

Semantic Theory — June 26th, 2018

A background image of a human brain in grayscale. Overlaid on the brain are several mathematical formulas and code snippets, suggesting a scientific or computational theme. The formulas include derivatives, loops, and optimization functions related to neural networks or machine learning. In the center, the title "Neural Semantics" is displayed in a large, bold, dark blue font. Below it, the subtitle "Bridging formal and probabilistic approaches to meaning" is written in a smaller, dark blue font.

**Neural Semantics**

Bridging formal and probabilistic approaches to meaning

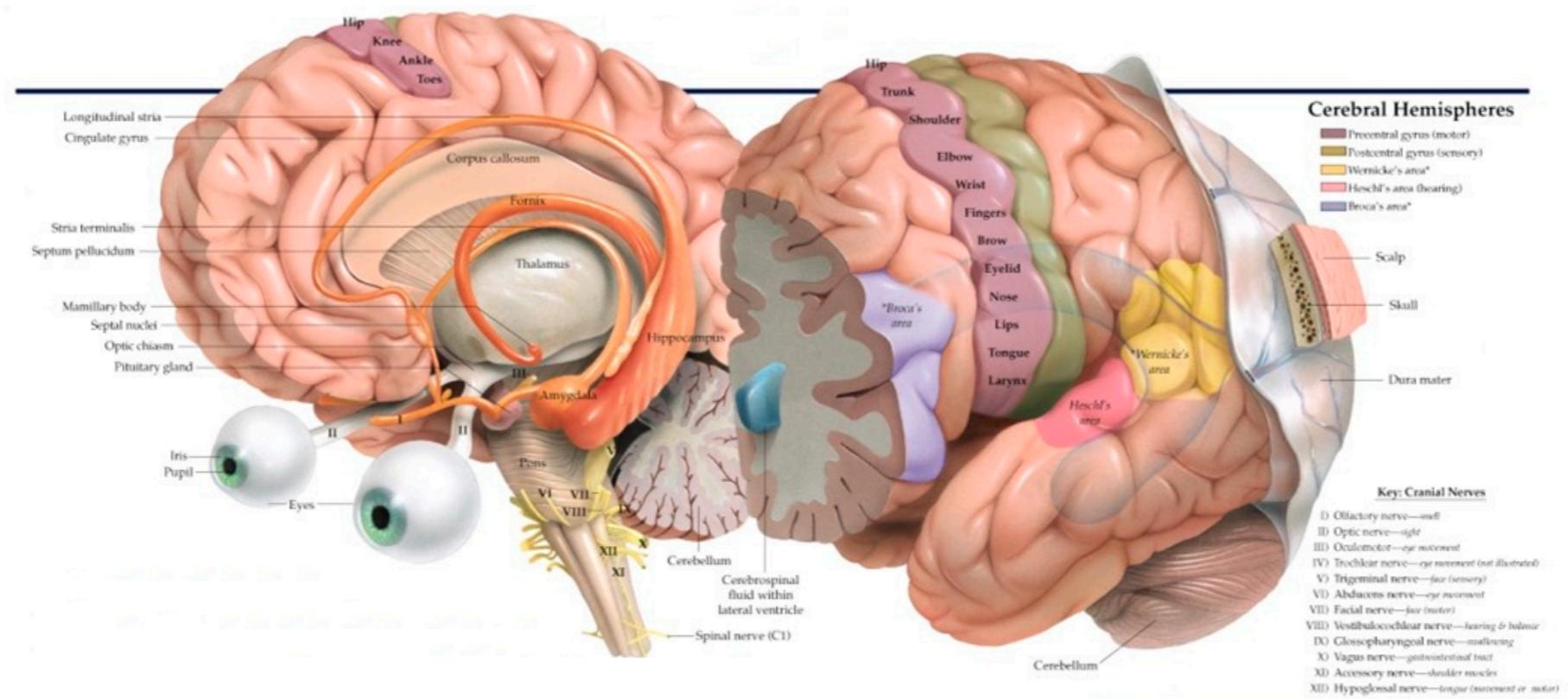
Harm Brouwer

```
1 foreach  $w_{ij}$  do
2    $\Phi \leftarrow \frac{\mu}{1-\mu}$ 
3   foreach  $w_{ij}$  do
4     if  $\frac{\partial E}{\partial w_{ij}}(t) > 0$  then
5        $w_{ij}(t+1) \leftarrow \min(\eta^+ \Gamma_{ij}(t-1), \Gamma_{max})$ 
6     else if  $\frac{\partial E}{\partial w_{ij}}(t) < 0$  then
7        $w_{ij}(t+1) \leftarrow \max(\eta^- \Gamma_{ij}(t-1), \Gamma_{min})$ 
8     else
9        $w_{ij}(t+1) \leftarrow w_{ij}(t)$ 
10    end
11  end
12   $\Gamma_{ij}(t+1) \leftarrow \Gamma_{ij}(t) + \Delta \Gamma_{ij}(t)$ 
13   $\Delta \Gamma_{ij}(t) = \text{sign}(\frac{\partial E}{\partial w_{ij}}(t)) \Gamma_{ij}(t)$ 
14   $\Delta w_{ij}(t) = \Delta \Gamma_{ij}(t) + \Delta w_{ij}(t-1)$ 
15   $w_{ij}(t+1) \leftarrow w_{ij}(t) + \Delta w_{ij}(t)$ 
16 end
```



UNIVERSITÄT  
DES  
SAARLANDES

# The Greatest Semanticist of them all ...



- > Our language comprehension system is highly effective and accurate at attributing meaning to unfolding linguistic signal (~word-by-word)
- >> This system's representations and computational principles are implemented in the neural hardware of the brain
- >>> We should understand meaning construction and representation in terms of "brain-style computation" by identifying a **neural semantics**

# Neural Semantics — Requirements

**Neural Plausibility:** assumed representations and computational principles should be implementable at the neural level [cf. Rumelhart, 1989]

**Expressivity:** representations should capture necessary dimensions of meaning, such as negation, quantification, and modality [cf. Frege, 1892]

**Compositionality:** the meaning of complex expressions should be derivable from the meaning of its parts [cf. Partee, 1984]

**Gradedness:** meaning representations are probabilistic, rather than discrete in nature [cf. Spivey, 2008]

**Inferential:** The derivation of utterance meaning entails (direct) inferences that go beyond literal propositional content [cf. Johnson-Laird, 1983]

**Incrementality:** As natural language unfolds over time, representations should allow for incremental construction [cf. Tanenhaus et al., 1995]

# Today's lecture

I. A Framework for Neural Semantics

II. A Neural Model of Language Comprehension

III. Neural Semantics — A fertile approach?

# I. A Framework for Neural Semantics

```
1 foreach  $w_{ij}$  do
2    $\Phi \leftarrow \frac{\mu}{1-\mu}$ 
3   foreach  $w_{ij}$  do
4     if  $\frac{\partial E}{\partial w_{ij}y_j}(t) > 0$  then
5       if  $\Delta w_{ij}(t-1) > \Gamma_{ij}(t)$ 
6          $\Gamma_{ij}(t) \leftarrow \min(\eta^+ \Gamma_{ij}(t-1), \Gamma_{max})$ 
7        $\Delta w_{ij}(t) \leftarrow \text{sign}(\frac{\partial E}{\partial w_{ij}}(t)) \Gamma_{ij}(t)$ 
8        $w_{ij}(t+1) \leftarrow w_{ij}(t) + \Delta w_{ij}(t)$ 
9     end
10    else if  $\frac{\partial E}{\partial w_{ij}}(t-1) \frac{\partial E}{\partial w_{ij}}(t) < 0$  then
11       $\Gamma_{ij}(t) \leftarrow \max(\eta^- \Gamma_{ij}(t-1), \Gamma_{min})$ 
12       $\Delta w_{ij}(t) \leftarrow \text{sign}(\frac{\partial E}{\partial w_{ij}}(t)) \Gamma_{ij}(t)$ 
13    end
14  end
15   $\Delta w_{ij}(t) \leftarrow \frac{\partial E}{\partial w_{ij}}(t) + d \cdot w_{ij}(t)$ 
16   $w_{ij}(t+1) \leftarrow w_{ij}(t) + \Delta w_{ij}(t)$ 
17 end
```

