Probabilistic CKY Algorithm

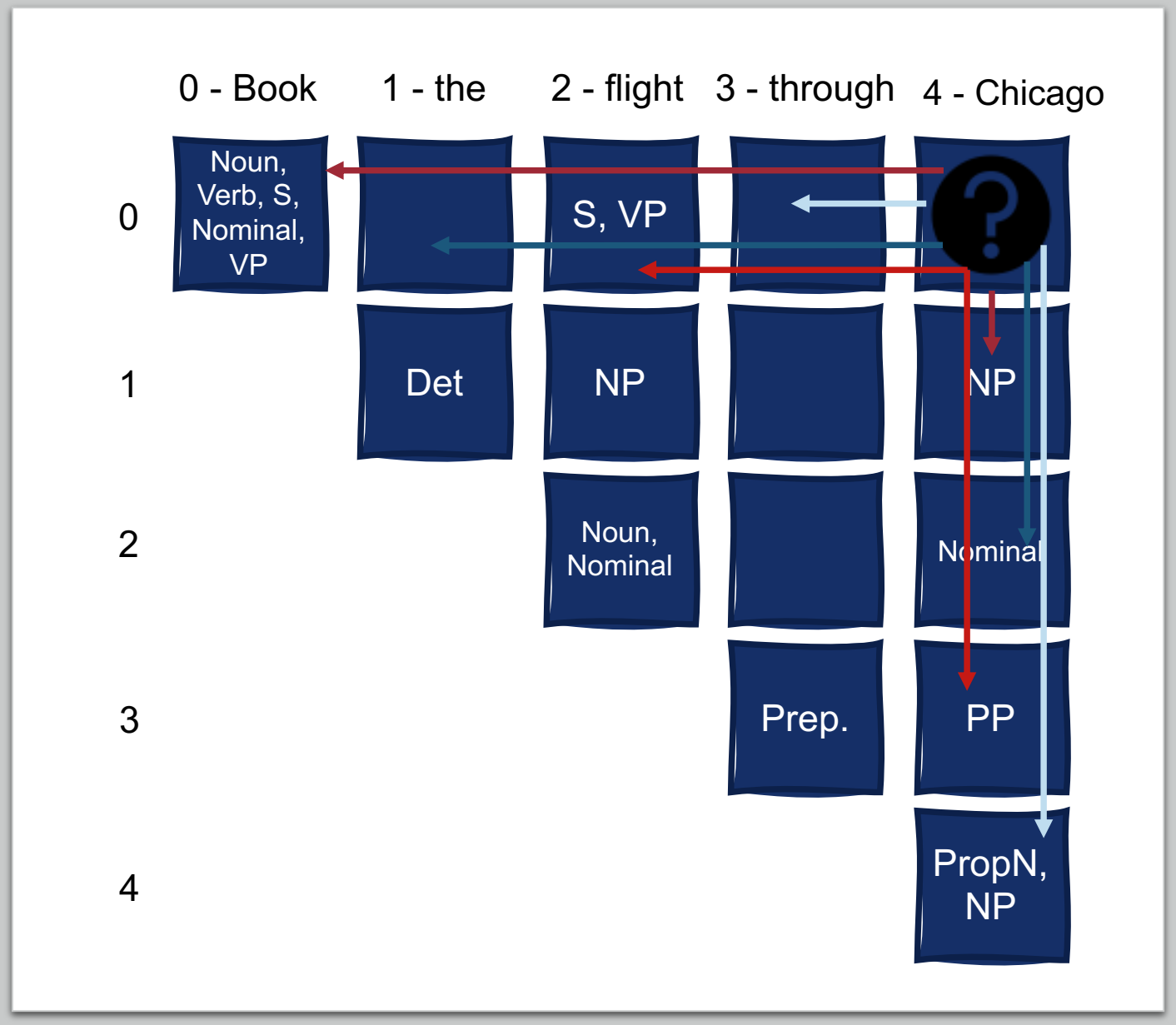
Natalie Parde
UIC CS 421

How do we compute the probability of a parse tree?

Simple solution: Extend the classic parsing algorithms we already have

Probabilistic CKY

- Still assume grammar is in Chomsky Normal Form
  - Right-hand side of production rule expands to two non-terminals or one terminal node
    - $A \rightarrow B\ C$
    - $A \rightarrow w$
- Still work with the upper triangular portion of a matrix



Probabilistic CKY

- Let *n* be the length of an input sentence, and *V* be the number of non-terminals in a grammar
- Consider the constituents *inside* the matrix cells to be part of a third dimension, of maximum length *V*
- Then, each cell $[i, j, A]$ in the $(n+1)\times(n+1)\times V$ matrix corresponds to the probability of constituent *A* spanning positions *i* through *j* of the input

Case Example: Probabilistic CKY

The price includes a facemask

Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02

Case Example: Probabilistic CKY

The price includes a facemask

Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02

Case Example: Probabilistic CKY

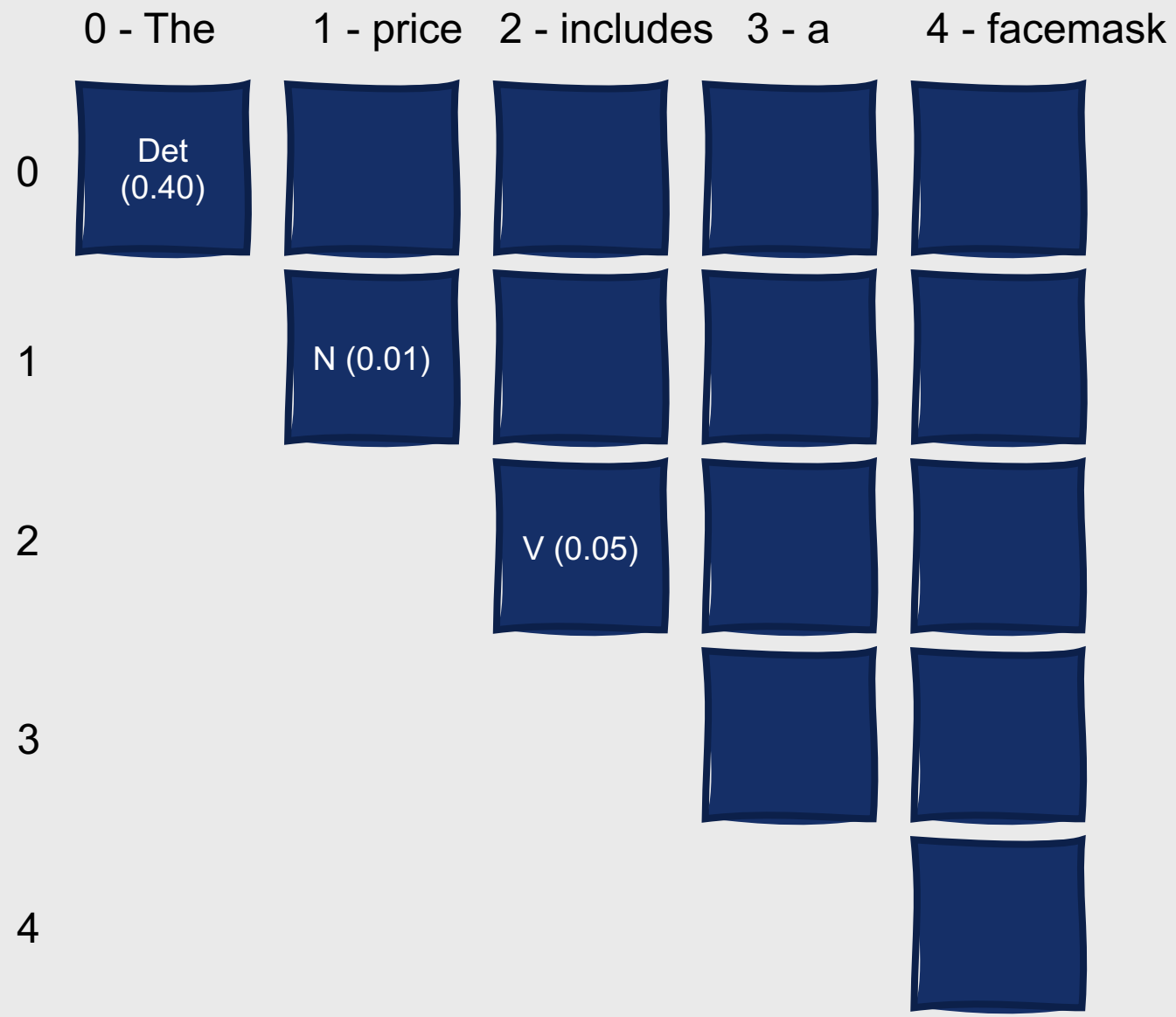
The price includes a facemask

Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02

Case Example: Probabilistic CKY

The price includes a facemask

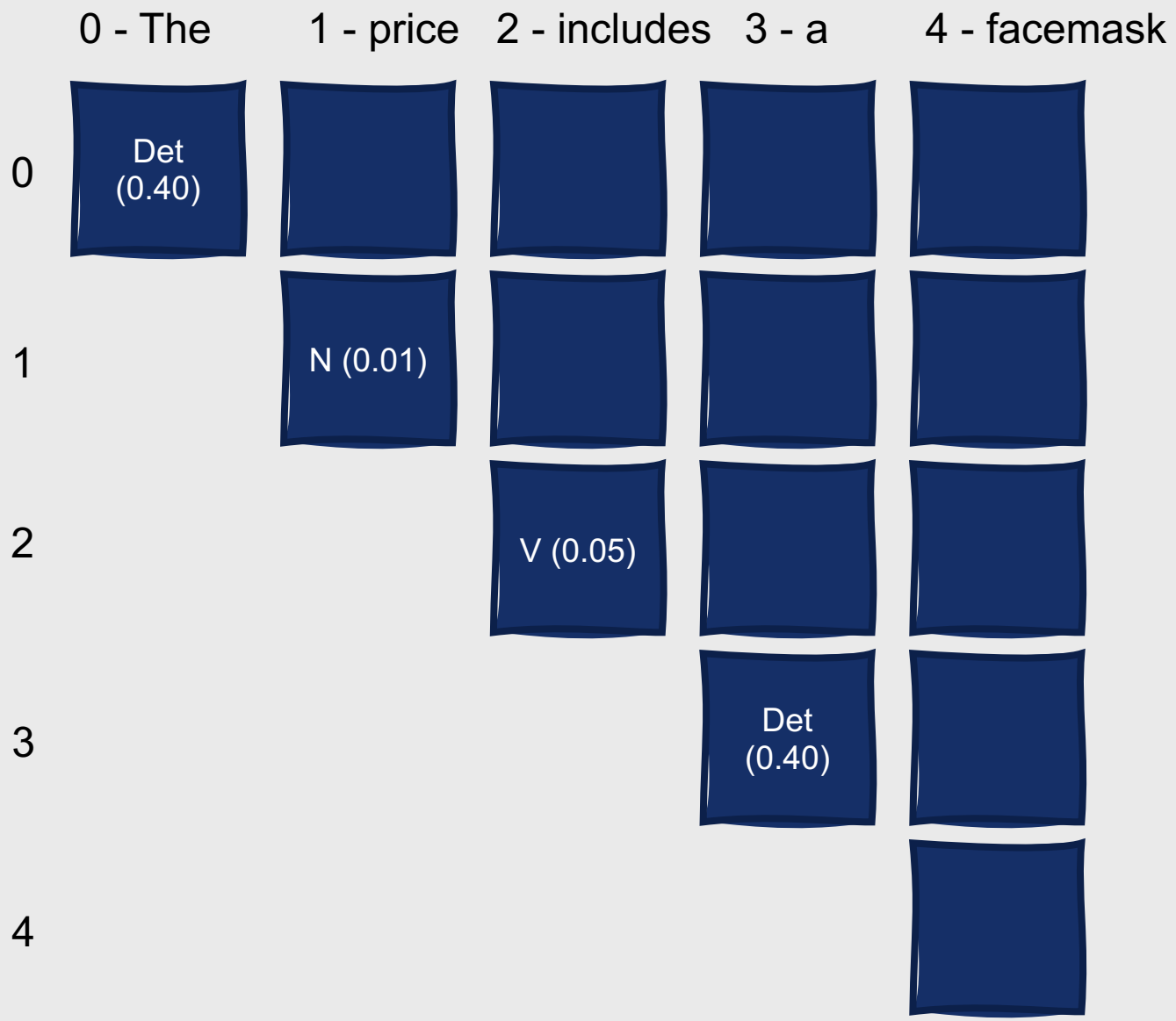
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

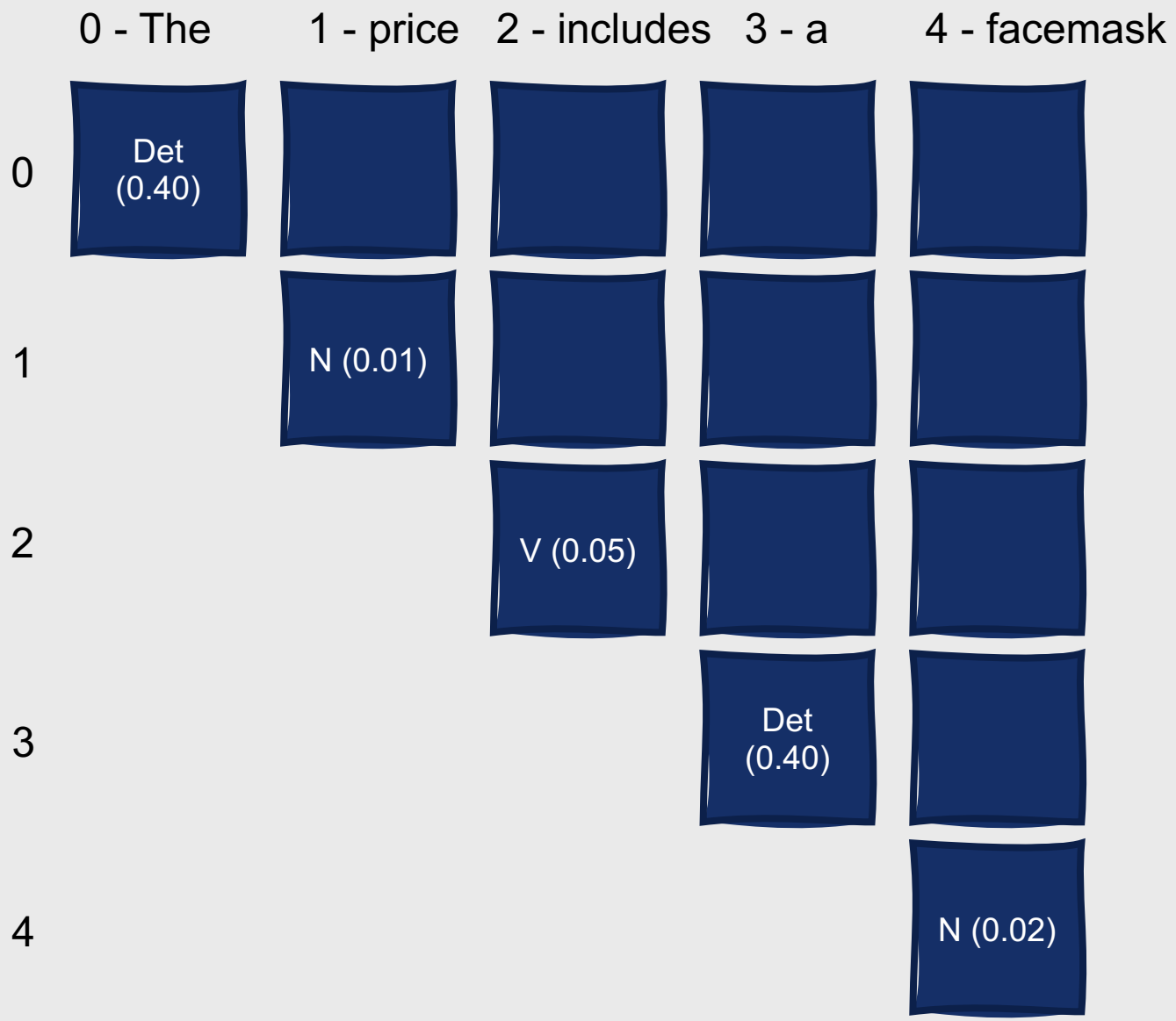
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

