The *“Crossword Effect”* in free word recall:

A retrieval advantage for words encoded in line with their spatial associations

Christoph Scheepers

Ricarda Senger

University of Glasgow

christoph.scheepers@glasgow.ac.uk

Theoretical context

- **Embodied Cognition**
- **Perceptual symbol hypothesis (e.g., Barsalou, 1999; 2008)**
  - Conceptual representations of symbols (words, numbers, etc.) are grounded in *perception* and *action.*
  - Concepts are understood via the mental re-enactment of perceptual, motor and introspective states acquired during experience with the world, body, and mind.
- **Here, we are specifically interested in how word concepts relate to horizontal / vertical space**
- Indeed, there is evidence that spatial associations of words affect sensori-motor processing (e.g.) in tasks involving eye-movements…

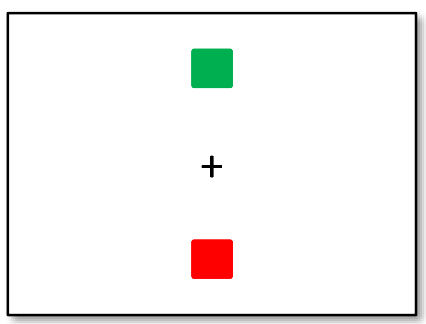
Example I

Ben

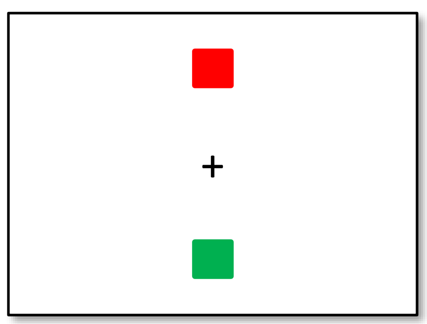
- Eye-tracking experiments by my former PhD student, Dr. Benjamin Dunn (photo), showed that saccades towards spatially compatible locations (e.g., hearing ‘moon’ and looking up; hearing ‘boots’ and looking down) are facilitated relative to a neutral control condition…

Example I

- Among plenty of non-word fillers trials (e.g., "blosh"), participants encountered trials where they would hear either:
  - A spoken word with *upward* spatial association (e.g., "moon")
  - A spoken word with *downward* spatial association (e.g., "boots")
  - A *spatially neutral* spoken word (e.g., "cheese")
- ***Ocular* lexical decision task (eye-tracking):** *"Look at the green square if it's a word, red square if it's not a word"*

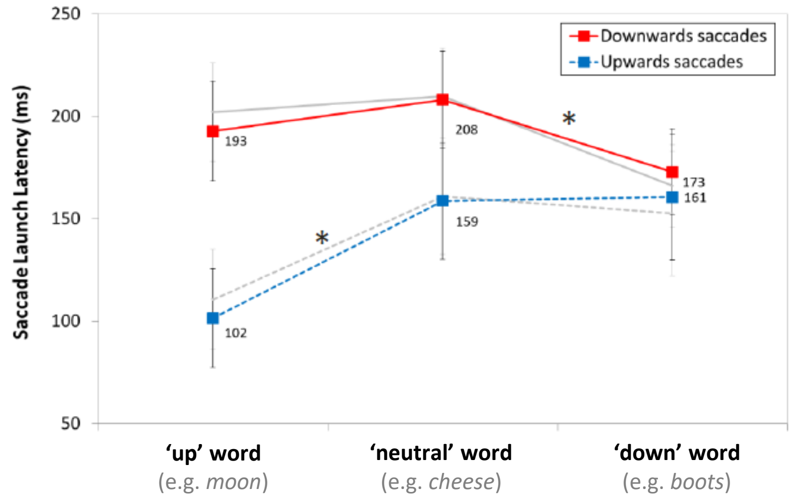


50% trials



50% trials

Example I



In grey: EM means without adjusting for control predictors

- **Effects in *Saccade Launch Latency:***
  - Generally, correct downwards saccades take longer to launch.
  - Relative to 'neutral' words (e.g., *cheese*):
    - Correct upwards saccades are launched quicker following 'up' words (e.g., *moon*)
    - Correct downwards saccades are launched quicker following 'down' words (e.g., *boots*)
- **Facilitation of *concept-compatible* saccadic response!**

Example II

- **Spivey & Geng (2001)** played pre-recorded descriptions of spatiotemporally dynamic scenes (see table on the right) and recorded participants eye-movements while they listened to the descriptions and faced a **blank screen.**
- They ensured that participants were not aware of their eye-movements being monitored.

Table 1 Pre-recorded scene descriptions from Experiment 1 (text in italics indicates sentences during which eye movements were analyzed)

Upward story	"Imagine that you are standing across the street from a 40 story apartment building. At the bottom there is a doorman in blue. *On the 10th floor, a woman is hanging her laundry out the window. On the 29th floor, two kids are sitting on the fire escape smoking cigarettes. On the very top floor, two people are screaming.*"
Downward story	"Imagine you are standing at the top of a canyon. Several people are preparing to rappel down the far canyon wall across from you. *The first person descends 10 feet before she is brought back to the wall. She jumps again and falls 12 feet. She jumps another 15 feet. And the last jump, of 8 feet, takes her to the canyon floor.*"
Leftward story	"Imagine a train extending outwards to the left. It is pointed to the right, and you are facing the side of the engine. It is not moving. *Five cars down is a cargo holder with pink graffiti sprayed on its side. Another six cars down is a flat car. The train begins to move. Further down the train you see the caboose coming around a corner.*"
Rightward story	"Imagine a fishing boat floating on the ocean. It's facing leftward from your perspective. *At the back of the boat is a fisherman with a fishing pole. The pole extends about 10 feet to the right beyond the edge of the boat. And from the end of the pole, the fishing line extends another 50 feet off to the right before finally dipping into the water.*"
Control story	"Imagine you are on a hill looking at a city through a telescope. *Pressing a single button zooms a specific block into view. Another button brings a gray apartment building into focus. Finally a third button zooms in on a single window. Inside you see a family having breakfast together. A puppy appears and begs for a piece of French toast.*"

Example II

- Results showed **significant saccadic biases** (on an uninformative blank screen!) in the direction implied by the described scenario
- These saccadic biases emerged 'spontaneously' in response to the scene descriptions.