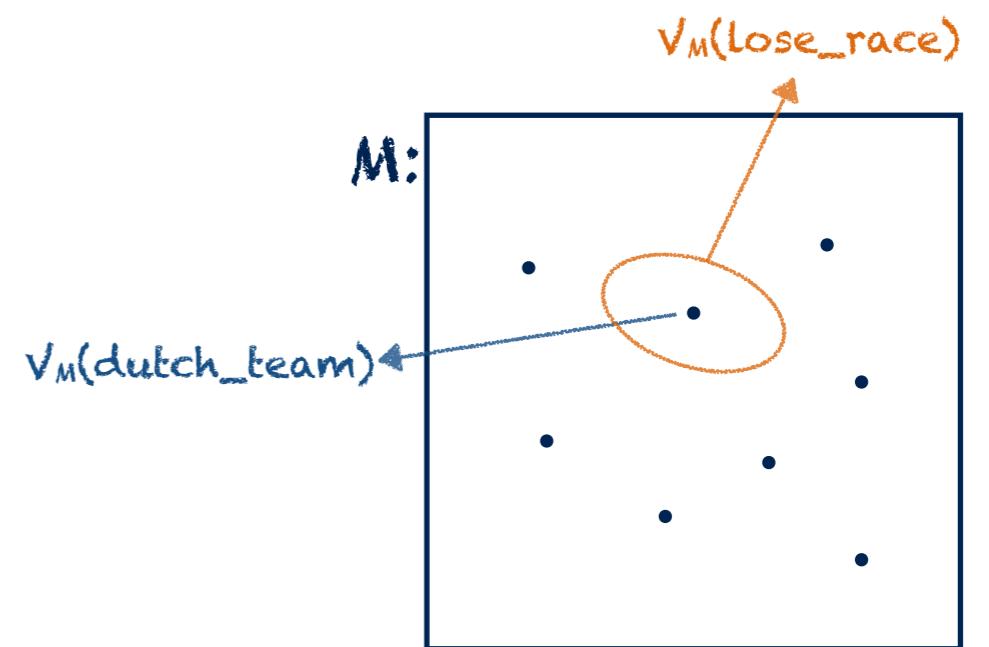
# Meaning in formal semantics

Meaning is defined as **truth-conditions** over logical models

$\llbracket \text{The Dutch team lost the race} \rrbracket^{M,g} = 1$

iff  $\llbracket \text{lose\_race(dutch\_team)} \rrbracket^{M,g} = 1$

iff  $V_M(\text{dutch\_team}) \in V_M(\text{lose\_race})$



Intuition: Meaning is defined by the **situations** that satisfy an expression

Cognitively, truth-conditions derive from **experience** with the world

Idea: Model experience as **observations** of states-of-affairs in the world

# Experiences as cues to meaning

Observations of SoAs can be captured by the set of logical models  $\mathcal{M}$



$M_1$ : enter(beth,restaurant), order(beth,dinner), eat(beth,dinner),  
enter(dave, cinema), order(dave,cola), ....



$M_2$ : enter(beth,cinema), order(beth,popcorn), order(dave,cola),  
pay(dave), enter(thom, restaurant), leave(thom), ....



$M_3$ : enter(beth,cinema), order(beth,popcorn), eat(beth,popcorn),  
order(dave,cola), pay(dave), order(thom, cola), ....

....



$M_{150}$ : enter(beth,cinema), leave(beth), enter(dave,restaurant),  
pay(dave), enter(thom, cinema), order(thom, water), ....

> Each  $M \in \mathcal{M}$  provides a **cue** toward the underlying truth-conditions

# Experiences as cues to meaning

Observations of SoAs can be captured by the set of logical models  $\mathcal{M}$



$M_1$ : enter(beth,restaurant), order(beth,dinner), **eat(beth,dinner)**,  
enter(dave, cinema), order(dave,cola), ....



$M_2$ : enter(beth,cinema), order(beth,popcorn), order(dave,cola),  
pay(dave), enter(thom, restaurant), leave(thom), ....



$M_3$ : enter(beth,cinema), order(beth,popcorn), **eat(beth,popcorn)**,  
order(dave,cola), pay(dave), order(thom, cola), ....

....



$M_{150}$ : enter(beth,cinema), leave(beth), enter(dave,restaurant),  
pay(dave), enter(thom, cinema), order(thom, water), ....

> Each  $M \in \mathcal{M}$  provides a **cue** toward the underlying truth-conditions

# Experiences as cues to meaning

Observations of SoAs can be captured by the set of logical models  $\mathcal{M}$



- > Each  $M \in \mathcal{M}$  provides a **cue** toward the underlying truth-conditions
- > Together, all  $M \in \mathcal{M}$  reflect the world **truth-conditionally** and **probabilistically**

# Neural Semantics — Meaning

Formally, we can define the model space  $S_{M \times P}$  for a set of models  $\mathcal{M}$  and propositions  $\mathcal{P}$

**Rows** represent models that reflect observations of ‘states-of-affairs’ in the world

**Columns** represent meaning vectors of individual propositions

$$\mathbf{v}_i(p) = 1 \text{ iff } \llbracket p \rrbracket^{M_i} = 1$$

Meaning is defined in terms of **co-occurrence**

> Propositions with related meanings will be true in many of the same models

	$p_1 = \text{enter(beth,restaurant)}$	$p_2 = \text{ask\_menu(beth)}$	$p_3 = \text{order(beth,cola)}$	$\dots$	$p_n = \text{leave(dave)}$
$M_1$	1	0	0		1
$M_2$	0	1	1		0
$M_3$	1	1	1		0
$M_4$	0	0	0		0
$M_5$	0	0	1		1
$\dots$					
$M_m$	0	1	1	$\dots$	1

$\mathbf{v}(p_2)$

# Neural Semantics — Compositionality

Meaning vectors of **atomic situations** (propositions) are columns in  $S_{M \times P}$

Meaning vectors of **complex situations** are derived compositionally:

$$\mathbf{v}_i(\neg p) = 1 \text{ iff } \mathbf{v}_i(p) = 0 \text{ for } 1 \leq i \leq m$$

$$\mathbf{v}_i(p \wedge q) = 1 \text{ iff } \mathbf{v}_i(p) = 1 \text{ and } \mathbf{v}_i(q) = 1, \text{ for } 1 \leq i \leq m$$

> which gives functional completeness [e.g.,  $\mathbf{v}_i(p \vee q) = \mathbf{v}_i(\neg(\neg p \wedge \neg q))$ ]

Quantification is defined over the combined universe of  $\mathcal{M}$ :  $U_{\mathcal{M}} = \{u_1, \dots, u_k\}$

$$\mathbf{v}_i(\forall x \phi) = 1 \text{ iff } \mathbf{v}_i(\phi[x \setminus u_1] \wedge \dots \wedge \phi[x \setminus u_k]) = 1 \text{ for } 1 \leq i \leq m$$

$$\mathbf{v}_i(\exists x \phi) = 1 \text{ iff } \mathbf{v}_i(\phi[x \setminus u_1] \vee \dots \vee \phi[x \setminus u_k]) = 1 \text{ for } 1 \leq i \leq m$$