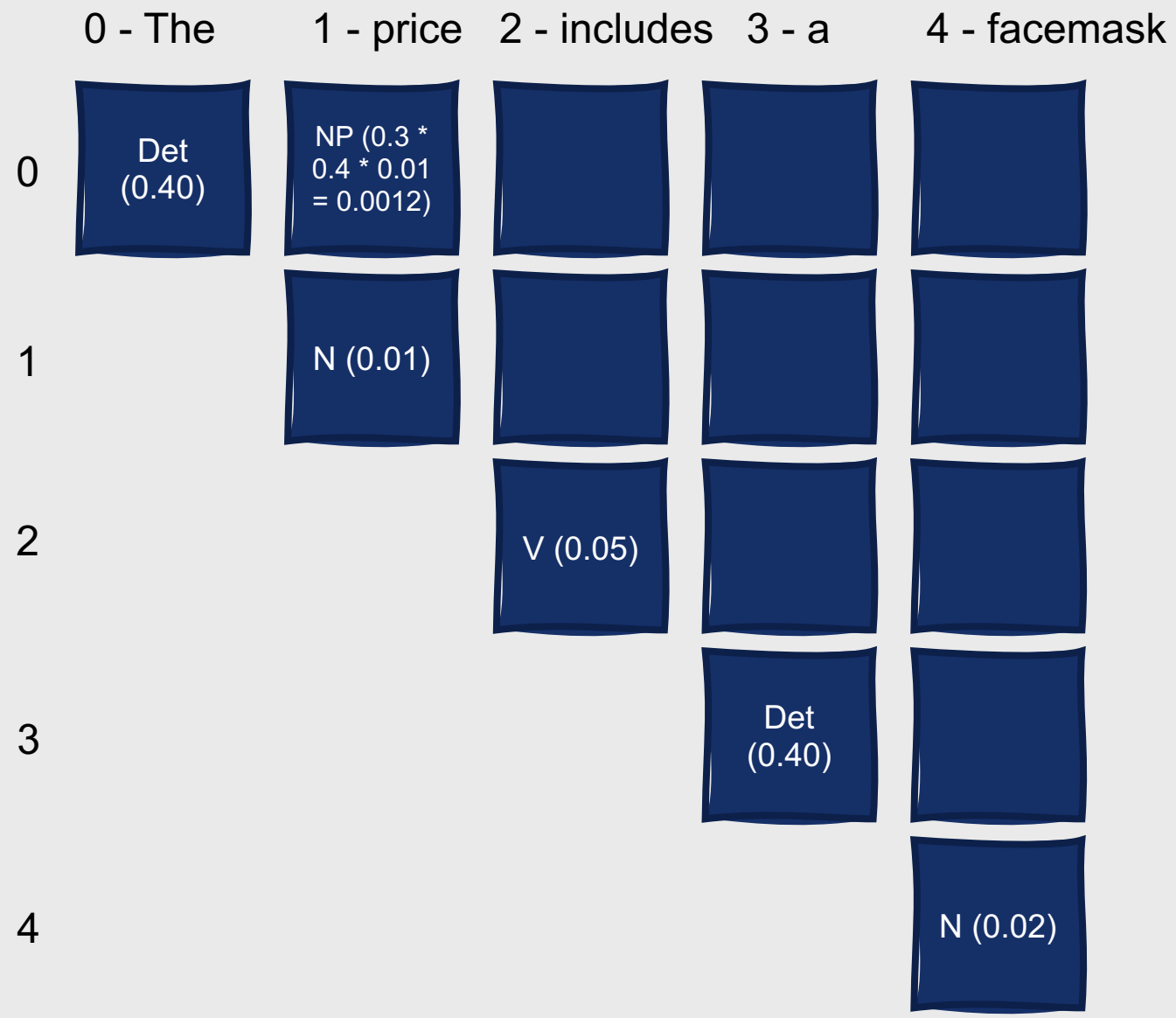
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

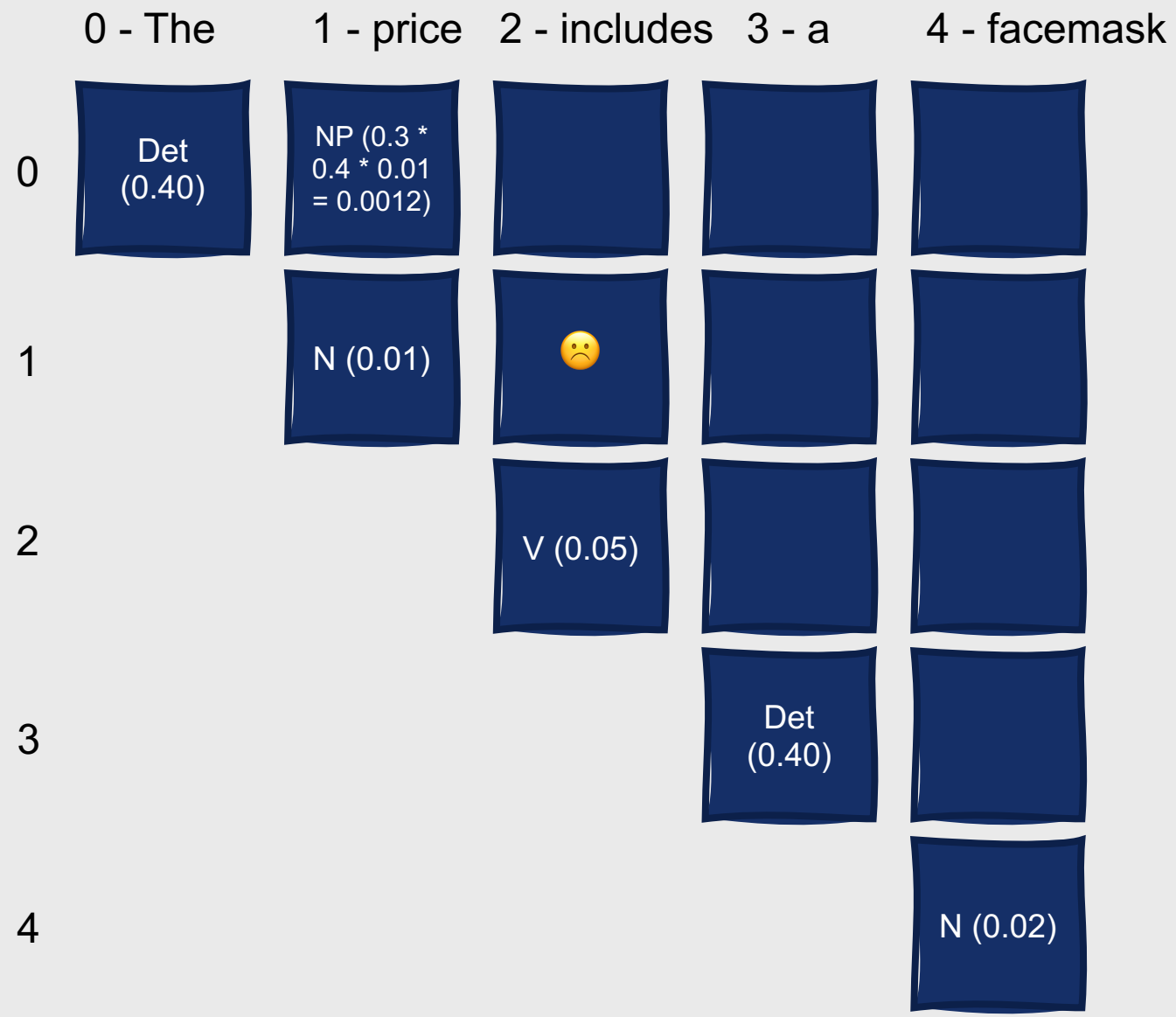
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

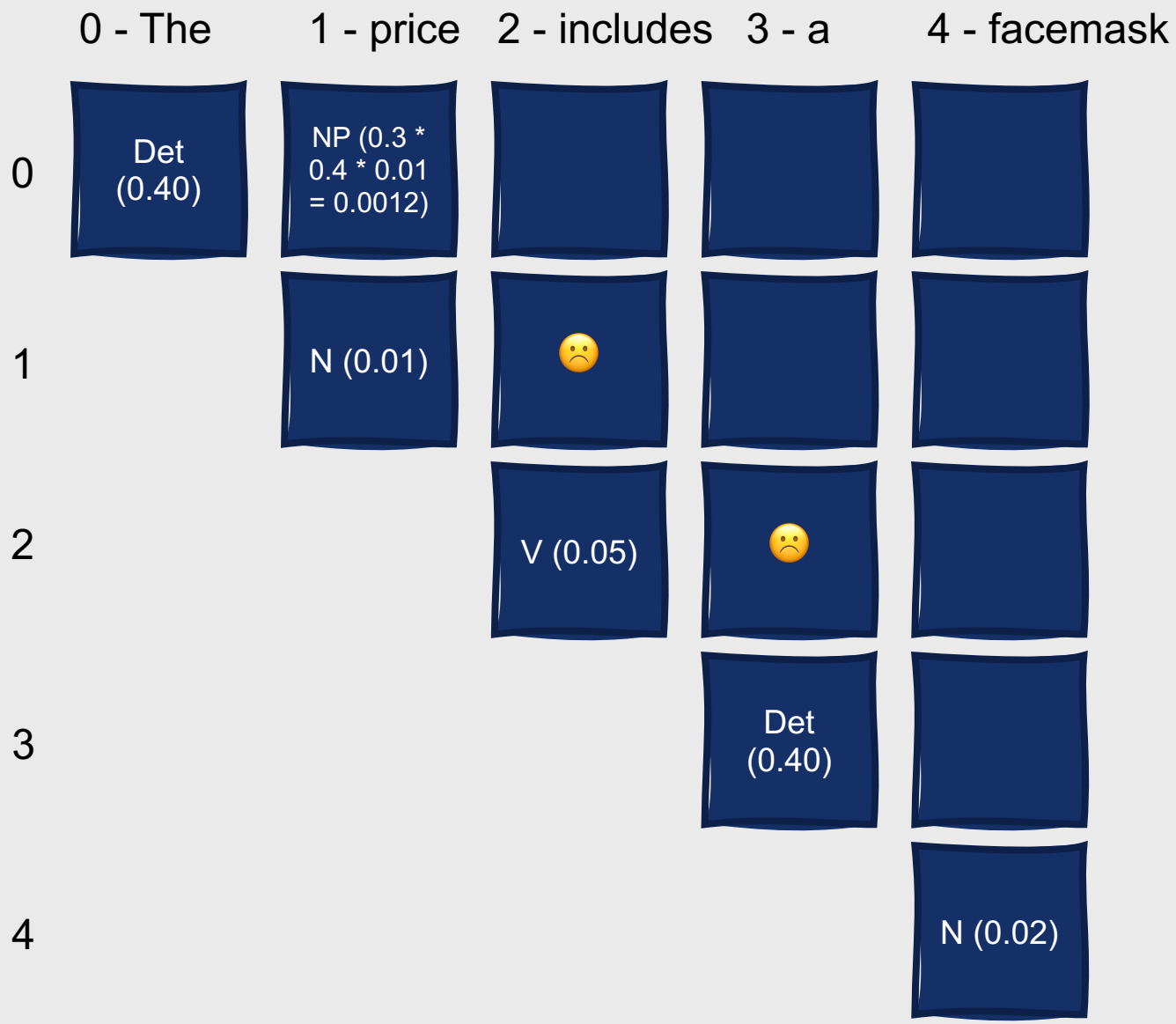
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