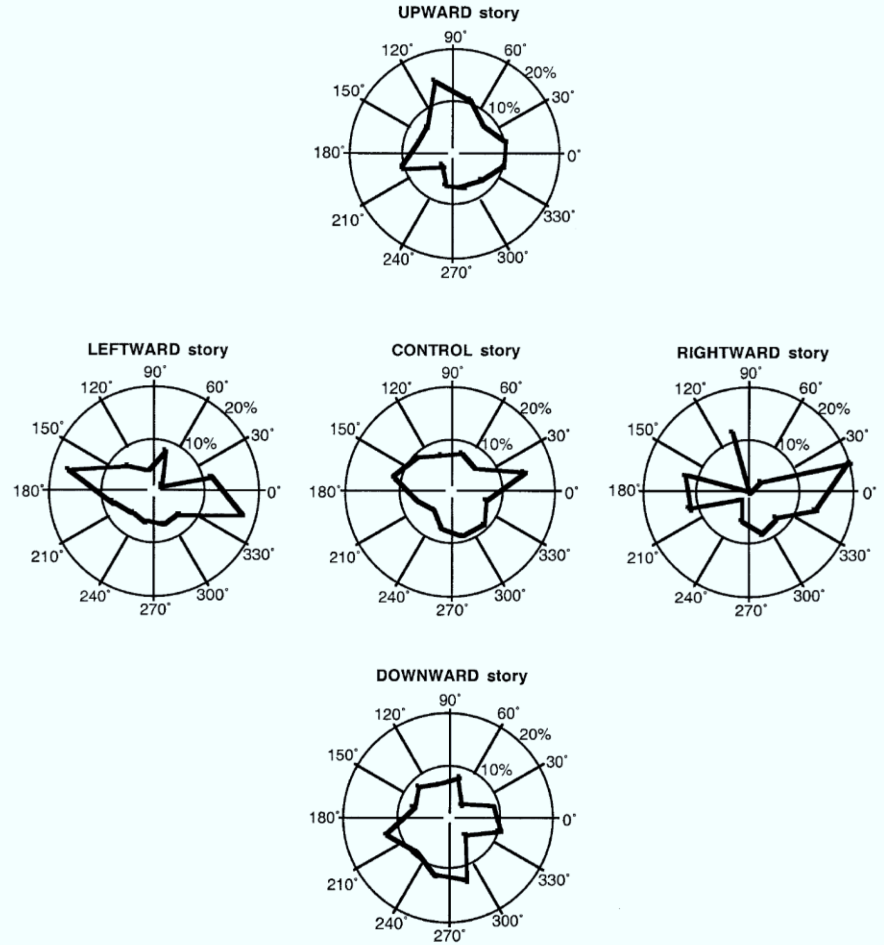


Fig. 1 Polar plots of the percentage of saccades in all directions during scene descriptions. Saccades were pooled over sentences in the scene description that exhibited explicit directionality (or over all sentences during the control scene description). For example, the upper plot is taken only from saccades during sentences 3–5 of the upward scene description (see Methods for scene description). While facing a blank screen, participants' eye movements showed a clear bias toward the direction of the spatio-temporal imagery in the scene description

Interim conclusion

- These examples and others (e.g., Dutschig et al., 2013,; Kamide et al., 2016) indicate that perceivers spontaneously generate spatial mental models of linguistic concepts which trigger (Example II) or facilitate (Example I) oculomotor responses in the direction implied by the spatial mental model.
- **Support for the notion of *perceptual symbols***
  - **Spivey & Geng (2001):** *More broadly,* […] *results point to a concrete embodiment of cognition, in that a construction of a mental image is almost "acted out" by the eye movements*
- However, while such findings demonstrate that linguistic concepts activate perceptual representations (measurable via eye-movements), **evidence for a link *in the other direction* (from perceptual encoding to accessing word concepts) is still rather sparse to date.**
- This is what the present study is trying to assess …

Present research question

"Crosswords"

- Words can be printed (and thus encoded…)

L	I	K	E		T	H	I	S
I								
K								
E								
T								
H								
I								
S								

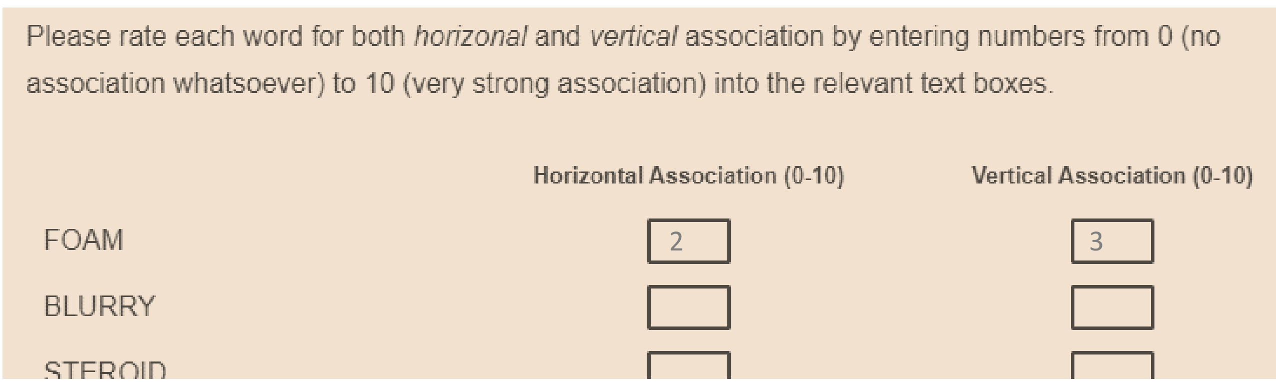
- **Does this matter for free word recall, particularly if words have different spatial associations?**
- Consider, e.g., **CARPET (↔)** vs. **LADDER (↕)**
- If word concepts are *amodal* symbols, there is no reason why it should matter, apart from vertical reading being more difficult in general (Yu et al., 2010).
- If word concepts are *perceptual*, it could matter in the sense that perceptual encoding of the word itself may support accessing its meaning.