# Neural Semantics — Probability

Meaning vectors inherently encode (co-)occurrence probabilities of situations—which may be atomic (~propositional) or complex (~conjunctive)

> The prior probability of situation a:

$$P(a) = \sum_i (\mathbf{v}_i(a)) / m$$

> The conjunctive probability of situations a and b:

$$P(a \wedge b) = \sum_i (\mathbf{v}_i(a)\mathbf{v}_i(b)) / m$$

> The conditional probability of situation a given b:

$$P(a | b) = P(a \wedge b) / P(b)$$

	$p_1 = \text{enter}(\text{beth}, \text{restaurant})$				
		$p_2 = \text{ask\_menu}(\text{beth})$			
			$p_3 = \text{order}(\text{beth}, \text{cola})$		
				$\dots$	
					$p_n = \text{leave}(\text{dave})$
$M_1$	1	0	0		
$M_2$	0	1	1		
$M_3$	1	1	1		
$M_4$	0	0	0		
$M_5$	0	0	1		
$\dots$					
$M_m$	0	1	1		
				$\dots$	

# Neural Semantics — Inference

How much is situation **a** understood to be the case from situation **b**?

$$\text{comprehension}(a,b) = \begin{cases} [P(a | b) - P(a)] / [1 - P(a)] & \text{if } P(a | b) > P(a) \\ [P(a | b) - P(a)] / P(a) & \text{otherwise} \end{cases}$$

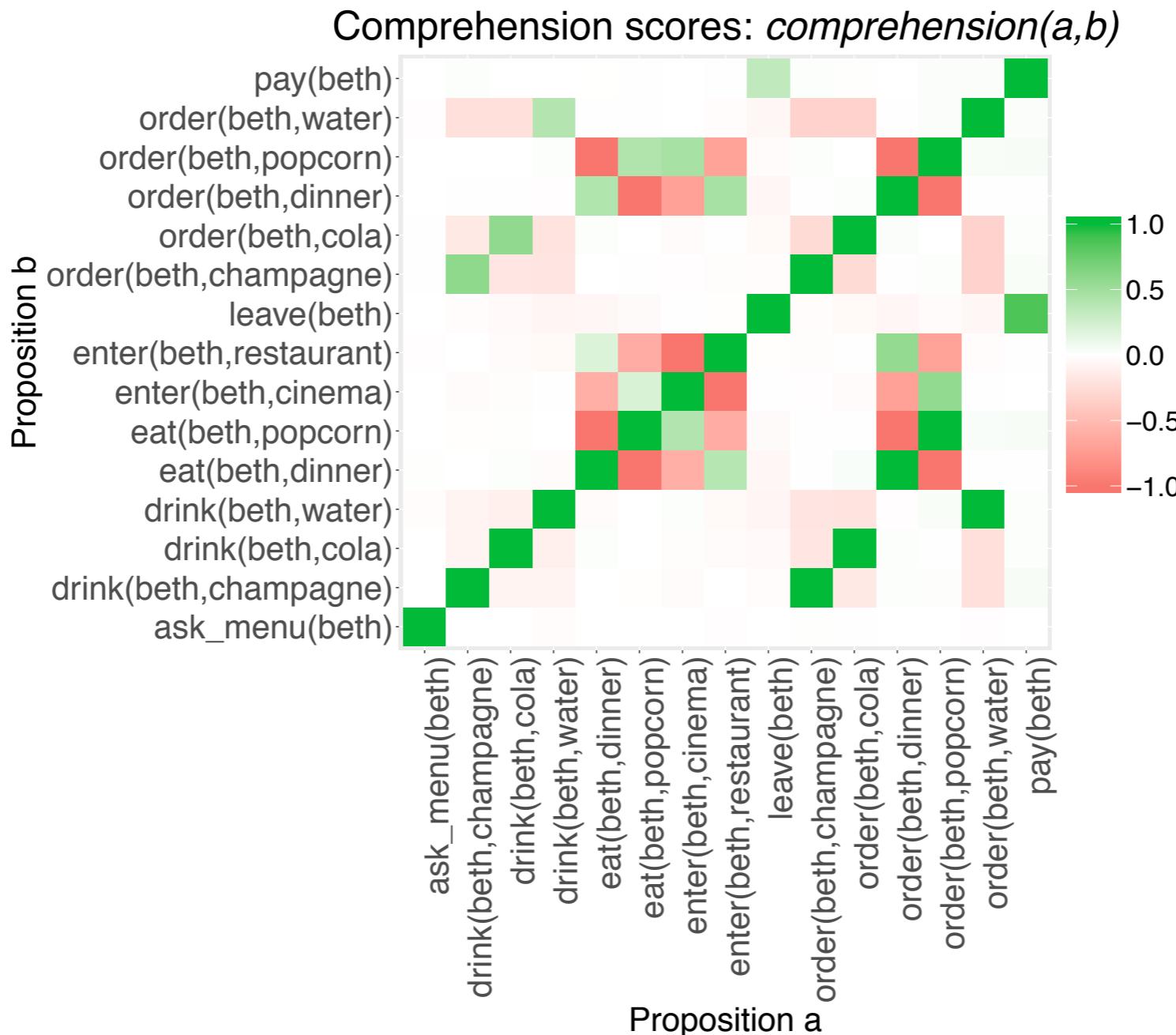
$P(a | b) > P(a)$  knowing **b** *increases* belief in **a**  
→ **a** is inferred to be the case from **b**

$\text{comprehension}(a,b) = 1$ : knowing **b** took away all ‘uncertainty’ in **a** →  $b \vDash a$

$P(a | b) \leq P(a)$  knowing **b** *decreases* belief in **a**  
→ **a** is inferred **not** to be the case from **b**

$\text{comprehension}(a,b) = -1$ : knowing **b** took away all ‘certainty’ in **a** →  $b \vDash \neg a$