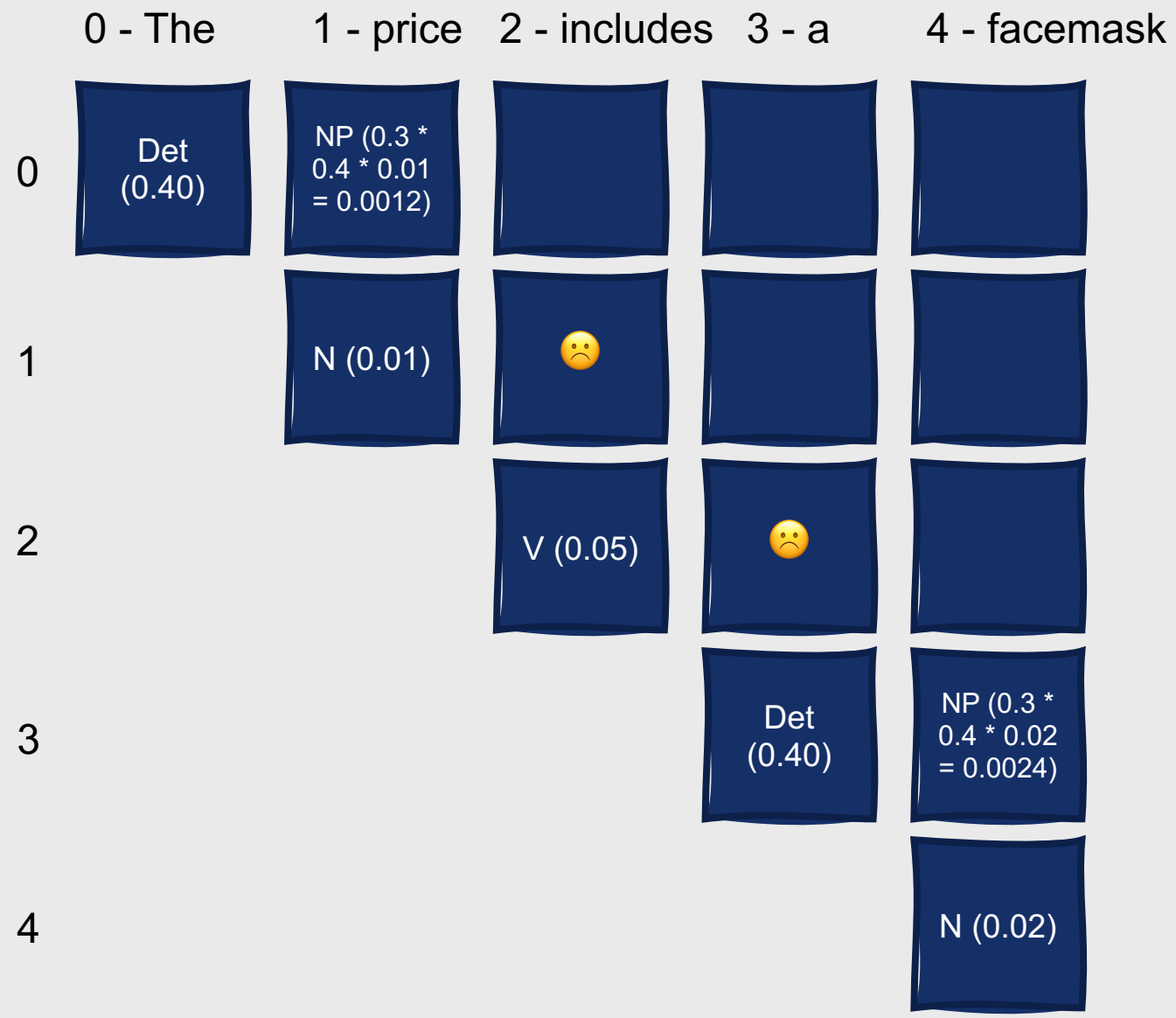
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

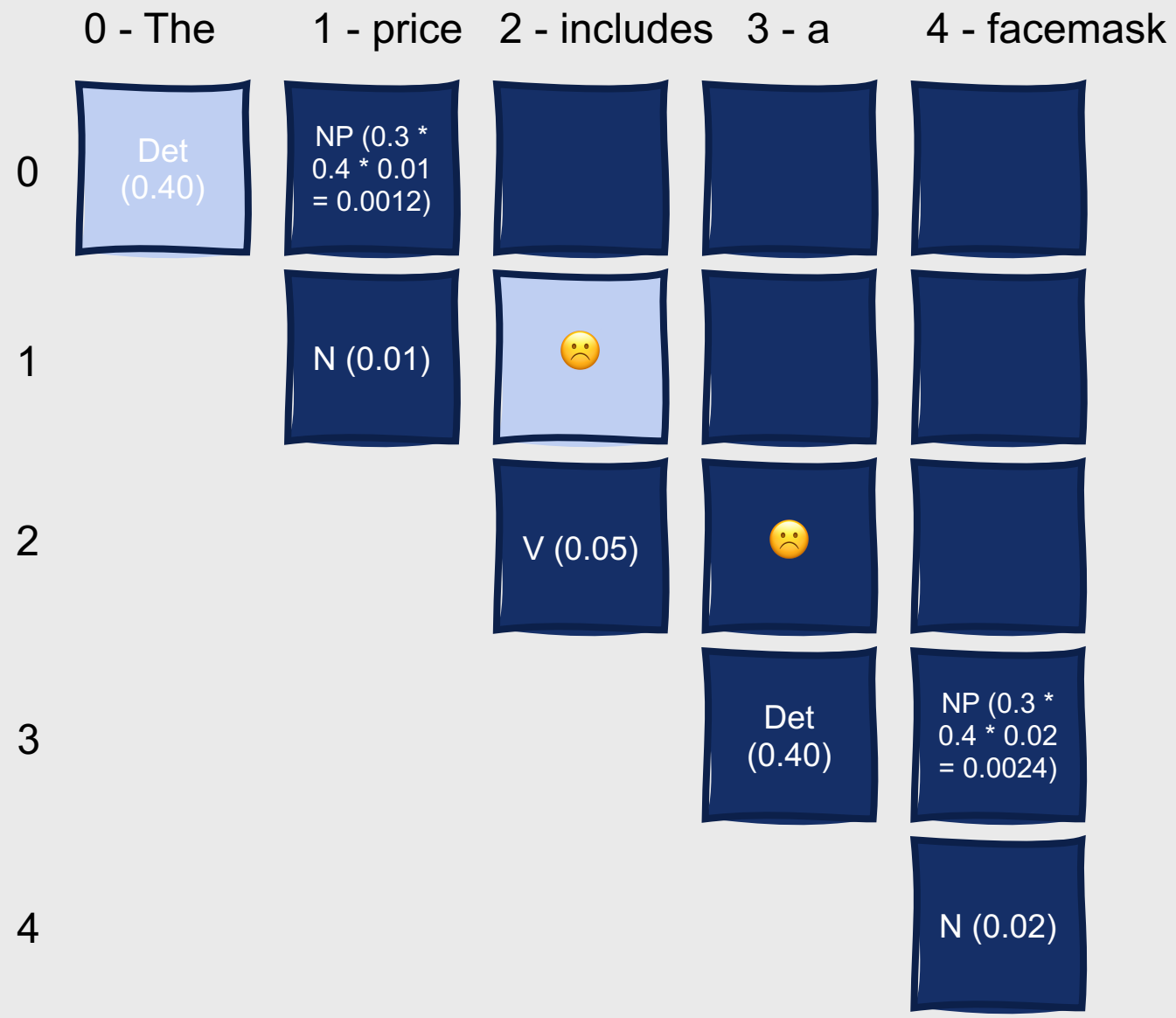
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

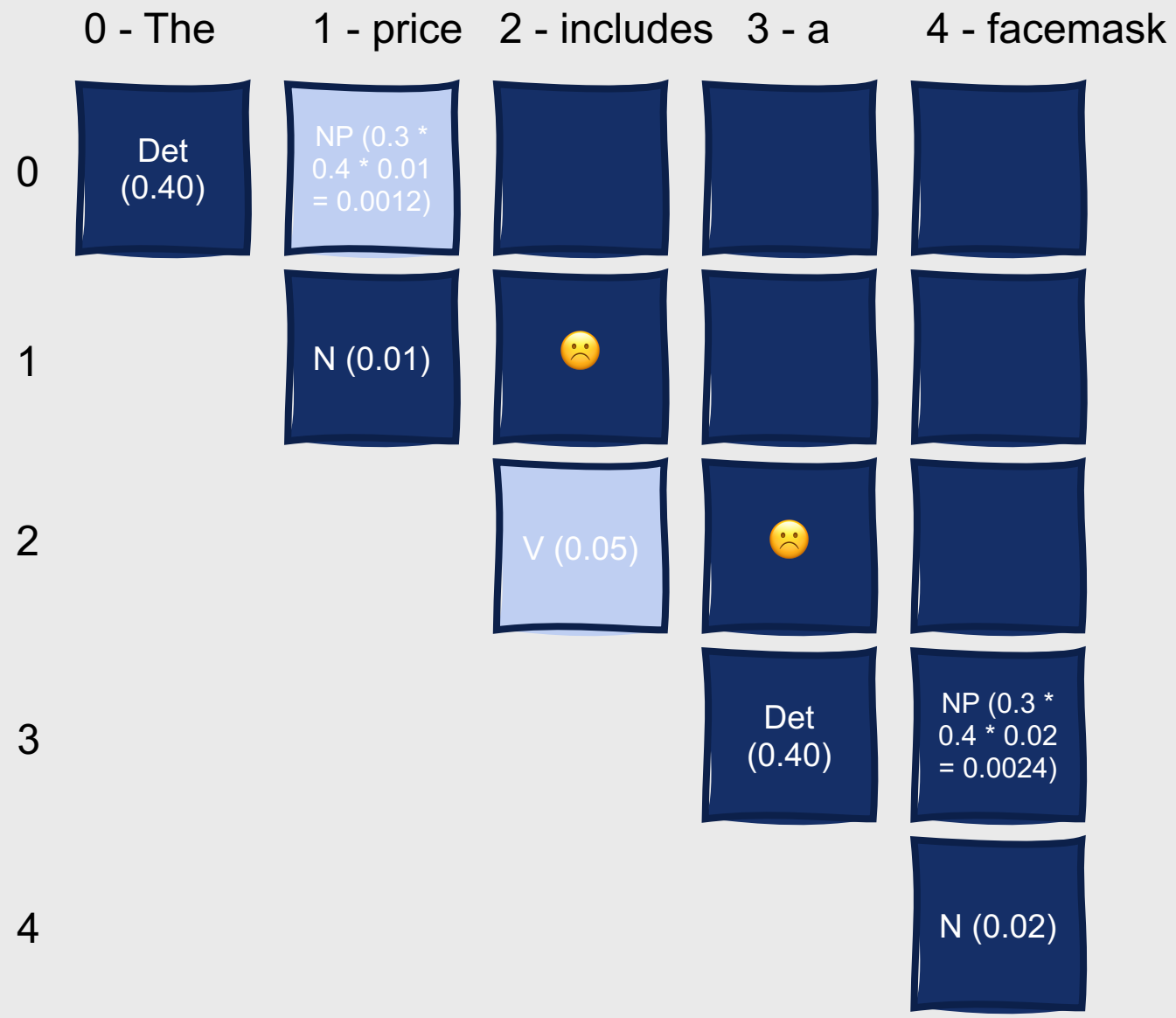
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

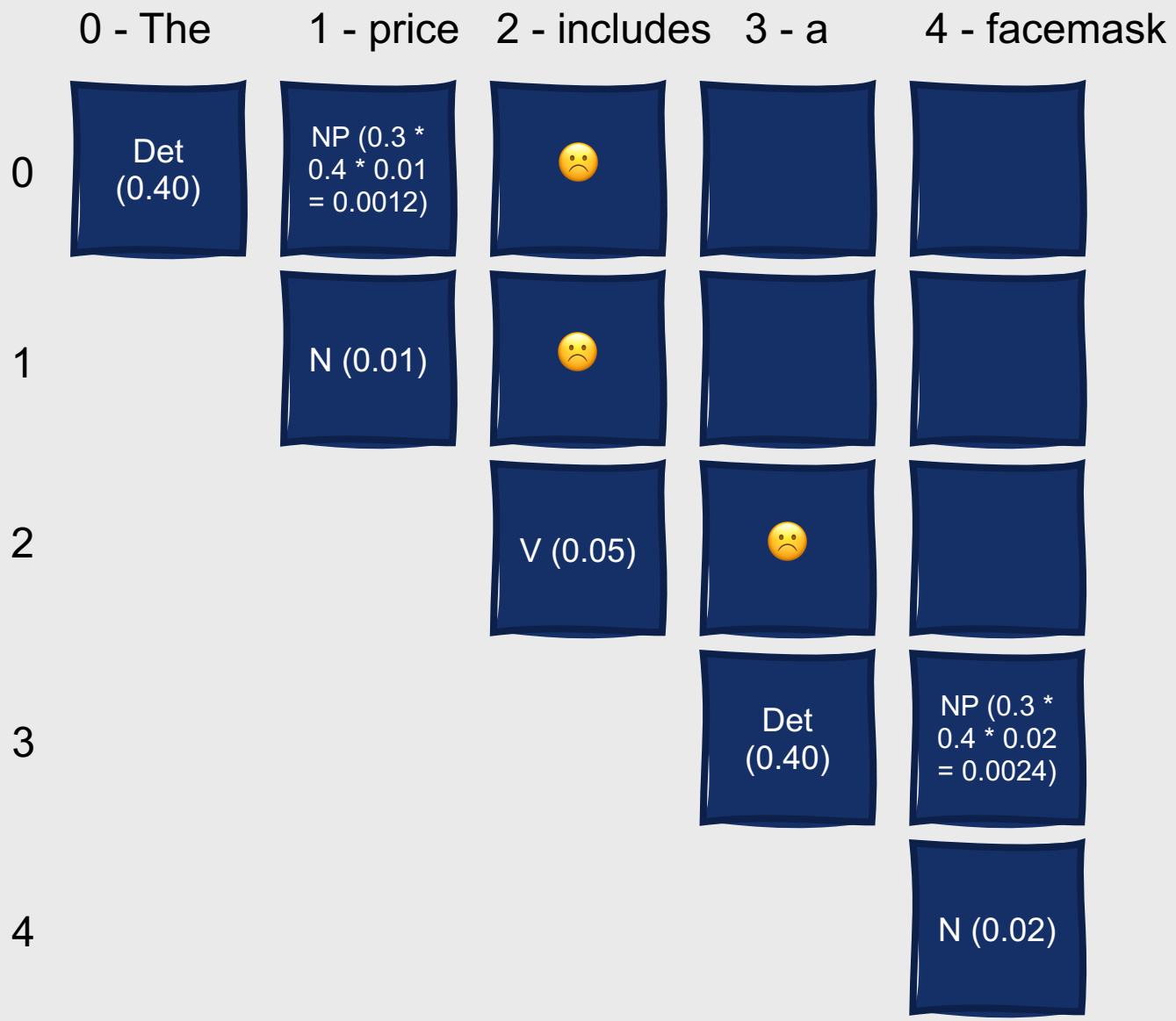
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