Hypothesis

- **A word's spatial association (horizontal vs. vertical) and how it is encoded (horizontal vs. vertical reading) should *interact* when predicting word recall performance:**
  - Recall performance should be better when words are presented *congruently* (rather than incongruently) with their spatial associations, i.e.
    - 'Horizontal words' (e.g., CARPET) presented/read horizontally
    - 'Vertical words' (e.g., LADDER) presented/read vertically
- Recall performance can be measured in two (mutually non-exclusive) ways:
  - (Likelihood of) **correct word recall** (yes / no); more = better
  - **Retrieval rank** of recalled word in the recall list; earlier = better
- We postulated that the predicted interaction must be confirmed at $p < .025$ in at least one of the two measures (or both)

Spatial associations of words

- 129 initial candidate words
- **Pre-tested in an on-line questionnaire (Qualtrics©)**
- 91 participants rated each candidate word (0 to 10) for both *horizontal* and *vertical* association, like so:


Please rate each word for both *horizonal* and *vertical* association by entering numbers from 0 (no association whatsoever) to 10 (very strong association) into the relevant text boxes.

	Horizontal Association (0-10)	Vertical Association (0-10)
FOAM	2	3
BLURRY		
STEROID		

- **Item selection**
- 43 Horizontal Association (HA) words, e.g. PUNCH, ROAD, RUNWAY, HORIZON, CARPET, …
  - mean horizontal rating > 7.0 and mean vertical rating < 3.0
- 51 Vertical Association (VA) words, e.g. RAIN, ROCKET, SHOWER, LADDER, TREE, …
  - mean horizontal rating < 3.0 and mean vertical rating > 7.0

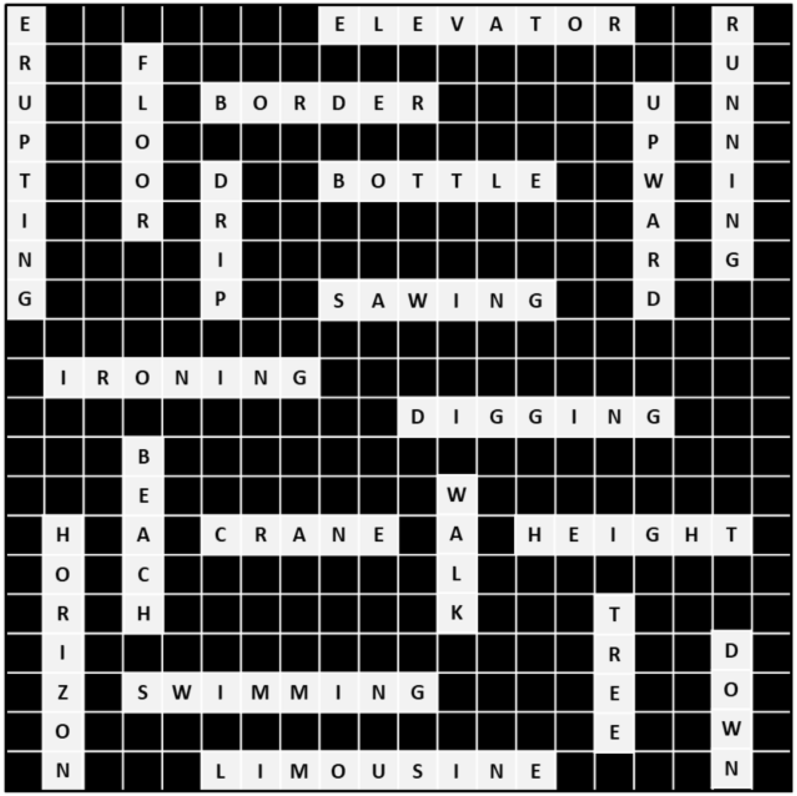
Item norms

Table 1. Mean item norms (*SD*s in brackets) for the 43 HA and 51 VA words used in Experiment 1. Shown are pre-test ratings for horizontal and vertical spatial association (H-rating and V-rating, respectively), as well as figures for the control variables *number of letters* (Letters), *number of syllables* (Syllables), *lexical frequency* (Freq.), and *concreteness* (Concr.). The bottom row shows *Hedge's g* (standardised effect size adjusted for unequal sample sizes) for the difference between HA and VA words in each variable.

	Spatial Assoc. Pre-test		Control Variables			
	H-rating	V-rating	Letters	Syllables	Freq.	Concr.
HA Words	7.63 (0.48)	2.15 (0.55)	6.28 (2.16)	1.81 (0.82)	0.65 (1.07)	4.35 (0.54)
VA Words	1.92 (0.57)	8.11 (0.56)	6.10 (1.70)	1.84 (0.76)	0.75 (0.93)	4.15 (0.68)
Hedge's g	10.76	–10.73	0.09	–0.04	–0.10	0.32

Word arrays for encoding

- For encoding, we wanted to present each participant with a ‘cross-word’ array of 20 words like this (where words actually don’t cross each other for better legibility):
- **10 horizontal association (HA) words**
  - 5 horizontally presented (here *border*, *sawing*, *ironing*, *swimming*, *limousine*)
  - 5 vertically presented (here *horizon*, *floor*, *beach*, *running*, *walk*)
- **10 vertical association (VA) words**
  - 5 horizontally presented (here *elevator*, *bottle*, *digging*, *crane*, *height*)
  - 5 vertically presented (here *erupting*, *drip*, *upward*, *tree*, *down*)