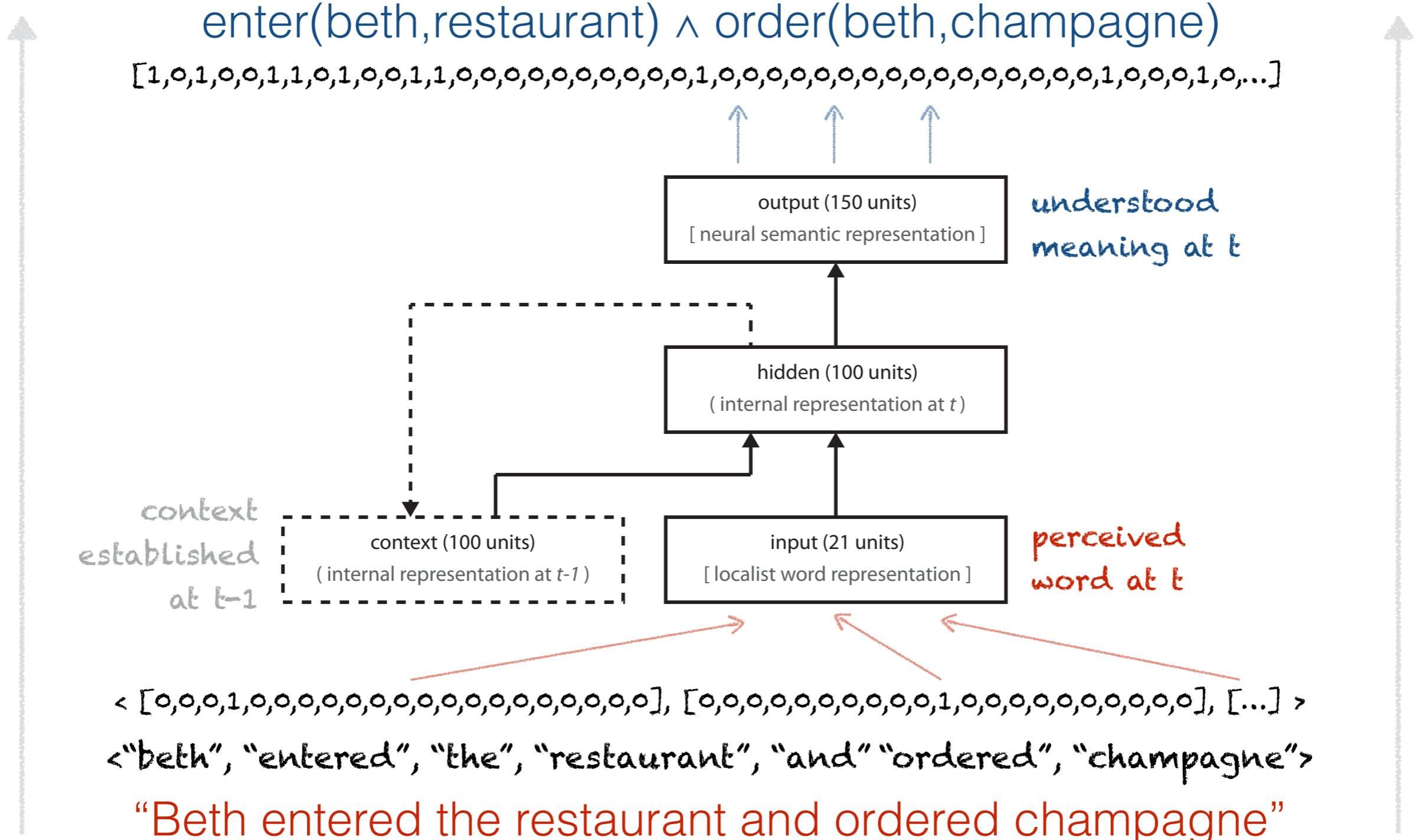
# World knowledge inferencing in $S_{M \times P}$



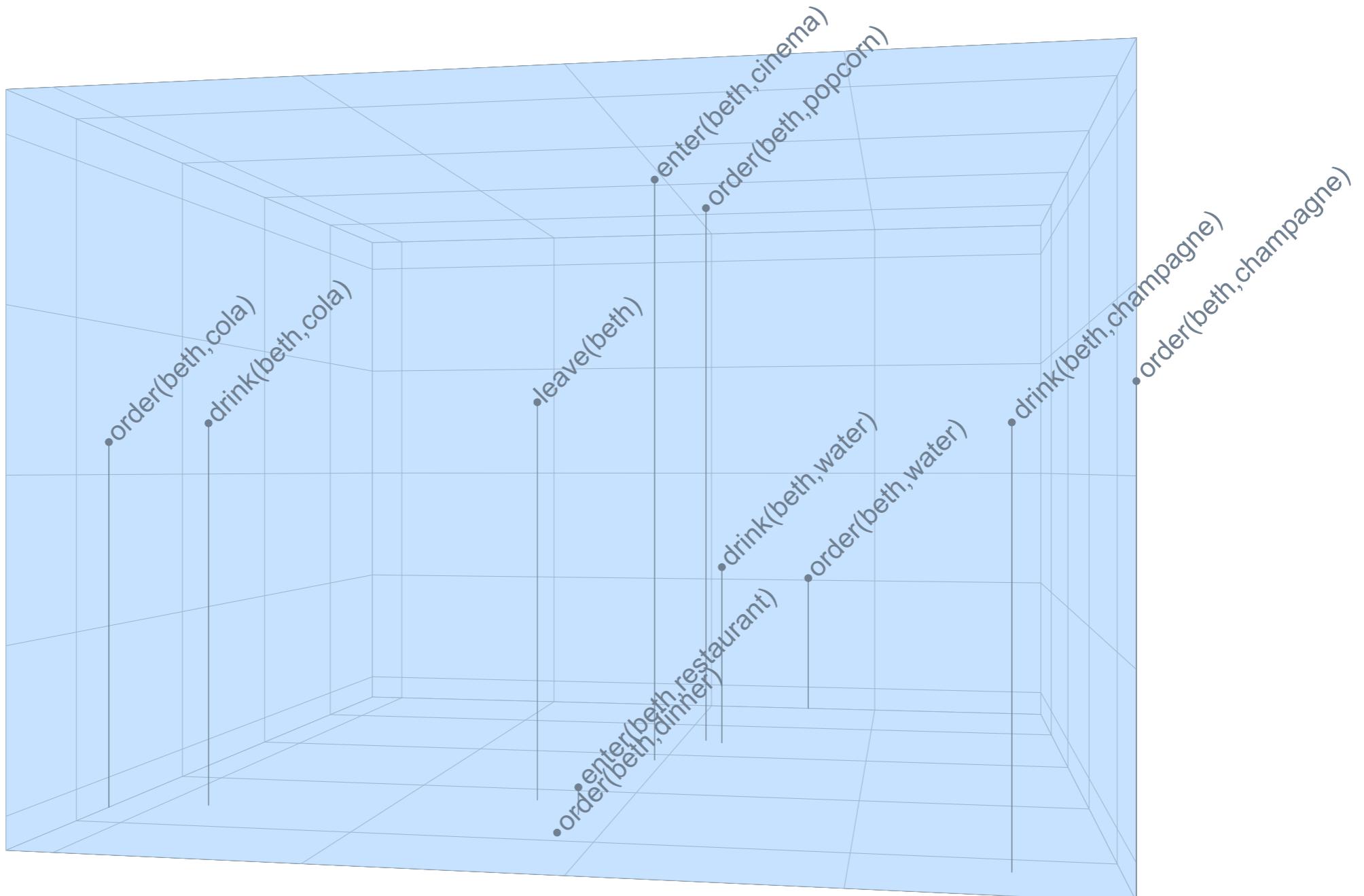
Next: Let's employ  $S_{M \times P}$  in an incremental comprehension model

