
- [RKT16] Sandro Romani, Mikhail Katkov, and Misha Tsodyks. Practice makes perfect in memory recall. *Learning & Memory*, 23(4):169–173, 2016. doi:10.1101/lm.041178.115.
- [SBW+02] John L Stricker, Gregory G Brown, John T Wixted, Juliana V Baldo, and Dean Delis. New semantic and serial clustering indices for the California Verbal Learning Test–Second Edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, 8:425–435, 2002. doi:10.1017/S1355617702813224.

Symbols

`__init__()` (*psifr.measures.TransitionMeasure* method), 49

A

`analyze()` (*psifr.measures.TransitionMeasure* method), 50

`analyze_subject()` (*psifr.measures.TransitionMeasure* method), 50

B

`block_index()` (*in module psifr.fr*), 33

C

`category_clustering()` (*in module psifr.fr*), 46

`category_crp()` (*in module psifr.fr*), 40

`check_data()` (*in module psifr.fr*), 24

`count_category()` (*in module psifr.transitions*), 54

`count_distance()` (*in module psifr.transitions*), 55

`count_lags()` (*in module psifr.transitions*), 51

`count_lags_compound()` (*in module psifr.transitions*), 52

`count_outputs()` (*in module psifr.outputs*), 62

D

`distance_crp()` (*in module psifr.fr*), 41

`distance_rank()` (*in module psifr.fr*), 44

`distance_rank_shifted()` (*in module psifr.fr*), 45

F

`filter_data()` (*in module psifr.fr*), 29

I

`item_query` (*psifr.measures.TransitionMeasure* attribute), 49

K

`keys` (*psifr.measures.TransitionMeasure* attribute), 49

L

`lag_crp()` (*in module psifr.fr*), 37

`lag_crp_compound()` (*in module psifr.fr*), 38

`lag_rank()` (*in module psifr.fr*), 43

M

`merge_free_recall()` (*in module psifr.fr*), 25

`merge_lists()` (*in module psifr.fr*), 28

O

`outputs_masker()` (*in module psifr.outputs*), 64

P

`percentile_rank()` (*in module psifr.transitions*), 56

`pli_list_lag()` (*in module psifr.fr*), 36

`plot_distance_crp()` (*in module psifr.fr*), 48

`plot_lag_crp()` (*in module psifr.fr*), 48

`plot_raster()` (*in module psifr.fr*), 47

`plot_spc()` (*in module psifr.fr*), 48

`plot_swarm_error()` (*in module psifr.fr*), 48

`pnr()` (*in module psifr.fr*), 35

`pool_index()` (*in module psifr.fr*), 32

R

`rank_distance()` (*in module psifr.transitions*), 58

`rank_distance_shifted()` (*in module psifr.transitions*), 59

`rank_lags()` (*in module psifr.transitions*), 57

`reset_list()` (*in module psifr.fr*), 30

S

`sequences_masker()` (*in module psifr.transitions*), 61

`spc()` (*in module psifr.fr*), 34

`split_lists()` (*in module psifr.fr*), 31

`split_lists()` (*psifr.measures.TransitionMeasure* method), 50

T

`table_from_lists()` (*in module psifr.fr*), 23

`test` (*psifr.measures.TransitionMeasure* attribute), 49

`TransitionMeasure` (class *in psifr.measures*), 49

`transitions_masker()` (*in module psifr.transitions*), 60