Word arrays for encoding

- To avoid results to be affected by particular word combinations, we wanted to present each participant with a **unique** combination of 10 HA and 10 VA words.
- **Problem:** How many subjects are required if we want to present *all possible* combinations of 10 HA (out of 43) plus 10 VA (out of 51) words per array?
- Answer:

$$\frac{43!}{10!\ (43-10)!} \times \frac{51!}{10!\ (51-10)!} = 2.450\ \times 10^{19}$$

  – That's about 100 million times more than the number of stars in the Milky Way galaxy (estimated to be between 100 – 400 billion)

- **Tough!...**

Word arrays for encoding

- Actual word samples for the crossword arrays were therefore determined via trial and error (brute-force random sampling with balance checks), but in a **reproducible** way using *R*
- We selected 80 samples, ensuring reasonably balanced occurrences of words
  - Each of the 94 words occurred at least 11 times and at most 24 times across the 80 samples
  - 10 HA and 10 VA words per sample
  - Half of the words per category were randomly allocated to the *horizontal* and the other half to the *vertical* presentation condition
- We then created a counterbalanced batch of 80 samples in which presentation conditions per word were flipped
- **Thus, 160 samples altogether, with reasonably balanced occurrences of words and fully counterbalanced presentation conditions per word**

Word arrays for encoding

- The item × presentation condition assignments from these samples were then used to *manually* (poor Ricarda!) create 160 crossword arrays for encoding in PowerPoint, like these two counterbalanced examples:

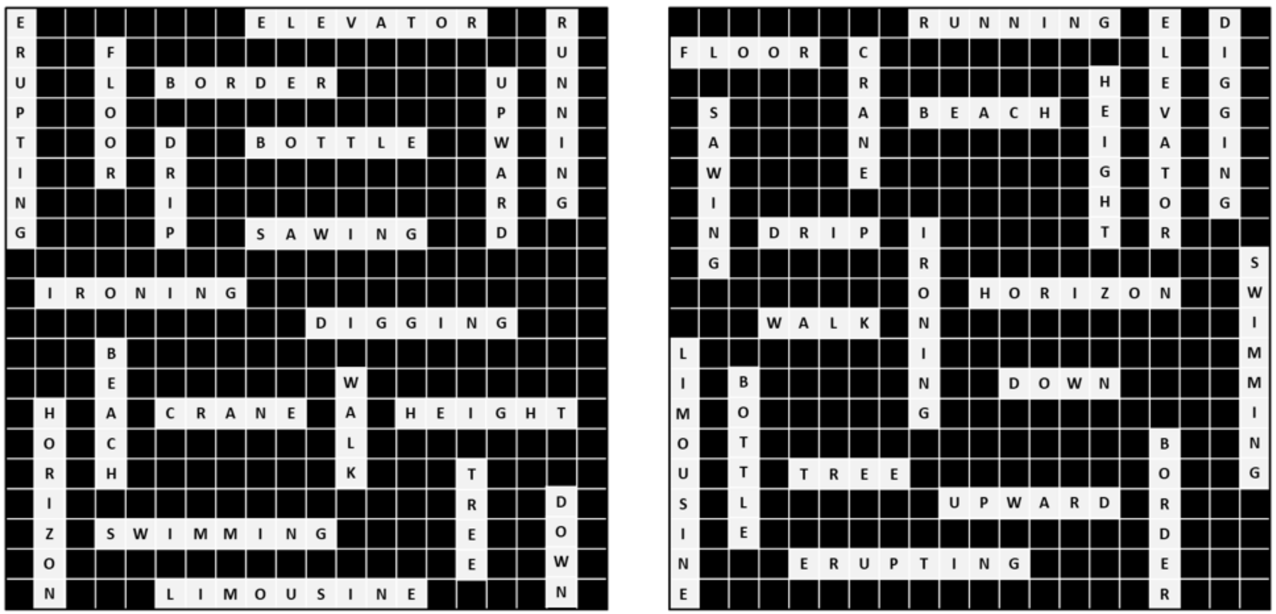


Figure 1. Example crossword arrays for encoding. The panel on the left shows an example from the first batch of 80 arrays; the panel on the right shows its second-batch counterpart in which horizontal vs. vertical presentation conditions per word were swapped. Each array comprised 10 HA words (here: *limousine*, *swimming*, *horizon*, *running*, *ironing*, *beach*, *sawing*, *floor*, *border*, and *walk*) and 10 VA words (here: *erupting*, *elevator*, *drip*, *bottle*, *upward*, *digging*, *tree*, *down*, *crane*, and *height*). Crossword arrays from the second batch were not exact 90-degree rotations of their first-batch counterparts.

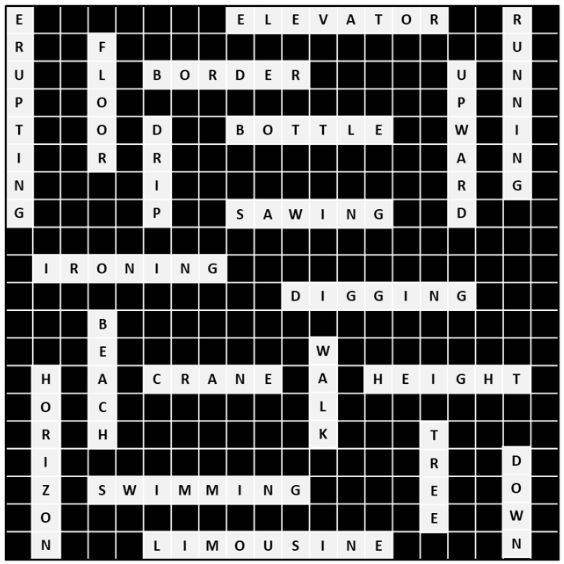
Participants and design

- Since there were 160 unique crossword arrays for encoding, we recruited **160 participants** for the experiment
  - Most of them (93%) were **not** psychology students!
- 2 × 2 design crossing the factors *word type* (HA vs. VA word) and word presentation (horizontal vs. vertical)
  - Both factors were manipulated *within-subjects*
  - Word category was *between-* and presentation *within-items* (where item = word)
  - Mixed effects model structure for analysis therefore looked like this:

```
Y ~ TYPE * PRES +
      (1 + TYPE * PRES | Subj) +
      (1 + PRES | Item)
```