## II. A Neural Model of Language Comprehension

# A Neural Model of Language Comprehension

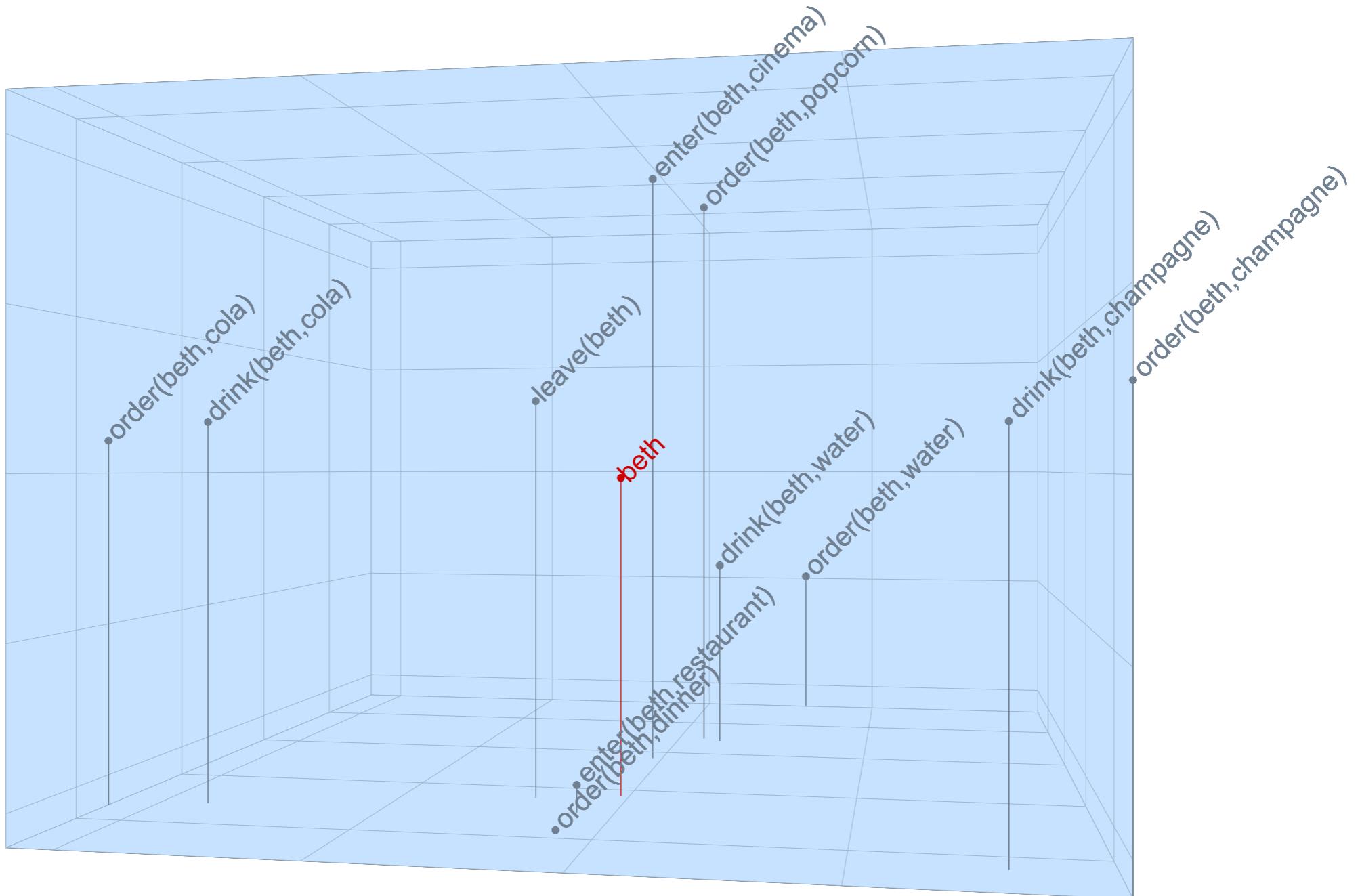


# Comprehension is meaning-space navigation



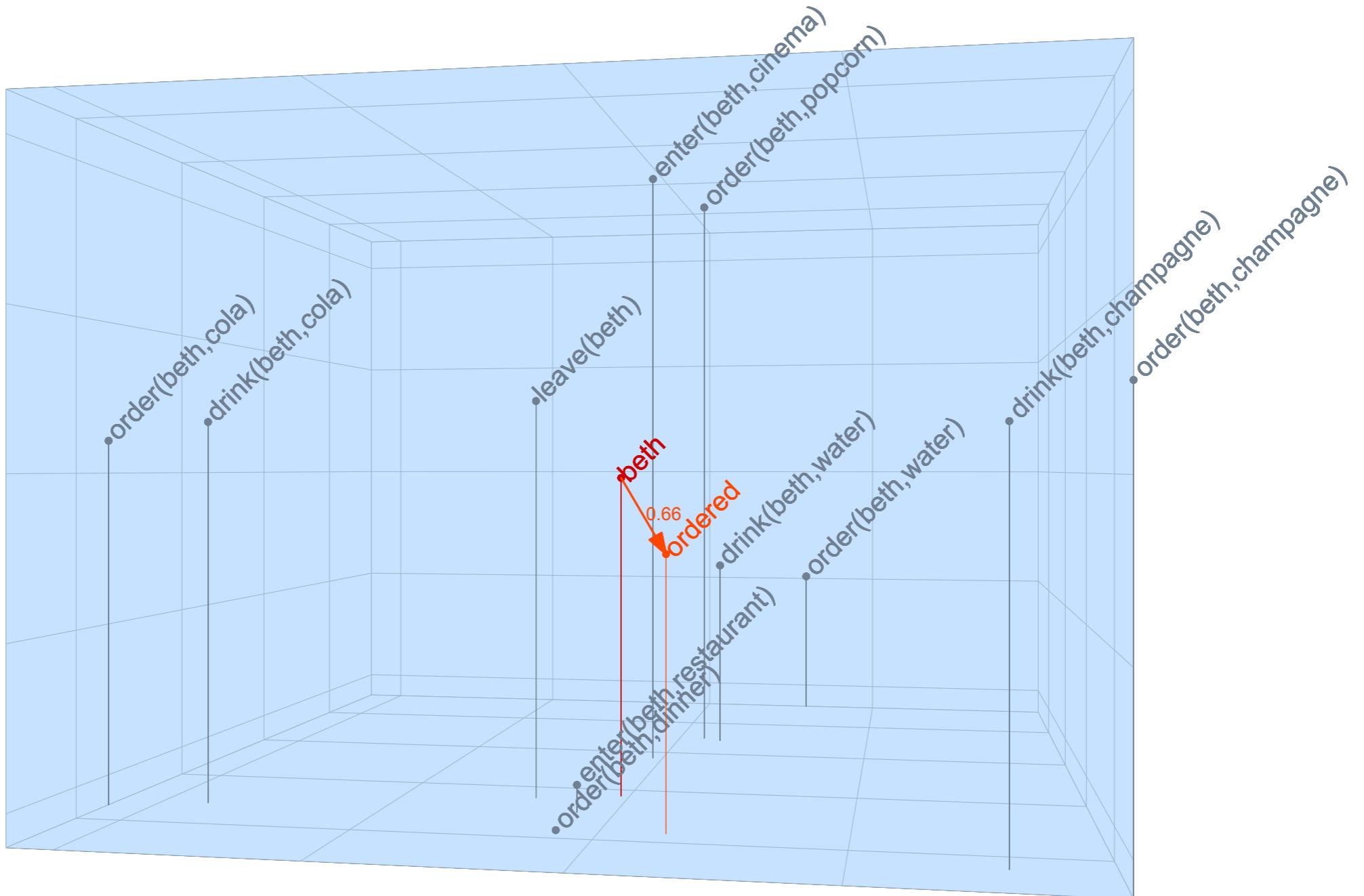
Multi-dimensional scaling:  $150\text{D} \mapsto 3\text{D}$

# Comprehension is meaning-space navigation



[“beth”]