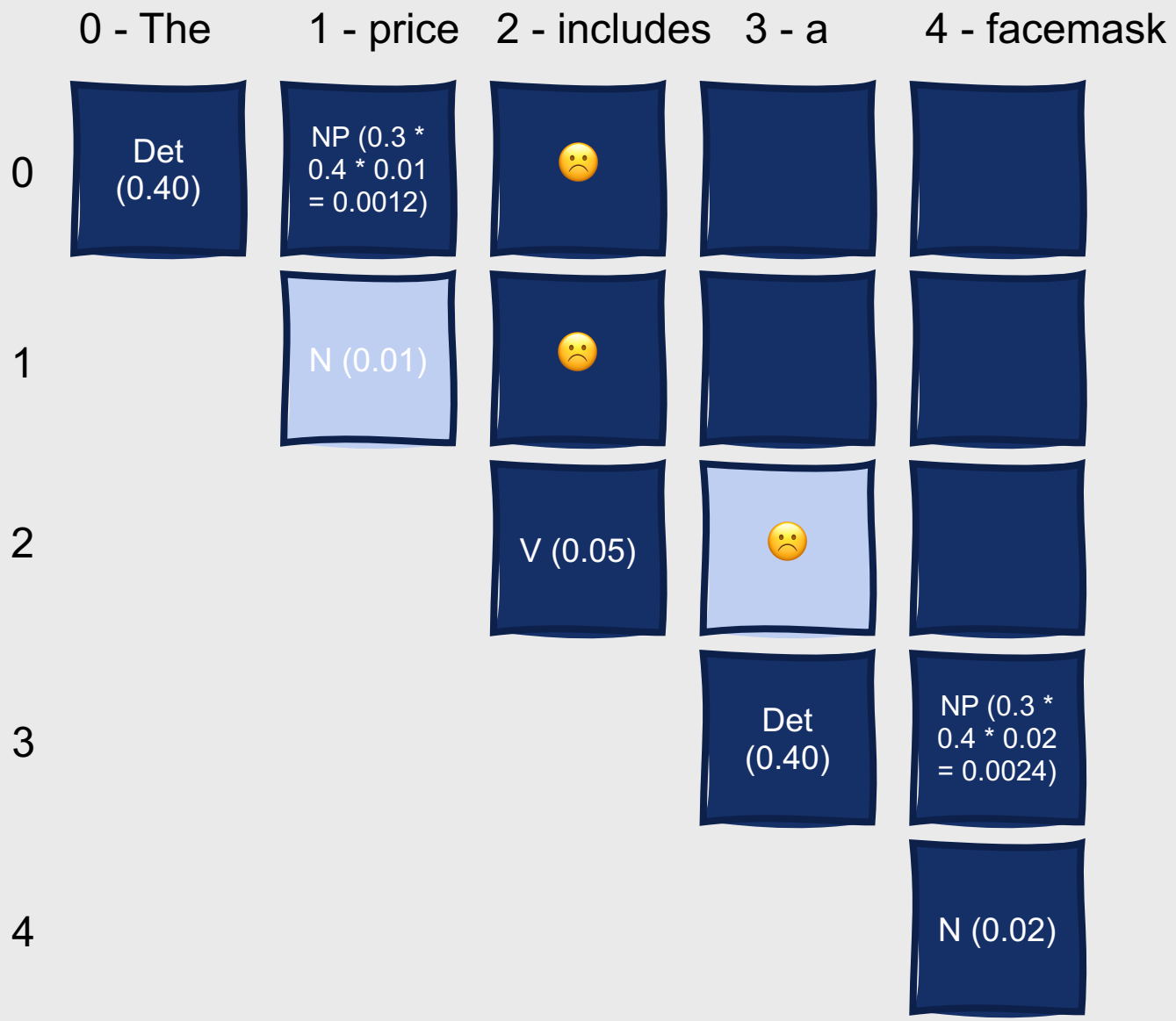
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

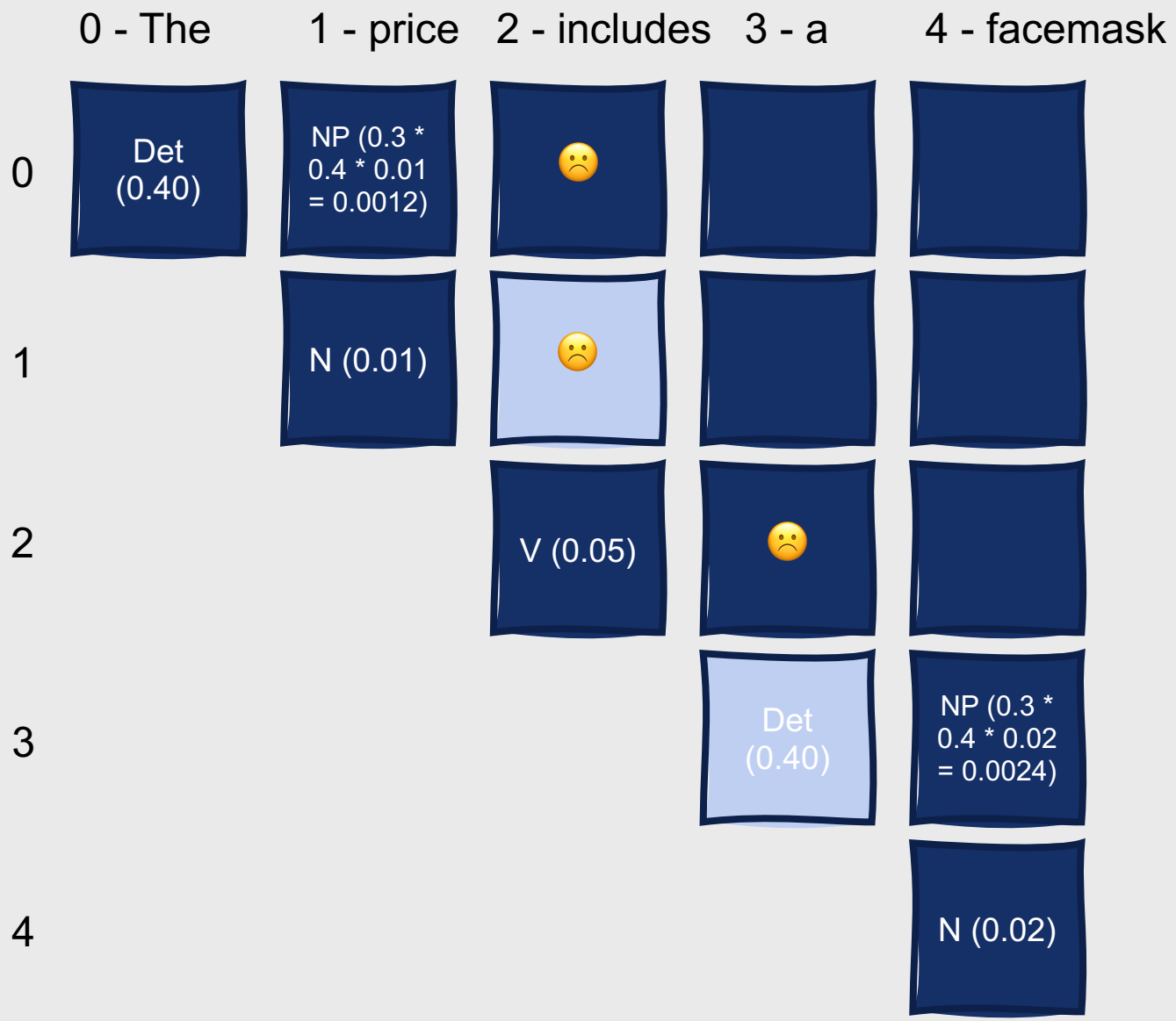
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

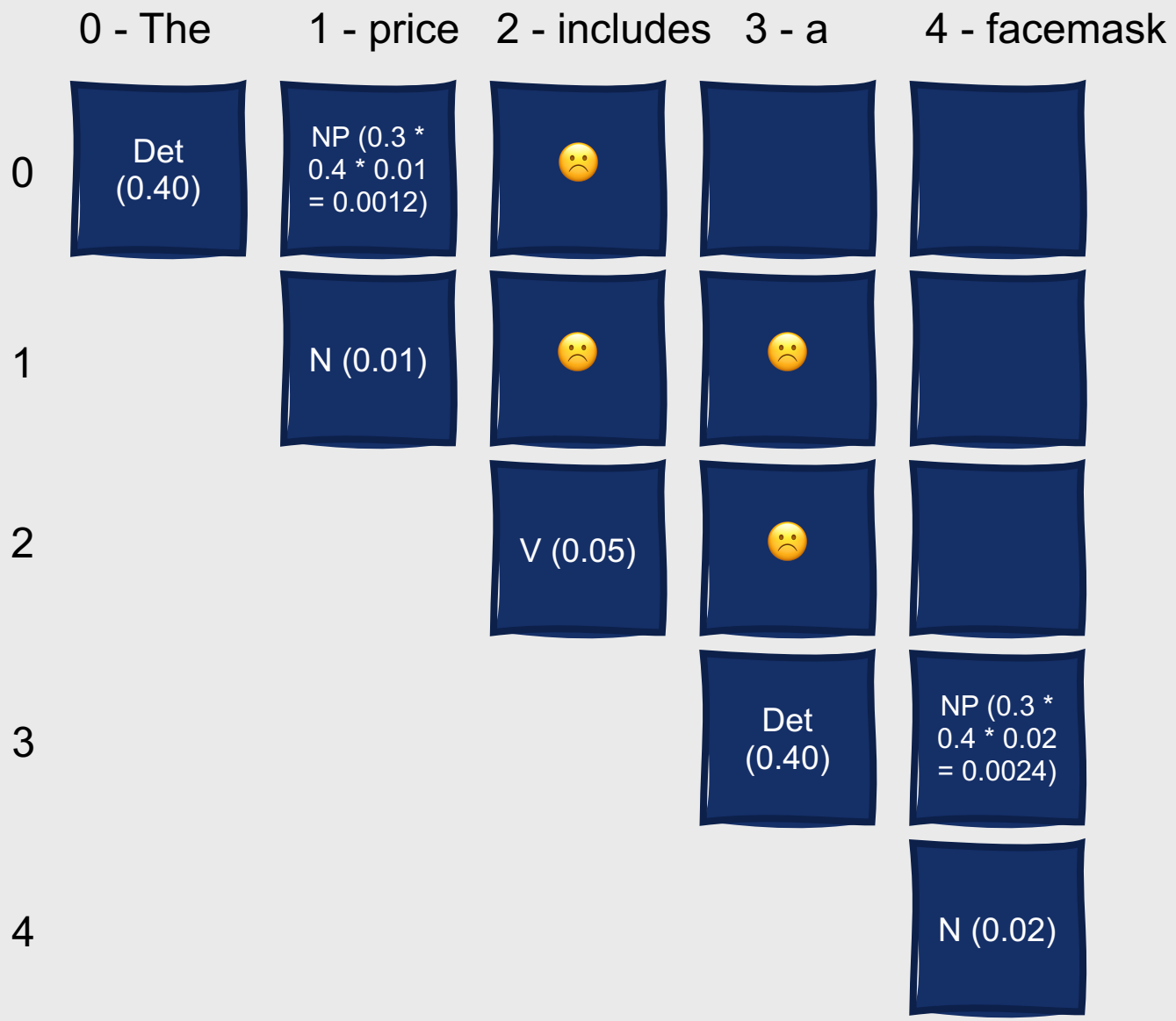
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

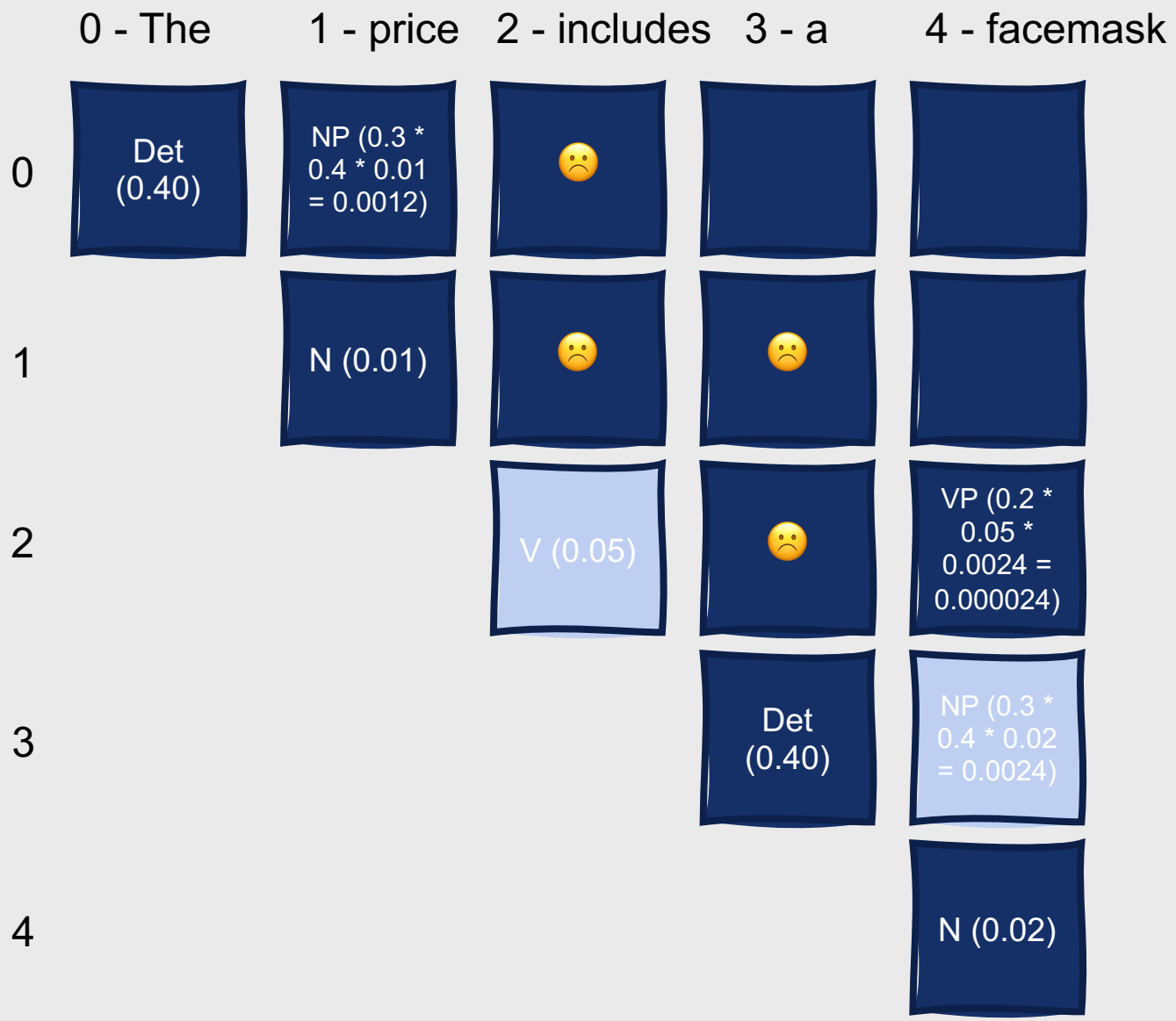
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