Procedure

Encoding phase



130 seconds

Distractor task

$(7 + 4) \times 10 =$

$11 - 2 + 6 - 3 =$

$(27 - (8 + 1)) / 9 =$

$73 - 50 + 5 =$

$3 \times 8 - 7 =$

$(13 + (20 - 1)) / 8 =$

$(6 - 4) \times 5 =$

$(17 + 11) / 7 =$

$(5 + (23 + 27)) / 5 =$

60 seconds

"Surprize" recall phase

1. ______________
2. ______________
3. ______________
4. ______________
5. ______________
6. ______________
7. ______________
8. ______________

. . .

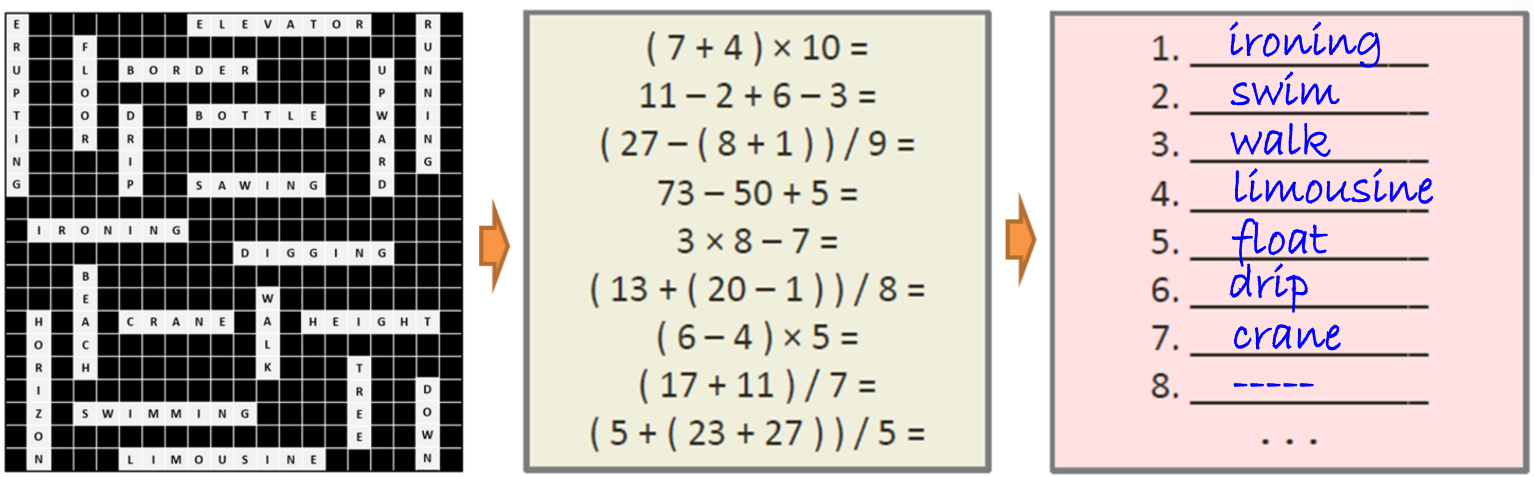
"List words you can remember"

- Presentation timing controlled in PowerPoint on a laptop.
- Experimental instructions included a mild form of deception:
  - Experiment disguised as studying "*the influence of word arrangements on mathematical problem solving*" (no mention of later word recall task)

Preregistration

- Before collecting any word recall data, we preregistered our hypotheses, design, materials, sample size, procedures, **and analyses** on the Open Science Framework, see https://osf.io/fb64q/
- Preregistration was ‘frozen’ on October 03, 2019
- Data collection took place between October 14, 2019 and January 10, 2020

Data coding (preregistered)



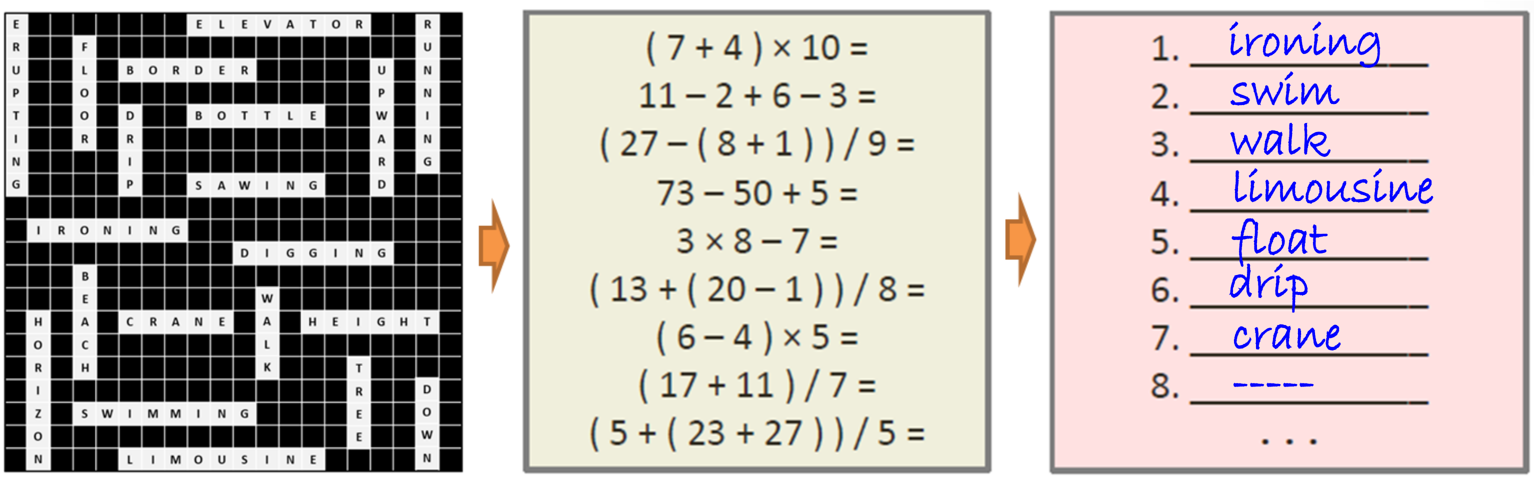
130 seconds

60 seconds

"List words you can remember"