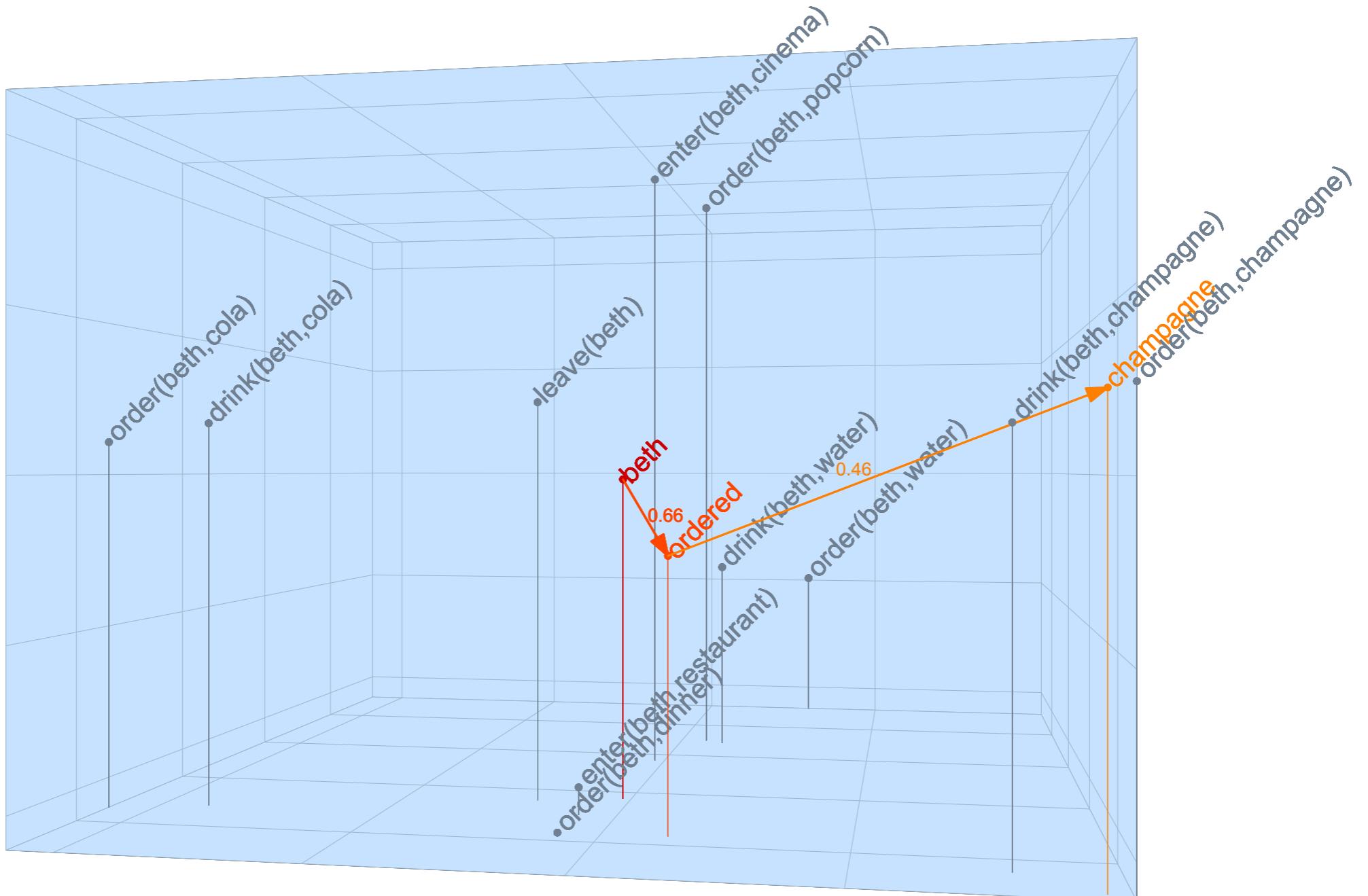
# Comprehension is meaning-space navigation



["beth", "ordered"]

(scalars  $\propto$  distance)

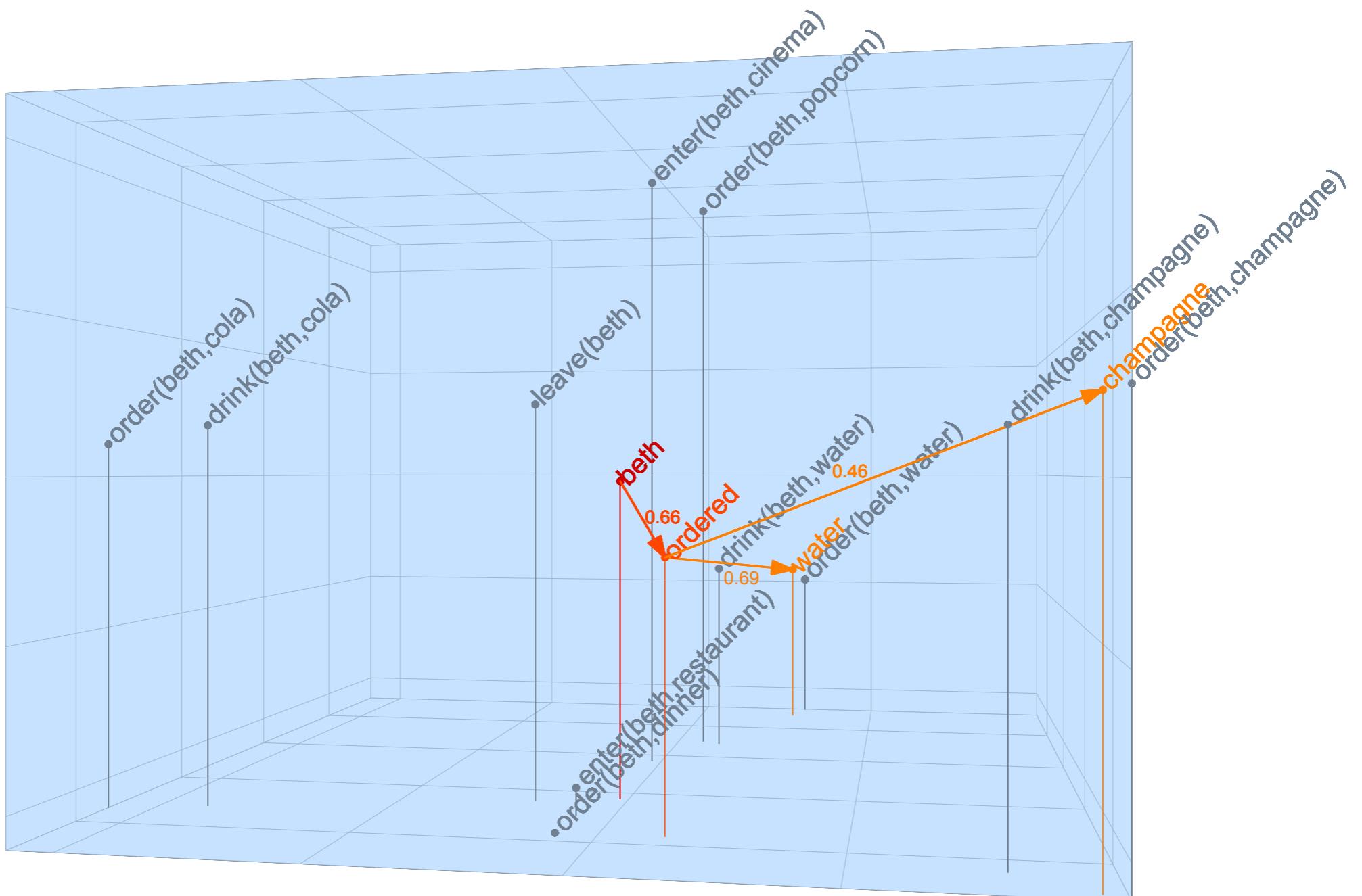
# Comprehension is meaning-space navigation



["beth", "ordered", "champagne"]

(scalars  $\propto$  distance)