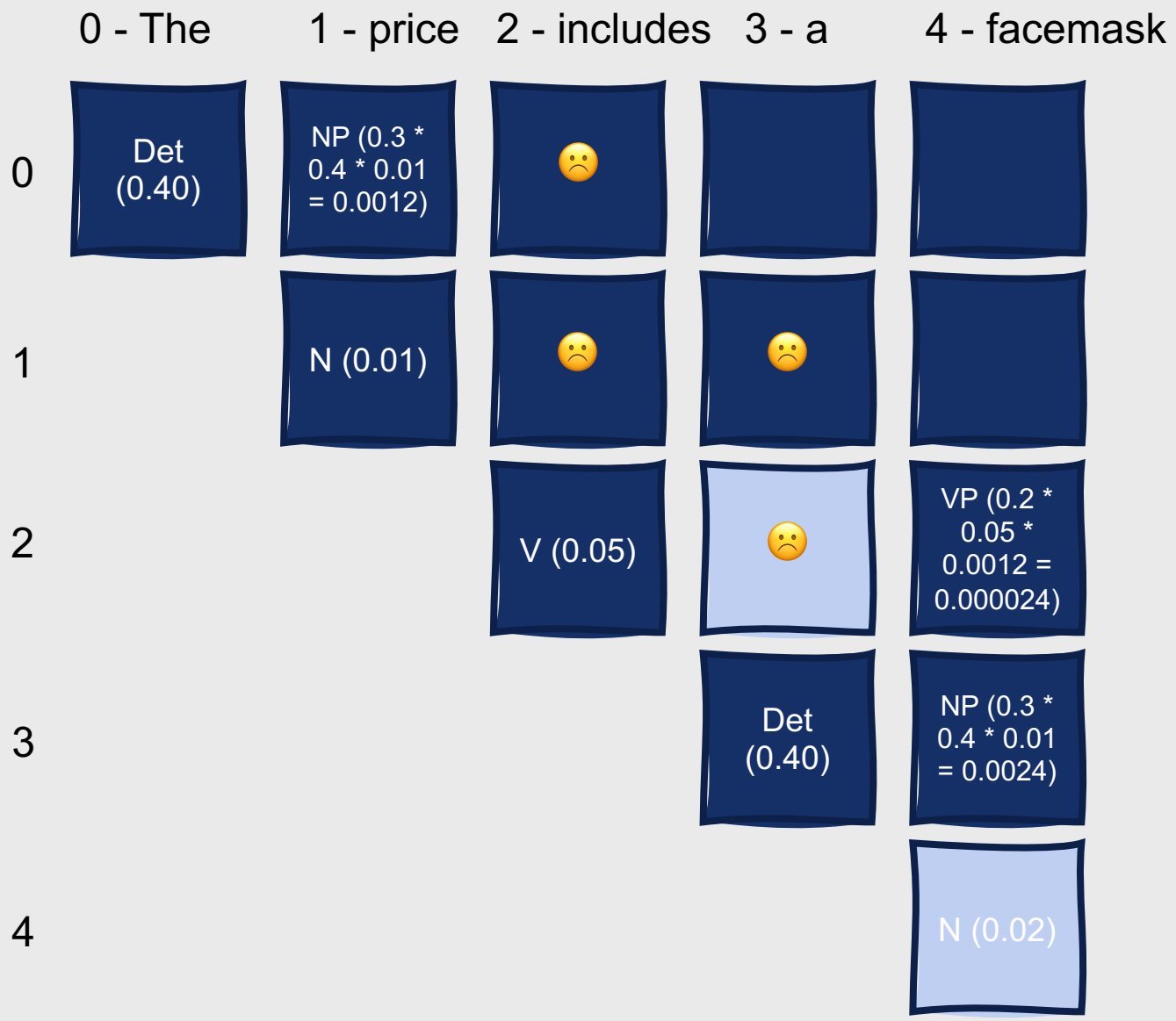
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

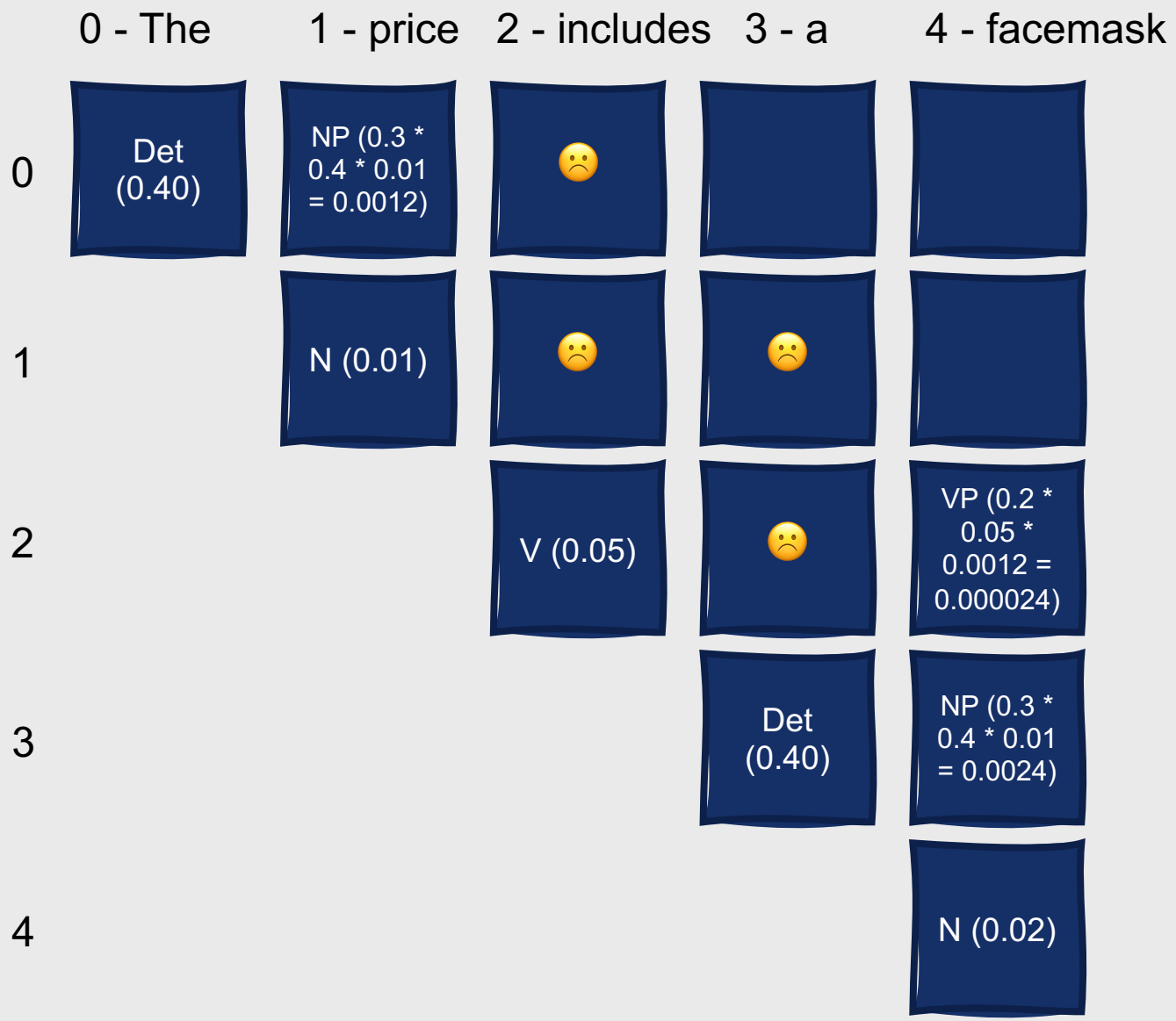
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

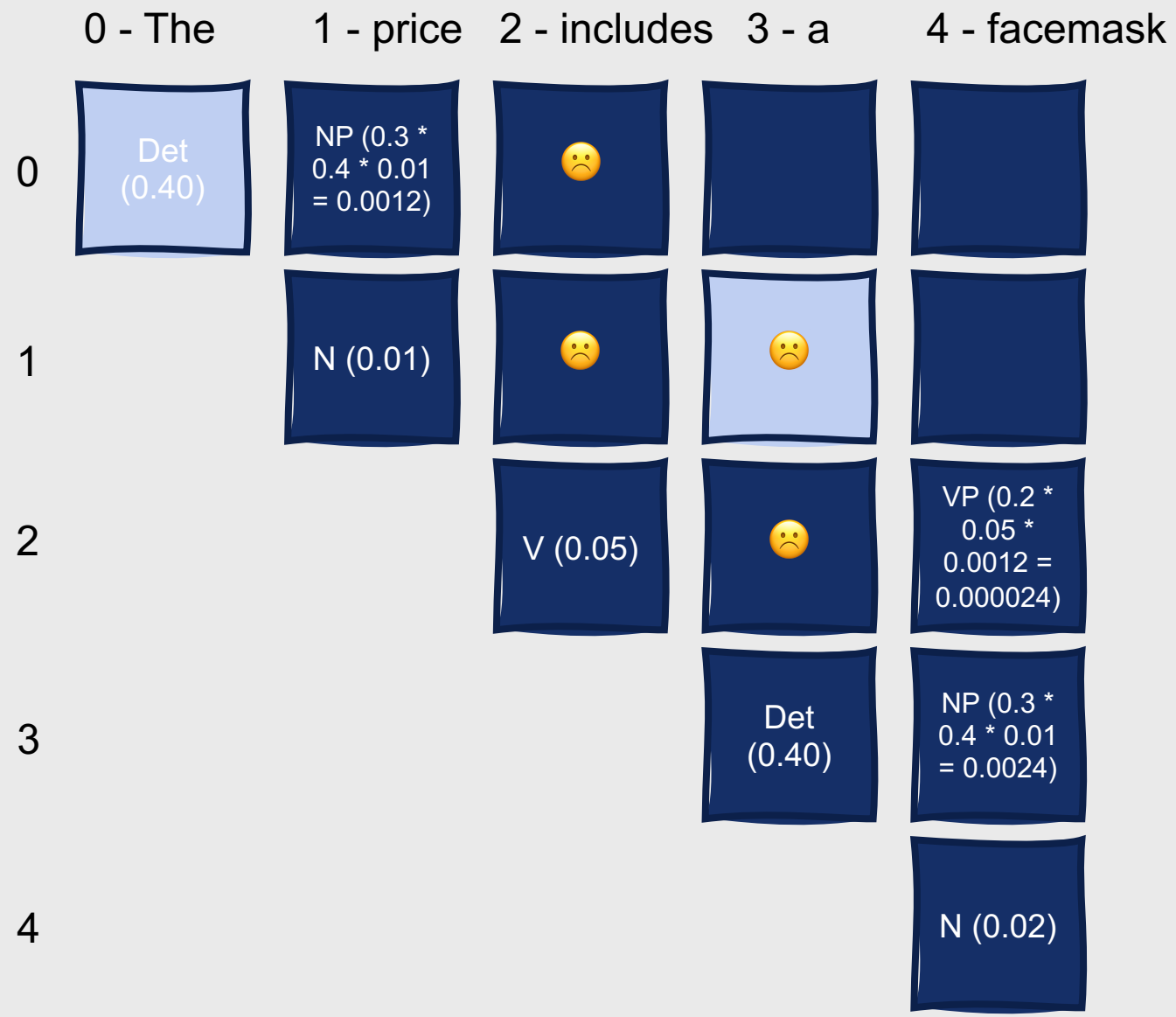
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

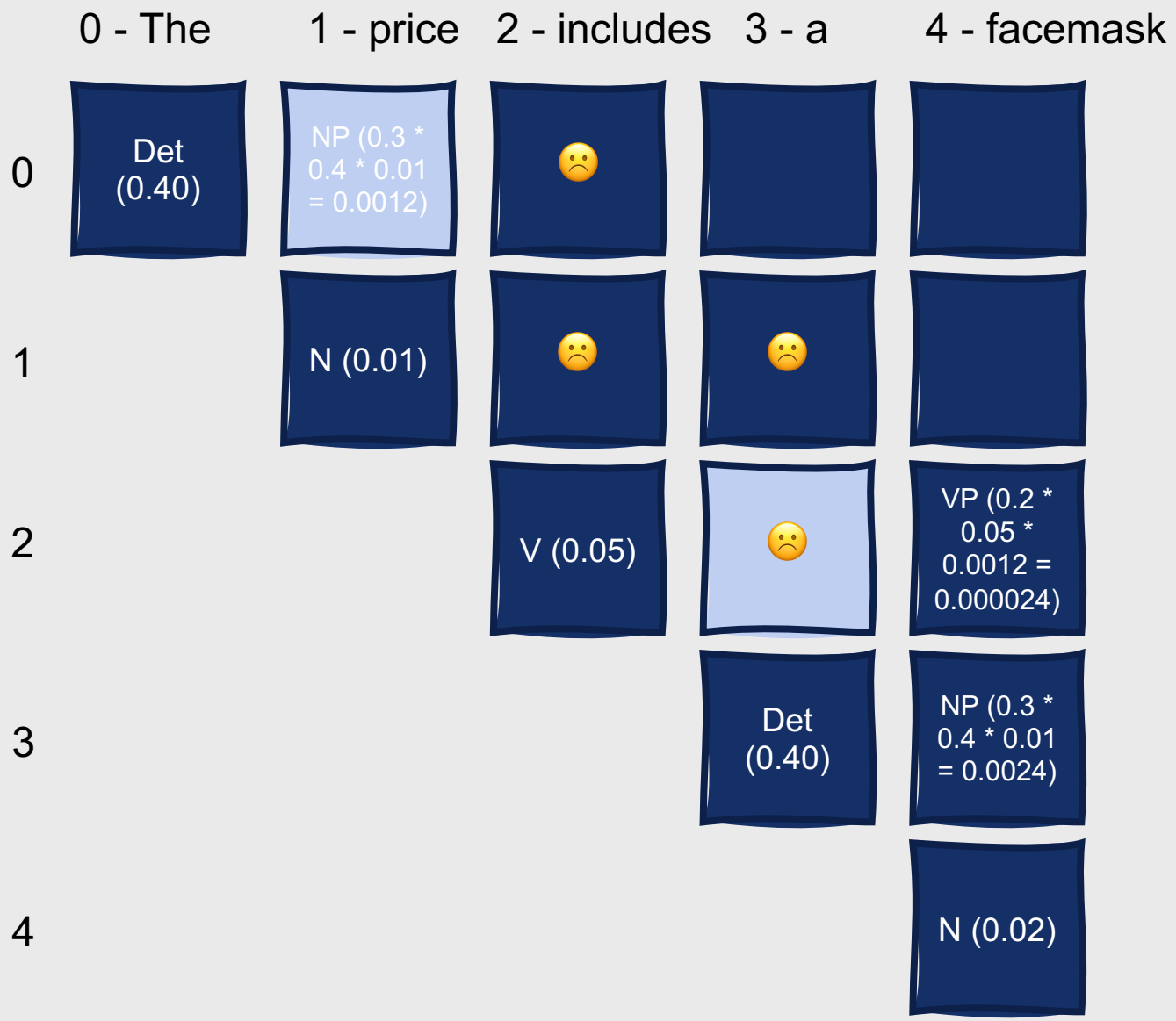
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