- **Correct Word Recall:** For each word in a participant's crossword array (20 per array), we coded whether it was correctly recalled (score of 1) or not (score of 0) => 20 codes per participant
  - Word-form related deviations (e.g. "swim" instead of "swimming") were treated as **correct** (lenient coding)
  - Words in the recall list that were **not presented** (e.g., "float") or **repeated** were treated as incorrect *intrusions*.

Data coding (preregistered)



130 seconds — 60 seconds — “List words you can remember”

- **Retrieval Rank:** For each correctly recalled word in the recall list, we recorded its serial position (without excluding intrusions or repetitions first), e.g.
  - *ironing* = 1, *swim(ming)* = 2, *walk* = 3, *limousine* = 4, *drip* = 6, *crane* = 7
  - The remaining 14 words from the presentation all score *NA* on retrieval rank as they were not correctly recalled.

Data transcription

- Information from each participant's handwritten recall list was manually entered into pre-prepared data sheets

	A	B	C	D	E	F	G
1	subject	category	present	item	recalled	rank	intr.rep
2	B009	H	H	BORDER	0	0	1
3	B009	H	H	DRAGGING	0	0	1
4	B009	H	H	DRIVING	0	0	1
5	B009	H	H	RIVERBANK	0	0	1
6	B009	H	H	WATERSKIING	1	3	1
7	B009	H	V	CANOE	1	2	1
8	B009	H	V	CONVEYOR	0	0	1
9	B009	H	V	EEL	0	0	1
10	B009	H	V	PLANK	0	0	1
11	B009	H	V	PUNCH	0	0	1
12	B009	V	H	BOTTLE	0	0	1
13	B009	V	H	BUNGEE	1	4	1
14	B009	V	H	DESCENT	0	0	1
15	B009	V	H	LIFTING	0	0	1
16	B009	V	H	RAISING	0	0	1
17	B009	V	V	CLIFF	0	0	1
18	B009	V	V	DOWN	0	0	1
19	B009	V	V	TALL	0	0	1
20	B009	V	V	UPWARD	0	0	1
21	B009	V	V	YOYO	1	1	1
22	A003	H	H	BENCH	1	3	0
23	A003	H	H	DRIVING	0	0	0
24	A003	H	H	EEL	0	0	0

General Results

- Of the 20 words presented during the encoding phase, the **average participant recalled ca. 6 ± 3 words correctly**
- Only four participants (2.5%) recalled fewer than two words, and only two participants (1.3%) achieved the observed maximum of 16 correctly recalled words
- The total set of recalled words contained all 94 items (i.e., there were no “non-retrievable” items)
- Retrieval ranks of correctly recalled words ranged from 1 to 16
- 27 participant data sets (ca. 17%) had intrusions and/or repetitions in the recall list