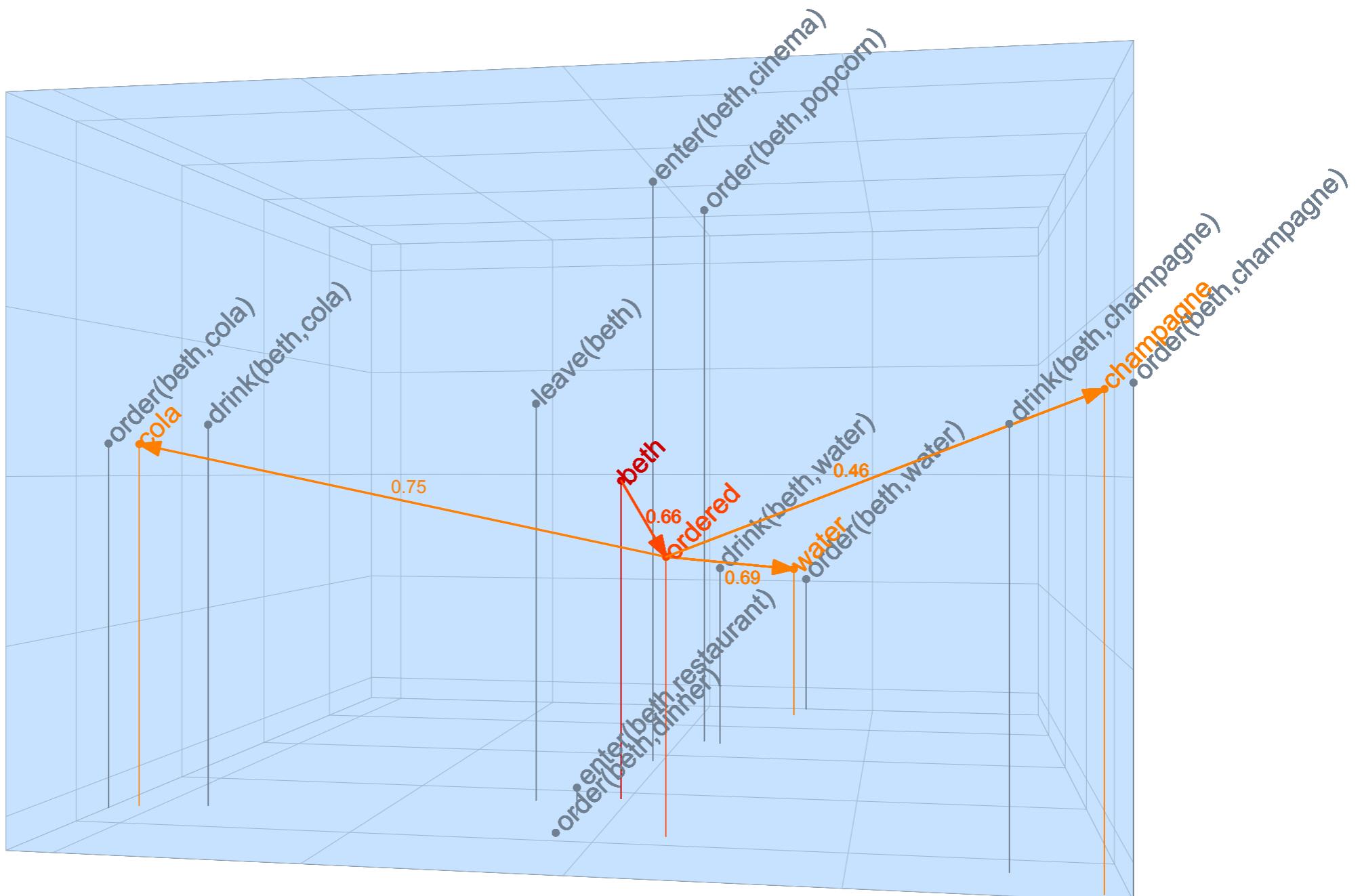
# Comprehension is meaning-space navigation



["beth", "ordered", "water"]

(scalars  $\propto$  distance)

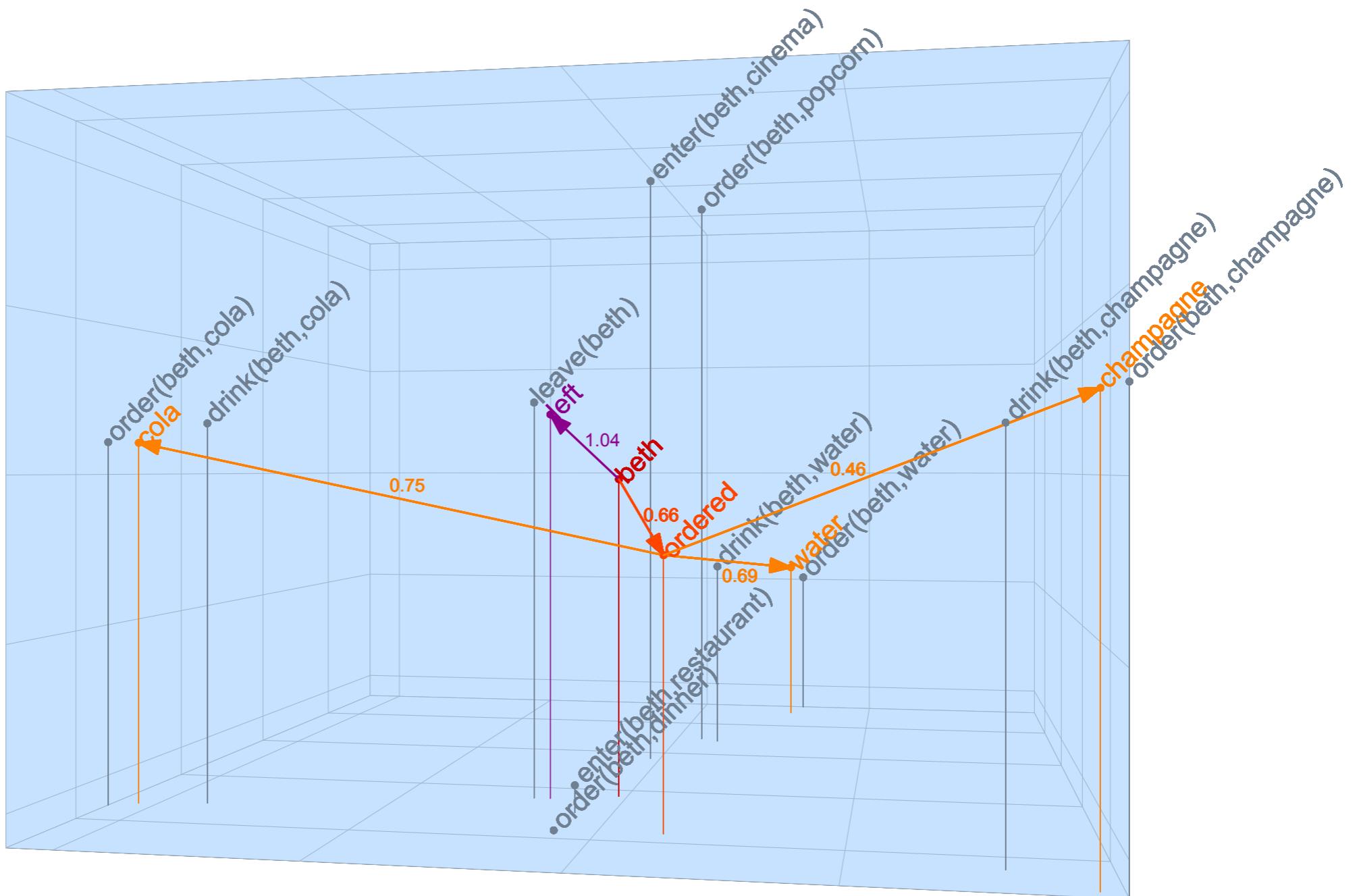
# Comprehension is meaning-space navigation



