

## Supplemental Material S2. Response confusions.

Analysis of accuracy in the main text focused largely on overall accuracy. However, on most trials there was one or more lexical competitors on the screen. Here we report the pattern of confusions: when subjects were incorrect were they more likely to select a cohort or rhyme competitor than an unrelated? If they were not, this might imply that the lack of accuracy was due to guessing.

Note that for the speech response trials in the dual task and for all the pupillometry trials, one complicating factor in computing these analyses is that not all trials had both types of competitors. Consider a set like *sandal*, *sandwich*, *candle*, and *necklace*. When *sandal* was the target, there would be both a cohort (*sandwich*) and rhyme (*candle*)—we term this a TCR trial. However, when *sandwich* is the target, there is a cohort (*sandal*) but no rhyme (a TC trial). And when *candle* is the target there is a rhyme (*sandal*) but no cohort.

Thus, to compute confusions, we computed the average likelihood of selecting each type of competitor separately on each trial type. We then averaged corresponding trial types (e.g., the proportion of cohort selections was the average of cohort selections on the TC and TCR trials). We always used the designated unrelated item (*necklace*) as the unrelated. As these analyses were exploratory and descriptive no inferential statistics were conducted.

Each experiment is broken down in detail below. Across experiments, we consistently see that the similar-sounding competitors are the source of more errors than the unrelated option.

### S2.1 Experiment 1

Table S2.1.1 presents the proportion of total responses from each trial type for each competitor option in the dual task. It is clear that the competitor grid in the grid-response trials was more distracting than the unrelated distractor grids. Across the speech response trials, the cohort and rhyme were more often the source of errors over the unrelated image. This is also true in the pupillometry task (Table S2.1.2).

Table S2.1.1 Proportion of total responses to each competitor for each trial type in the dual task.

Listening Condition	Grid response		Speech response		
	Competitor	Unrelated	Cohort	Rhyme	Unrelated
No audio	0.135	0.02			
Unmodified	0.136	0.02	0.001	0.001	0.0004
8-channel	0.14	0.02	0.01	0.04	0.02
4-channel	0.15	0.03	0.07	0.08	0.04

Table S2.1.2 Proportion of total responses to incorrect options for each listening condition in the pupillometry task.

Listening condition	Cohort	Rhyme	Unrelated
Unmodified	0.0006	0.0008	0.0003
8-channel	0.004	0.02	0.002
4-channel	0.02	0.06	0.007

## S2.2 Experiment 2

Table S2.2.1 presents the proportion of total responses for each listening condition to each of the competitor options. As in Experiment 1, the cohort and rhyme are more often the source of errors (with rhymes being chosen more often than cohorts) compared to unrelated images.

Table S2.2.1 Proportion of total responses to incorrect options for each listening condition in the pupillometry task.

Listening condition	Cohort	Rhyme	Unrelated
Unmodified	0.0005	0.001	0.0005
8-channel	0.04	0.09	0.03
4-channel	0.009	0.04	0.006

## S2.3 Experiment 3

Finally, Table S2.3.1 presents the proportions of total responses from each trial type to the competitor options in the dual task. In the grid-response trials, the competitor grid is selected in error more often than the unrelated distractor grids. In the speech-response trials, cohorts and rhymes are again selected more often than the unrelated image. This is also true in the pupillometry task (Table S2.3.2).

Table S2.3.1 Proportion of total responses to incorrect options for each trial type in the dual task.

Listening Condition	Grid response		Speech response		
	Competitor	Unrelated	Cohort	Rhyme	Unrelated
No audio	0.13	0.02			
Unmodified	0.14	0.02	0.0	0.001	0.0004
8-channel	0.14	0.03	0.02	0.06	0.009
4-channel	0.14	0.03	0.05	0.10	0.04

Table S2.3.2 Proportion of total responses to incorrect options for each listening condition in the pupillometry task.

Listening condition	Cohort	Rhyme	Unrelated
Unmodified	0.0008	0.001	0.0006
8-channel	0.02	0.04	0.008
4-channel	0.03	0.09	0.02