Hypothesis-relevant results (I): correct word recall

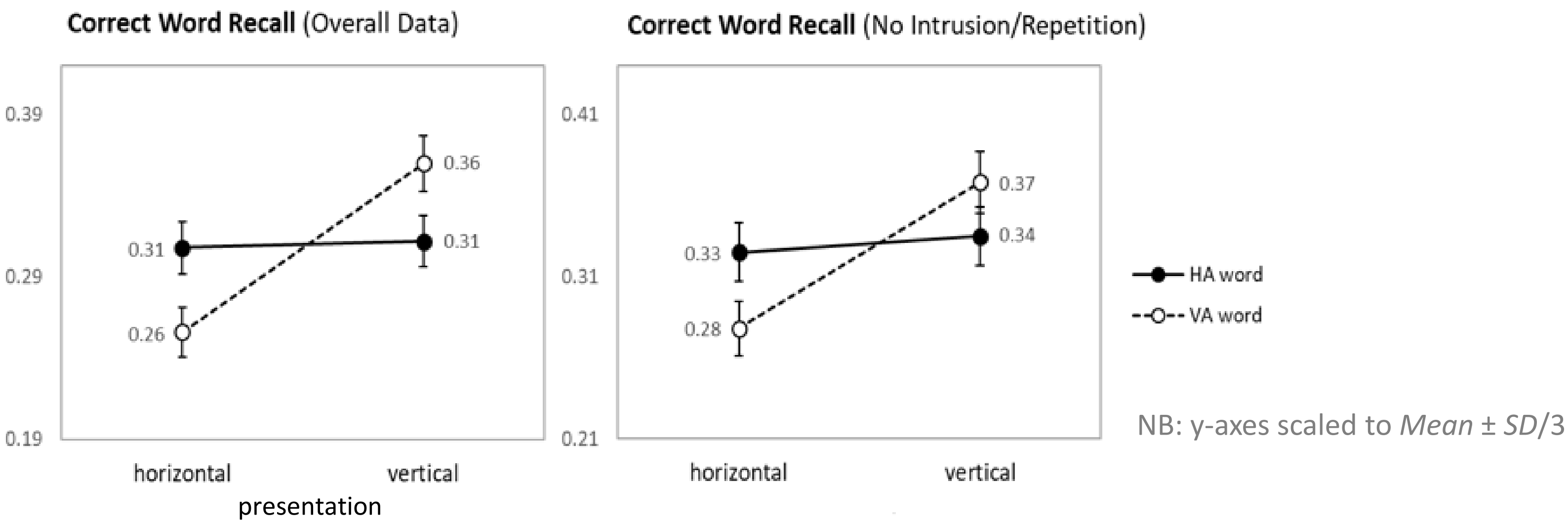
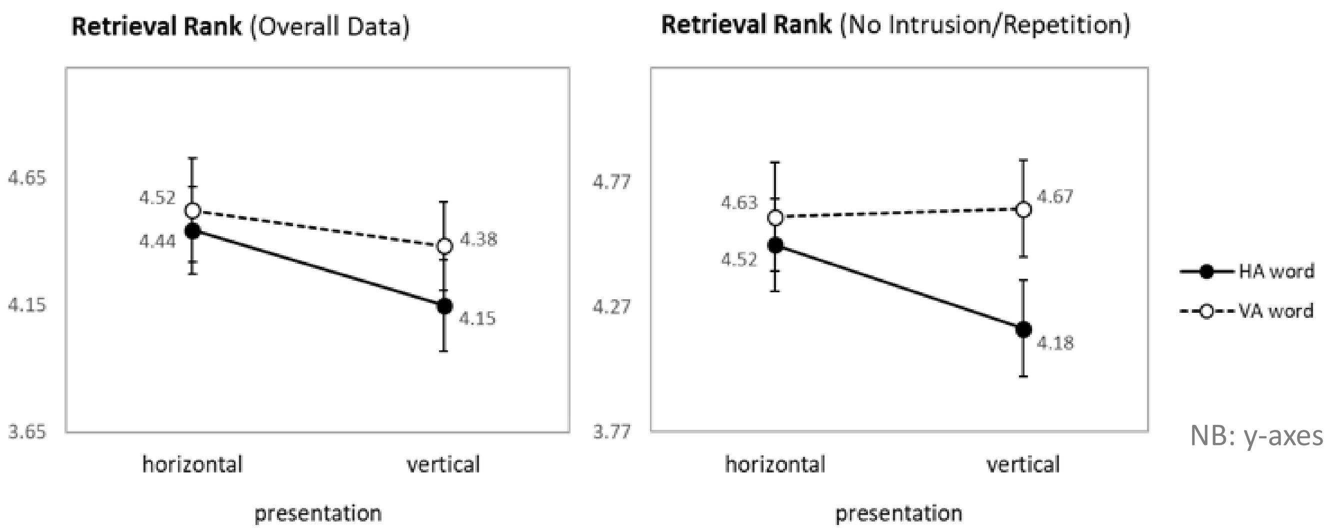


Table 2. Binary logistic mixed effects modelling results for *correct word recall* (*b* estimates and *SE*s are in *logit* units). Shown on the left are results from the overall data (3200 observations). Results after excluding 27 participant data sets with intrusions and/or repetitions in the recall list are shown on the right (2660 observations). TYPE: main effect of word category (VA – HA); PRES: main effect of presentation (vertical – horizontal); TYPE × PRES: word category × presentation interaction. Significant likelihood-ratio chi-squares are marked with asterisks: $*p \leq .025$; $**p \leq .0025$ (two-tailed). Actual *p*-values are rounded to three decimals.

	Overall Data				Without Intrusions and/or Repetitions			
Effect	*b*	*SE*	$_{LR}\chi^2(1)$	*P*	*b*	*SE*	$_{LR}\chi^2(1)$	*p*
TYPE	–0.080	0.087	0.318	.573	–0.093	0.140	0.233	.629
PRES	0.351	0.103	**11.015	< .001	0.312	0.108	*7.890	.005
TYPE × PRES	0.613	0.181	**11.262	< .001	0.515	0.193	*6.297	.012

Main hypothesis supported by the data! Boom!

Hypothesis-Relevant Results (II): Retrieval Ranks



NB: y-axes scaled to *Mean* ± *SD*/3

Table 3. Cumulative link mixed effects modelling results for *retrieval rank of recalled word* (*b* estimates and *SE*s are in *cumulative cloglog* units)[1]. Shown on the left are results for the overall data (989 observations). Results after excluding 27 participant data sets with intrusions and/or repetitions in the recall list are shown on the right (869 observations). TYPE: main effect of word category (VA – HA); PRES: main effect of presentation (vertical – horizontal); TYPE × PRES: word category × presentation interaction. *P*-values are rounded to three decimals. Significance is assumed at $p \leq .025$ (two-tailed).

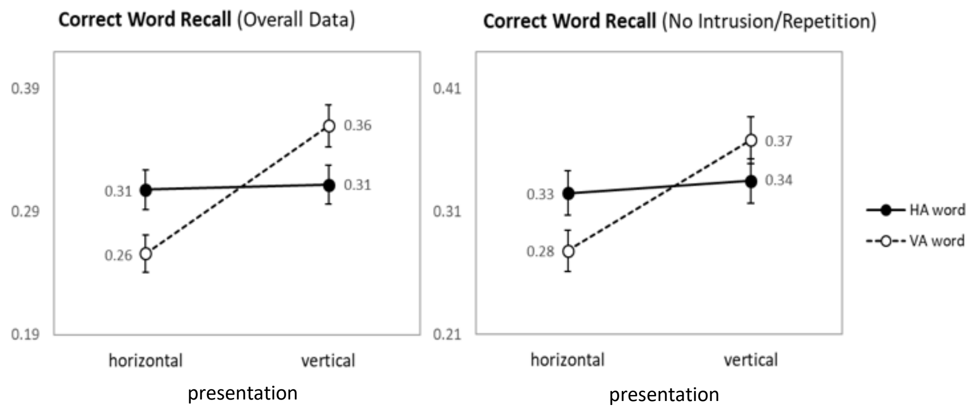
	Overall Data				Without Intrusions and/or Repetitions			
Effect	*b*	*SE*	$LR\chi^2(1)$	*p*	*b*	*SE*	$LR\chi^2(1)$	*p*
TYPE	0.001	0.078	0.001	.981	0.052	0.084	0.379	.538
PRES	–0.062	0.082	0.569	.451	–0.031	0.089	0.119	.730
TYPE × PRES	0.215	0.159	1.814	.178	0.306	0.172	3.150	.076

Not significant! Statistically inconclusive data!