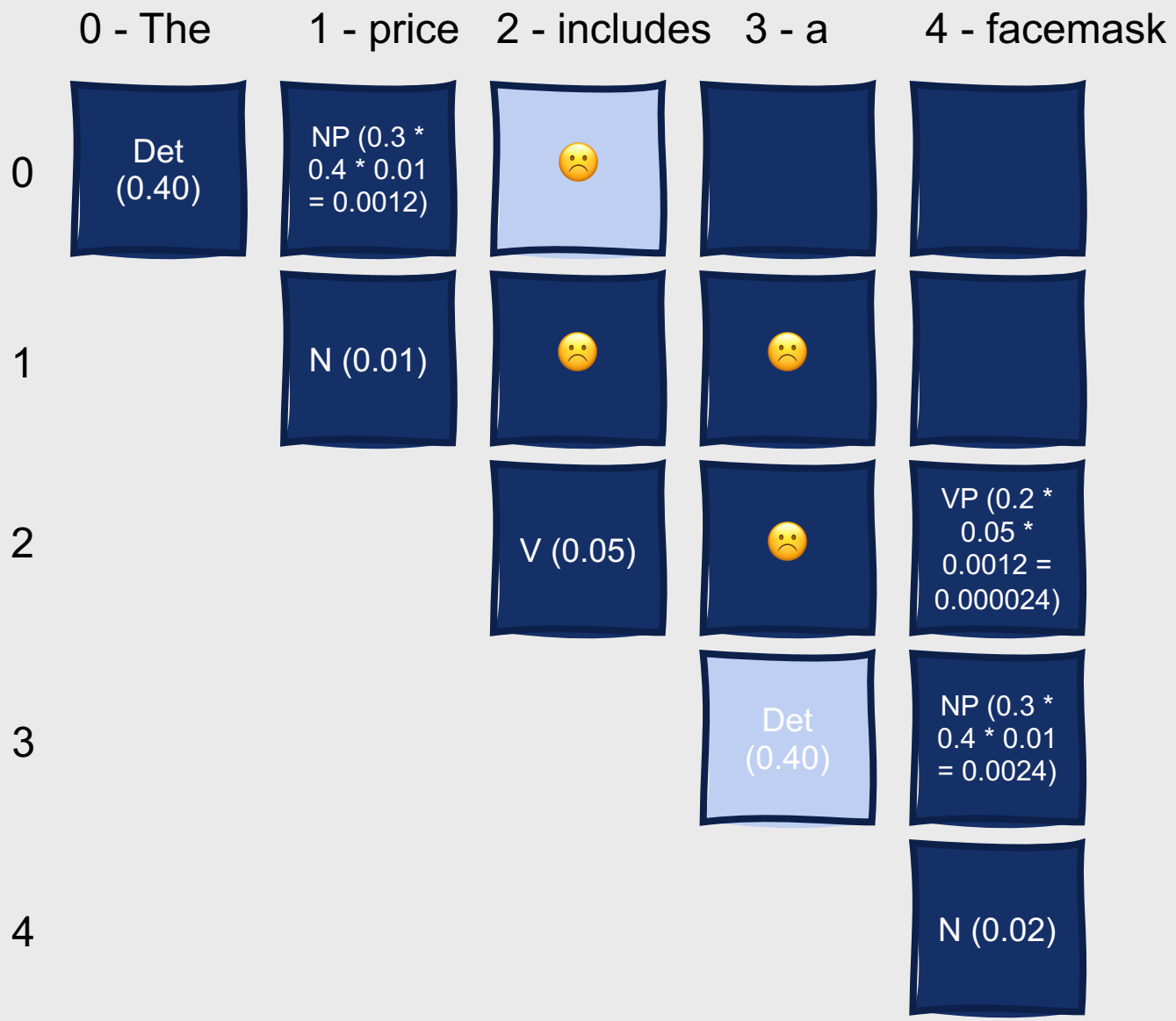
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

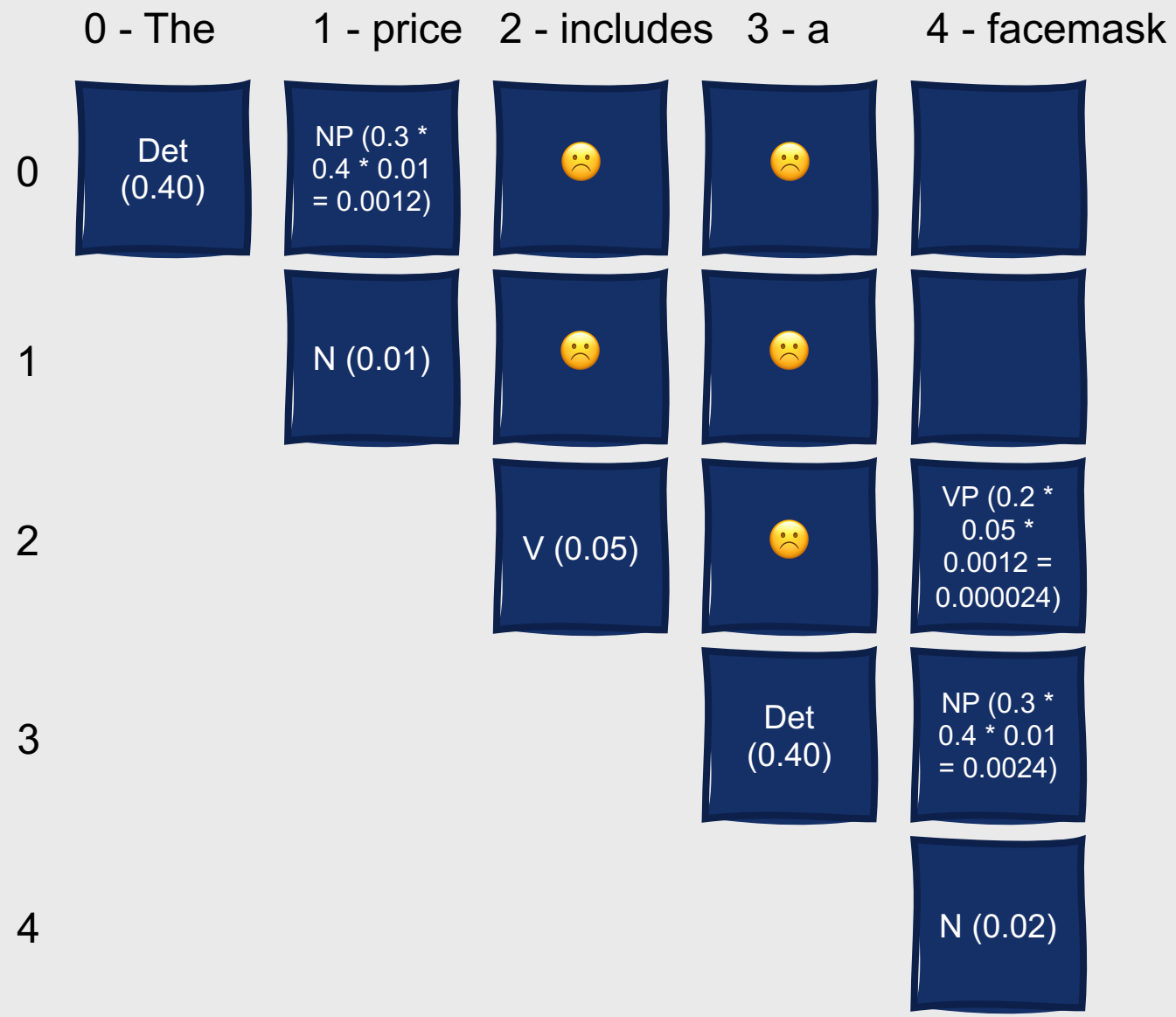
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

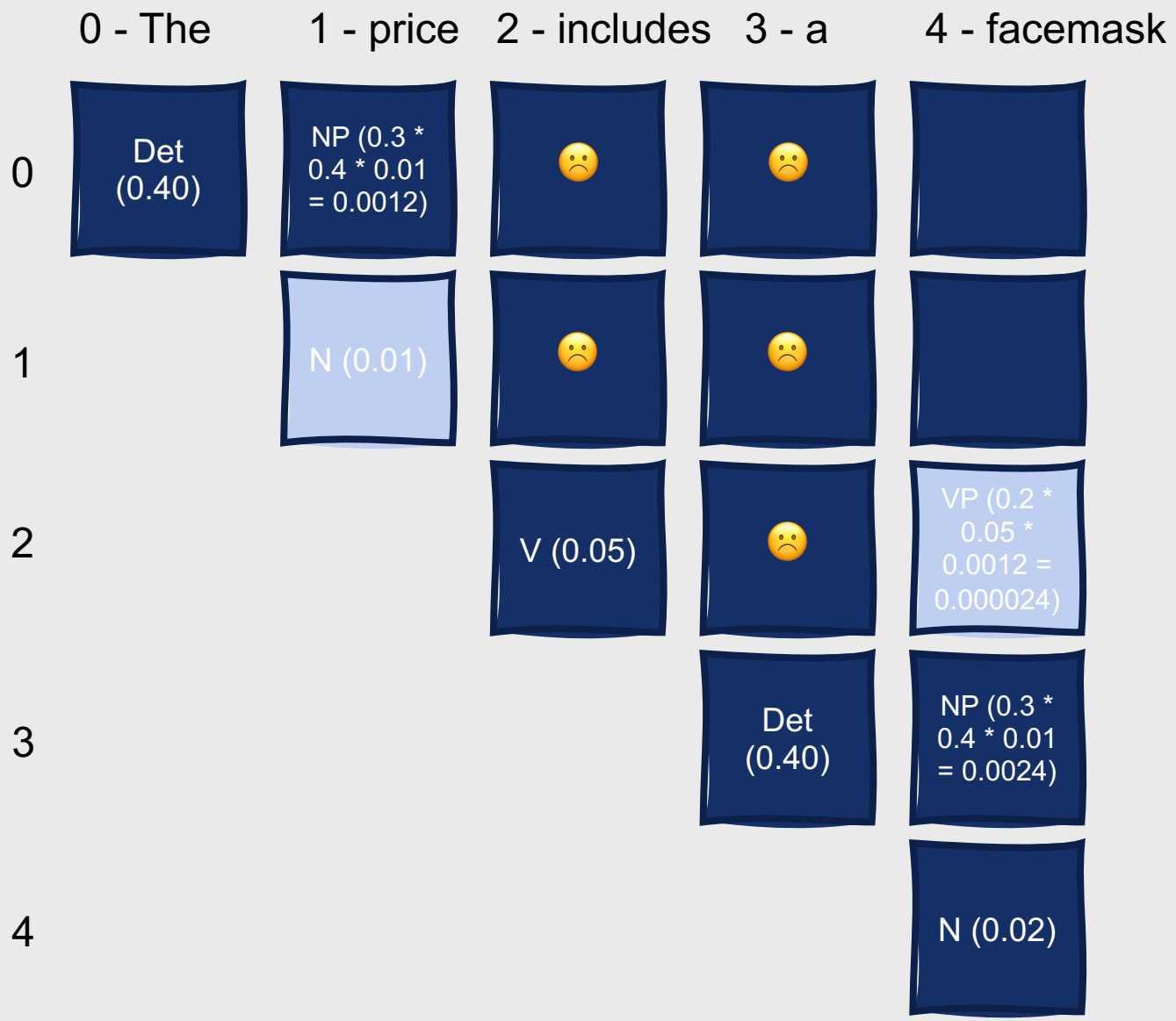
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

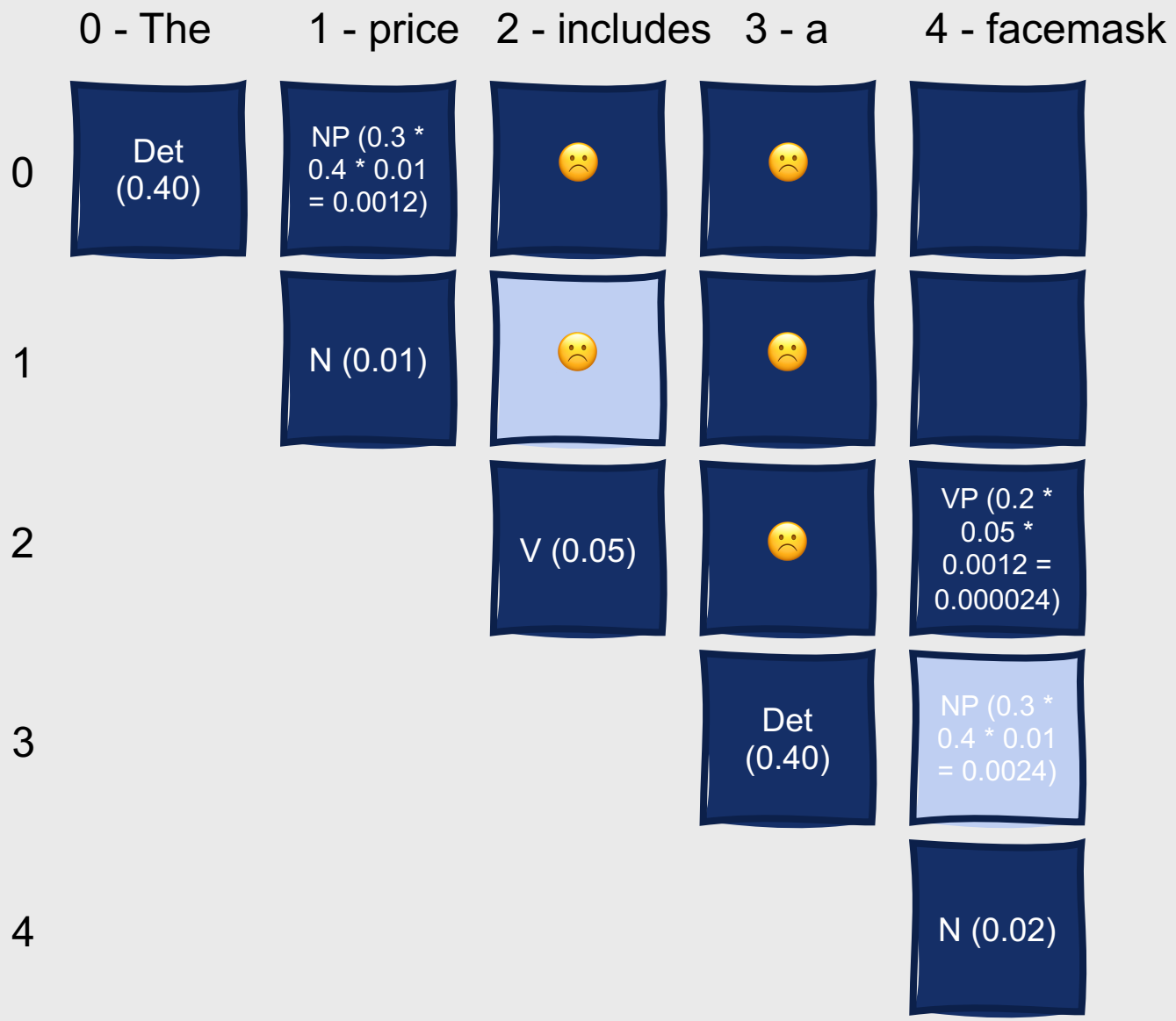
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