["beth", "ordered", "cola"]

(scalars  $\propto$  distance)

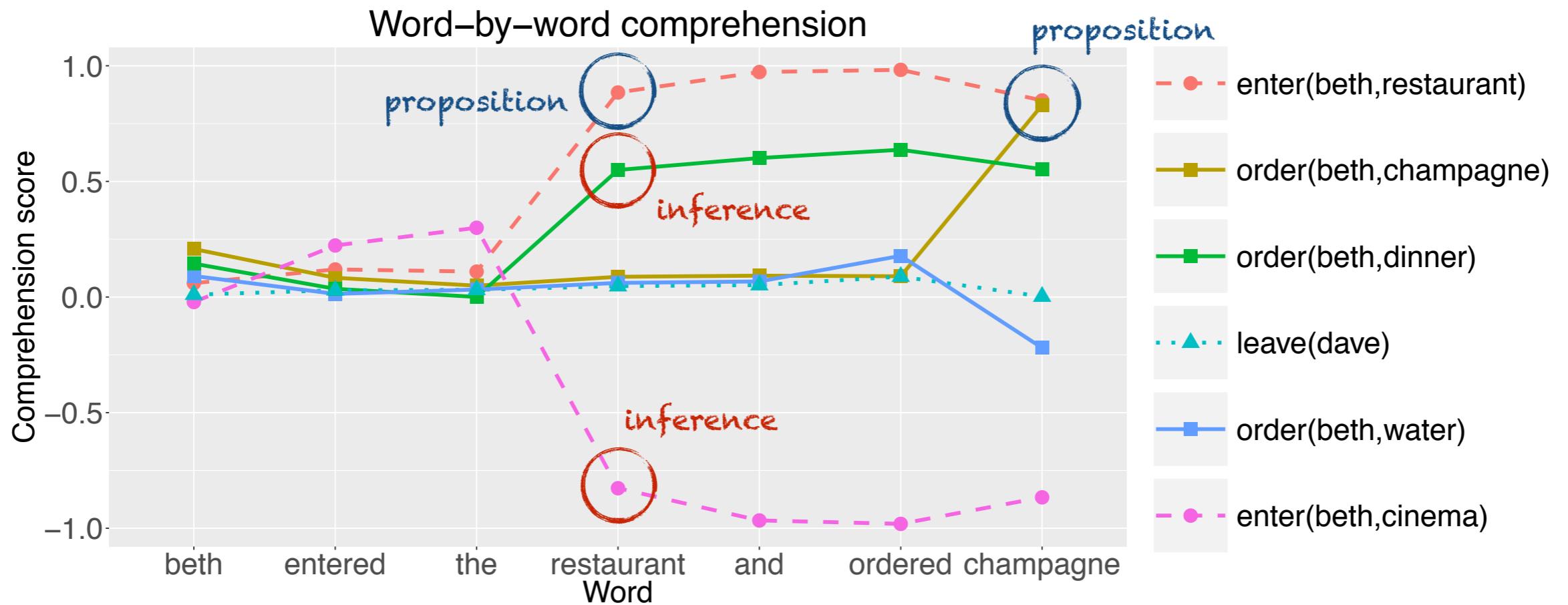
# Comprehension is meaning-space navigation



[“beth”, “left”]

(scalars  $\propto$  distance)

# What does the model ‘understand’?



> Points in meaning-space capture meaning beyond literal propositional content; i.e., model engages in direct knowledge-driven inferencing

# What does the model ‘understand’?

	beth	entered	the	restaurant	and	ordered	champagne	
	+0.07039	+0.00133	+0.07172	-0.01075	+0.06097	+0.29321	+0.35418	-0.0641
enter(beth, cinema)	-0.02148	+0.24441	+0.22293	+0.07659	+0.29951	-1.12	-0.125	-0.96616
enter(beth, restaurant)	+0.05997	+0.05947	+0.11943	-0.00959	+0.10984	-0.050	-0.050	-0.01477
enter(dave, cinema)	-0.02625	-0.01673	-0.04299	-0.01351	-0.050	-0.050	-0.050	+0.97301
enter(dave, restaurant)	-0.04534	+0.03629	-0.00905	-0.00655	-0.050	-0.050	-0.050	+0.00947
enter(thom, cinema)	-0.00525	-0.03060	-0.03585	-0.02386	-0.050	-0.050	-0.050	+0.00439
enter(thom, restaurant)	-0.03036	+0.03124	+0.00087	-0.05578	-0.050	-0.050	-0.050	-0.0971
ask_menu(beth)	+0.05057	+0.11973	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.03386
ask_menu(dave)	-0.04209	+0.0257	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.09729
ask_menu(thom)	+0.01846	+0.0302	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.06977
order(beth, dinner)	+0.14434	-0.1089	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.16706
order(beth, popcorn)	-0.02546	+0.1089	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00894
order(dave, dinner)	+0.01575	-0.02546	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.13153
order(dave, popcorn)	-0.02546	+0.1089	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.10589
order(thom, dinner)	-0.01575	-0.02546	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.17818
order(thom, popcorn)	-0.02546	+0.1089	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00031
order(beth, water)	-0.01575	-0.02546	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.17724
order(beth, cola)	-0.01575	-0.02546	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.0501
order(beth, champagne)	-0.01575	-0.02546	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.0440
eat(beth, dinner)	+0.05633	-0.0295	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
eat(beth, popcorn)	+0.03360	-0.0079	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
eat(dave, dinner)	-0.00731	+0.0491	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
eat(dave, popcorn)	-0.00597	-0.0151	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
eat(thom, dinner)	-0.12249	-0.0724	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
eat(thom, popcorn)	+0.01243	-0.0950	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(beth, water)	+0.15517	-0.1344	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(beth, cola)	+0.12933	-0.0556	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(beth, champagne)	+0.13052	-0.0761	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(dave, water)	-0.0	-0.0	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(dave, cola)	-0.0	-0.0	-0.00000	-0.00000	-0.050	-0.050	-0.050	-0.00000
drink(dave, champagne)	+0.01509	-0.01130	+0.00379	+0.00610	+0.00989	-0.00919	+0.00069	-0.01031
drink(thom, water)	-0.05061	-0.02465	-0.07525	-0.01961	-0.09486	-0.08131	-0.17618	-0.00217
drink(thom, cola)	+0.00930	-0.01481	-0.00551	+0.00614	+0.00064	-0.11499	-0.11435	-0.02679
drink(thom, champagne)	+0.02685	+0.01670	+0.04355	+0.01531	+0.05886	-0.05493	+0.00393	-0.00401
pay(beth)	+0.15462	+0.05596	+0.21058	+0.02684	+0.23742	+0.04137	+0.27879	-0.00146
pay(dave)	+0.01319	+0.02897	+0.04217	+0.00715	+0.04932	+0.02936	+0.07868	+0.00097
pay(thom)	-0.00686	+0.12125	+0.11439	+0.00955	+0.12394	+0.00018	+0.12412	+0.00689
leave(beth)	+0.05375	+0.06625	+0.12000	-0.00191	+0.11809	+0.02816	+0.14625	+0.00750
leave(dave)	+0.01022	+0.01904	+0.02927	+0.00185	+0.03111	+0.01600	+0.04711	+0.00440
leave(thom)	+0.01520	+0.07592	+0.09112	+0.00861	+0.09974	-0.06078	+0.03896	-0.01819