Main conclusions

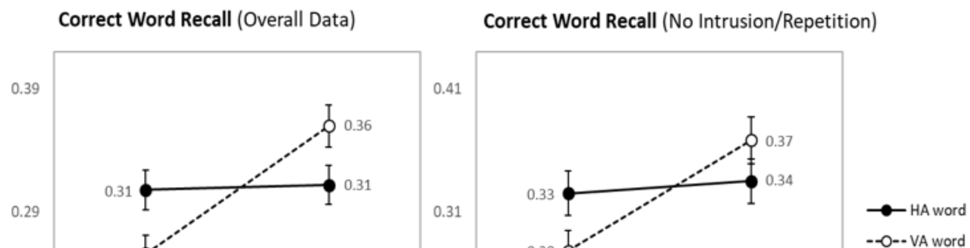
- Hypothesis confirmed in one of the two measures considered:
  - Spatial associations of words (e.g., CARPET vs. LADDER) interact with horizontal vs. vertical presentation during encoding such that **words that are encoded *congruently* with their spatial association are more likely to be recalled**.
  - Looking at simple effects, the interaction is mainly carried by VA words like LADDER
- No clear effect in retrieval ranks
  - Might suggest that the two factors are more relevant for encoding / retention of word information (=> recall likelihood) than for ease of retrieval (=> retrieval rank)
- **Overall, supporting the *perceptual symbol hypothesis* in a previously neglected domain (free word recall)**

Open questions (I)



- **Why no simple effect of presentation for HA words like CARPET?**
- One plausible explanation (currently being tested):
  - Recall performance is often enhanced when encoding requires more cognitive effort (e.g., Ellis, Thomas, & Rodriguez, 1984; Jacoby et al., 1979; Tyler, Hertel, McCallum, & Ellis, 1979)
  - **Vertical encoding is indeed very effortful** to western readers (see Yu et al. 2010), which could lead to **deeper encoding & better recall in general** (=> significant main effect of presentation)
  - The expected presentation contrast for HA words (e.g., CARPET) could have been offset by such a general recall advantage for vertically encoded words.

Open questions (I)



0.19

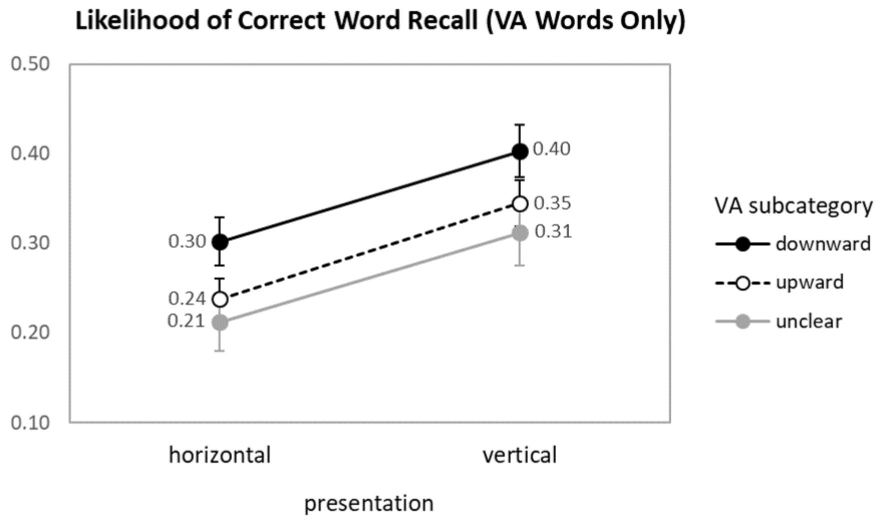
- **Why no simple**
- One plausible e
  - Recall perfor cognitive effo Tyler, Hertel,
  - **Vertical enco** al. 2010), whi (=> significan
  - The expected presentation contrast for HA words (e.g., CARPET) could have been offset by such a general recall advantage for vertically encoded words.

How to test?

- Replicate study in (say) Japanese, where readers are much more accustomed to vertical reading (general vertical presentation advantage should disappear). – Be my guest!... ☺
- Replicate study with ‘spatially unbiased’ words (e.g., DIESEL, PRODUCT, MEASURING, STENCH, SPHERE, …) – currently underway
  - If recall of spatially unbiased words benefits from vertical encoding as well, then this supports the idea that presentation effects for HA words were ‘suppressed’ by a general vertical encoding advantage.

Open questions (II)

- Is the strong presentation contrast for VA words mainly driven by *downward* association words (e.g., FALLING), in line with *vertical reading direction*?
- Descriptive analysis (based on 'intuitive' classification of VA words into *downwards*, *upwards*, *unclear*), does not seem to support this
  - *presentation* effect slopes are virtually the same:



downward association words (N = 19) include words like RAIN, BUNGEE, FALLING, DRIP, DIGGING, …

upward association words (N = 21) include words like ASCENDING, RISING, TOWER, UPRIGHT, ROCKET, …

unclear (N = 11) is a rest category including words like CANE, POLE, CANDLE, SPINE, LADDER, …

Final slide (yay!)

- The present study provided additional support for the *perceptual symbol hypothesis* (Barsalou, 1999; 2008) in the domain of short-term memory
  - Better recall performance when words are encoded in line with their (horizontal or vertical) spatial associations
- Numerous studies have shown that linguistic concepts activate spatial representations and related oculomotor responses
- Here, we have found evidence *in the other direction*, from perceptual representations of words (horizontal or vertical reading) to accessing word concepts in memory.