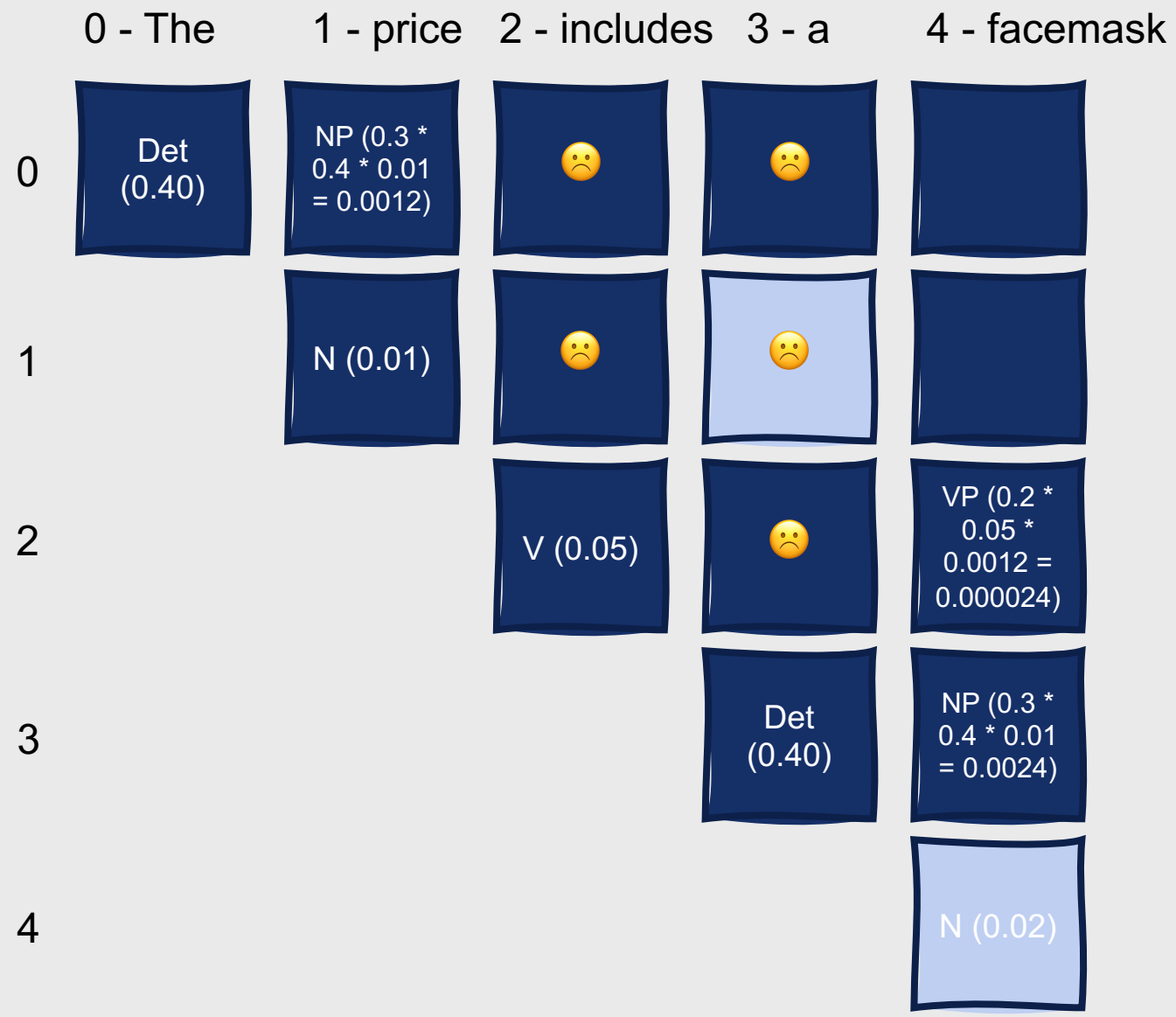
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

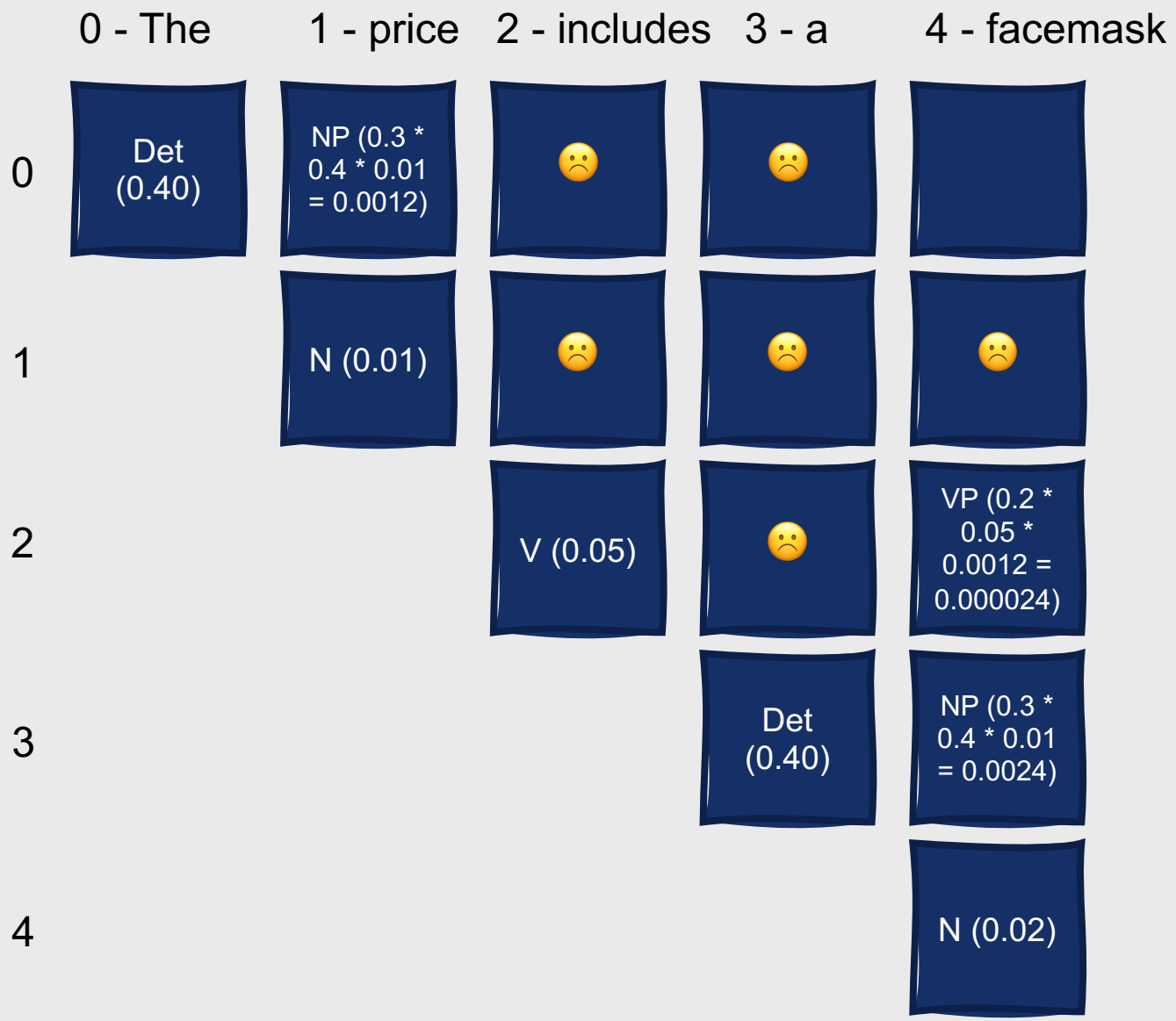
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

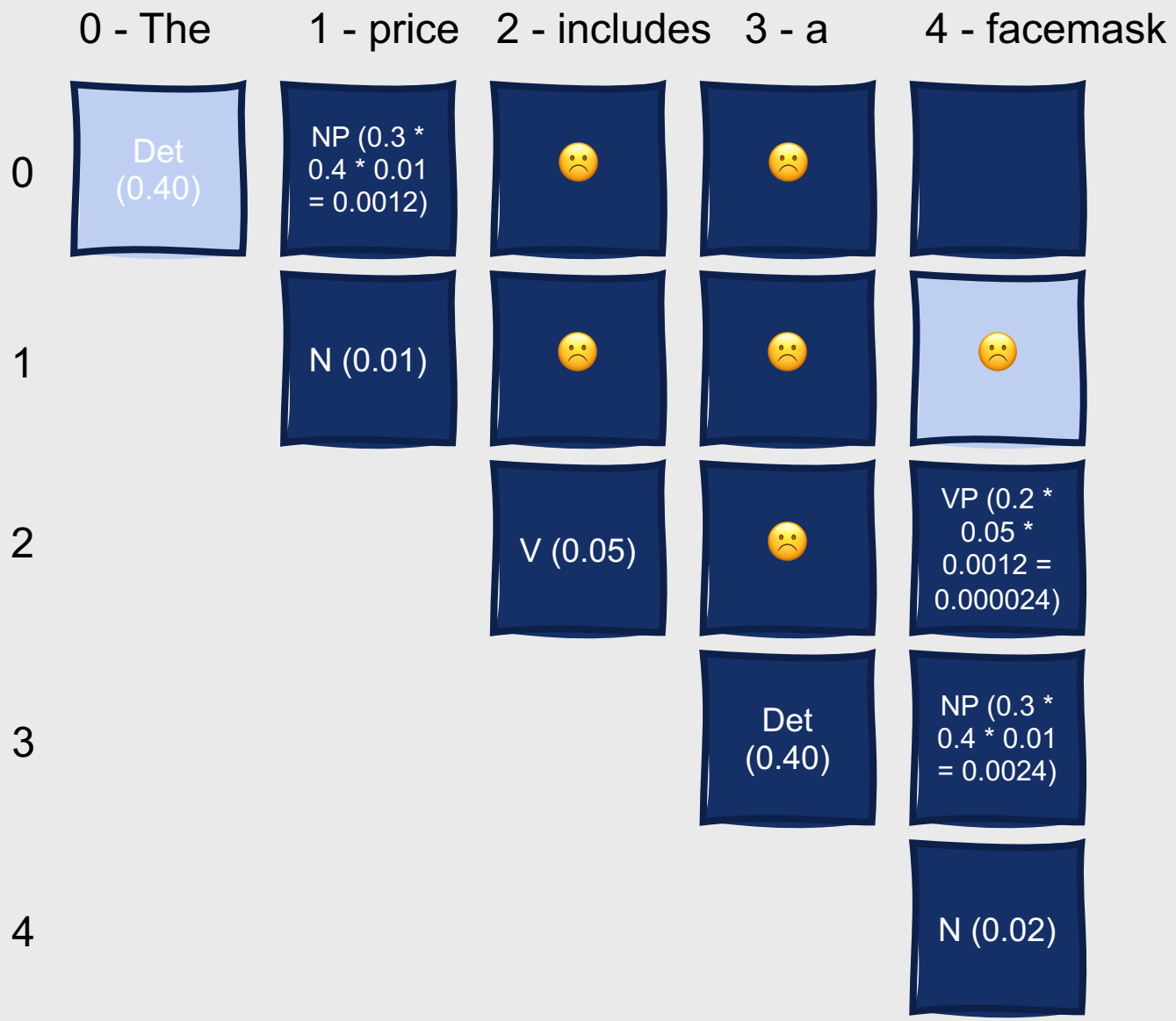
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

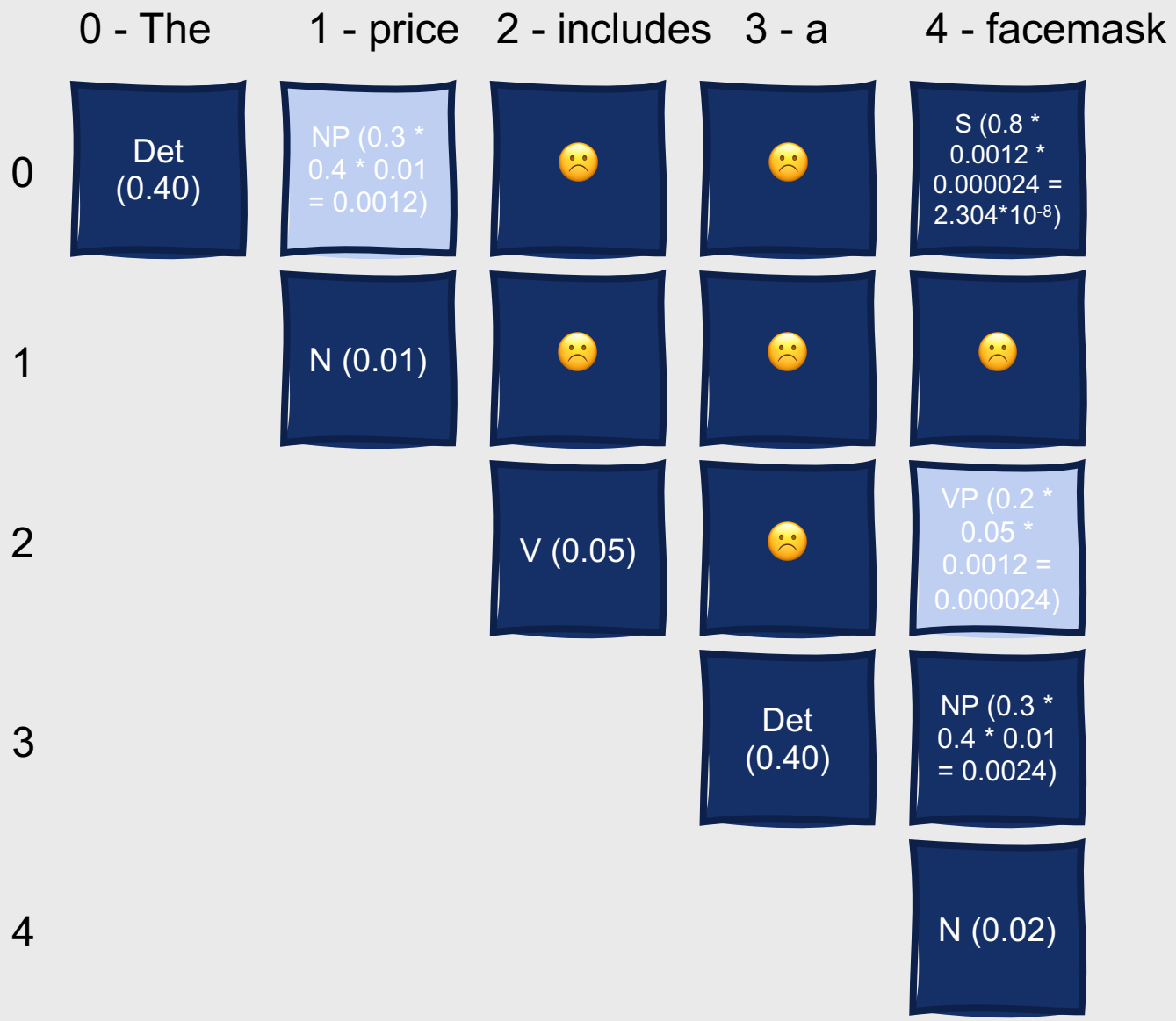
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

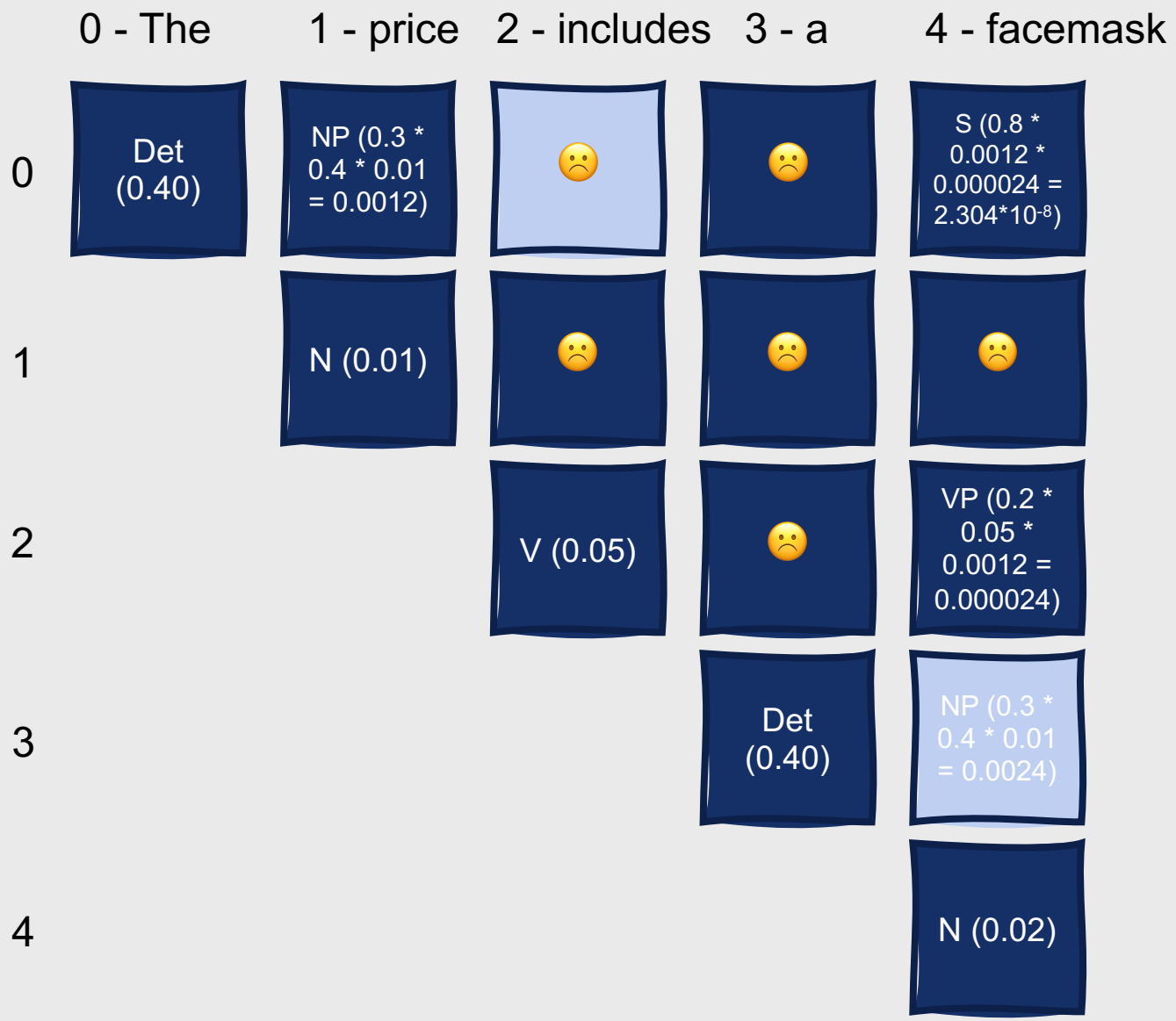
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

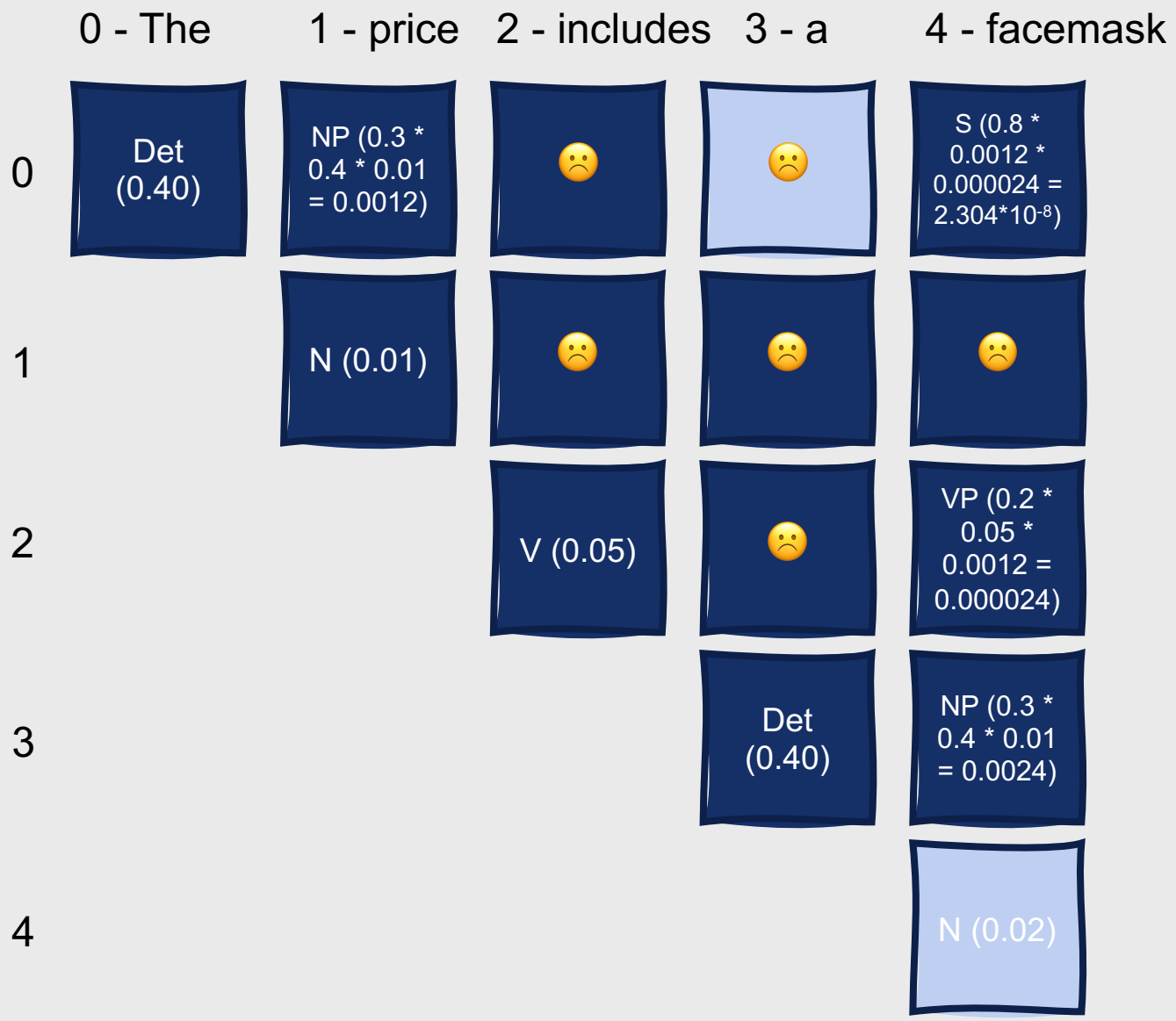
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

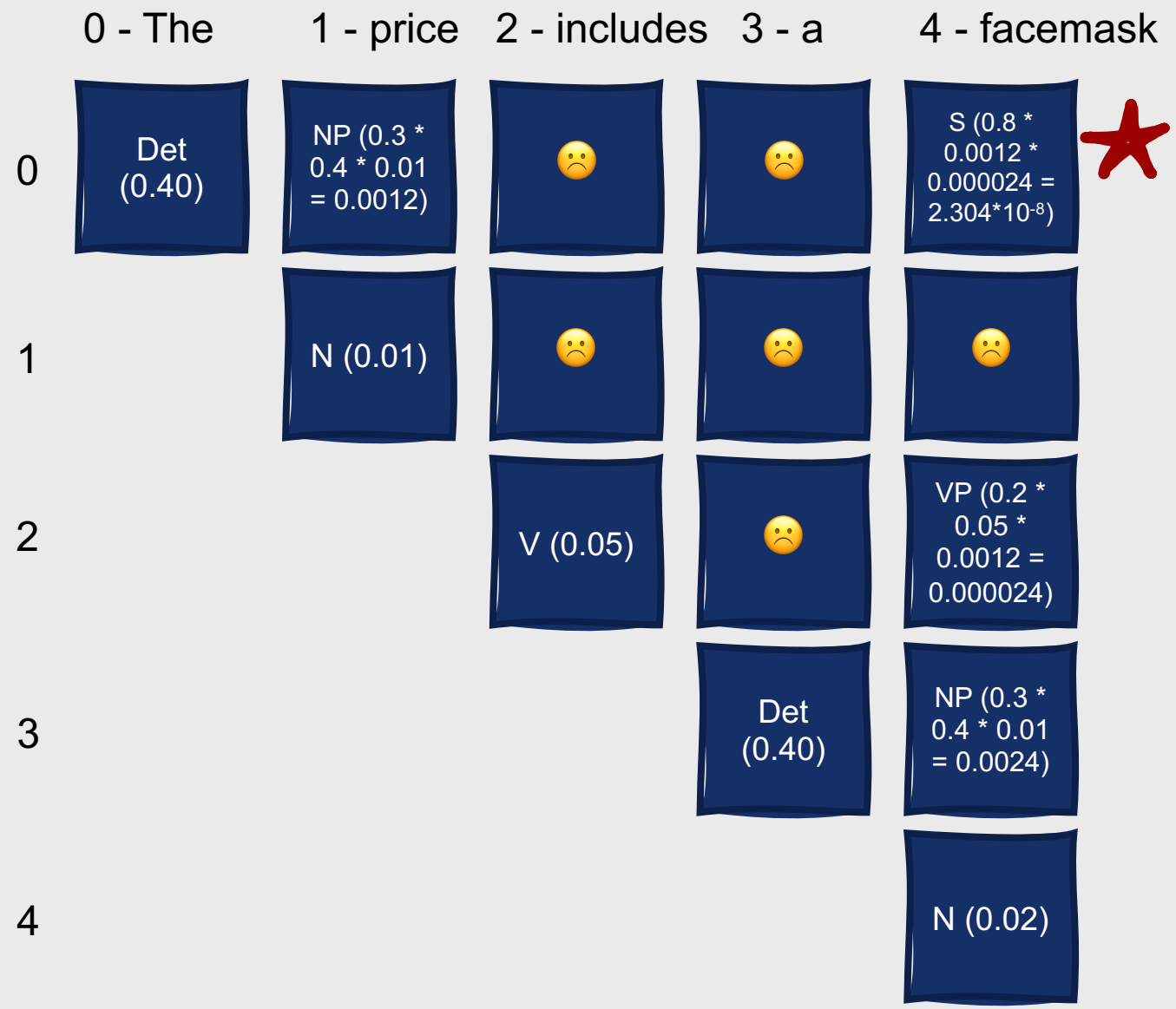
Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Case Example: Probabilistic CKY

The price includes a facemask

Production Rule	Probability
S → NP VP	0.80
NP → Det N	0.30
VP → V NP	0.20
V → includes	0.05
Det → the	0.40
Det → a	0.40
N → price	0.01
N → facemask	0.02



Where did these probabilities come from?

- Often, a corpus
  - $P(\alpha \rightarrow \beta|\alpha) = \frac{Count(\alpha \rightarrow \beta)}{\sum_{\gamma} Count(\alpha \rightarrow \gamma)} = \frac{Count(\alpha \rightarrow \beta)}{Count(\alpha)}$
- Or, if we don't have a labeled corpus, we can apply a generalization of the forward-backward algorithm called the **inside-out algorithm**
  - Start with equal probabilities for each rule
  - Parse the input
  - Compute a probability for each parse
  - Weight the counts based on these probabilities
  - Re-estimate the probabilities accordingly
  - Repeat until convergence