Probability and Language

Quick Recap

Quick Recap

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲呑	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲呑湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

Quick Recap

Develop a statistical *model* of translation that can be *learned* from *data* and used to *predict* the correct English translation of new Chinese sentences.

Quick Recap

- *Minimally*, our model must account for:
  - Lexical ambiguity.
  - One-to-many translation.
  - Many-to-many translation.
  - Untranslated words.
  - Word reordering.

Quick Recap

- Oh, and it would probably be good to include:
  - Fluent output.
  - Adequate transfer of source language meaning.

Quick Recap

Although north wind howls , but sky still very clear .

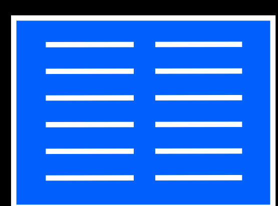
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

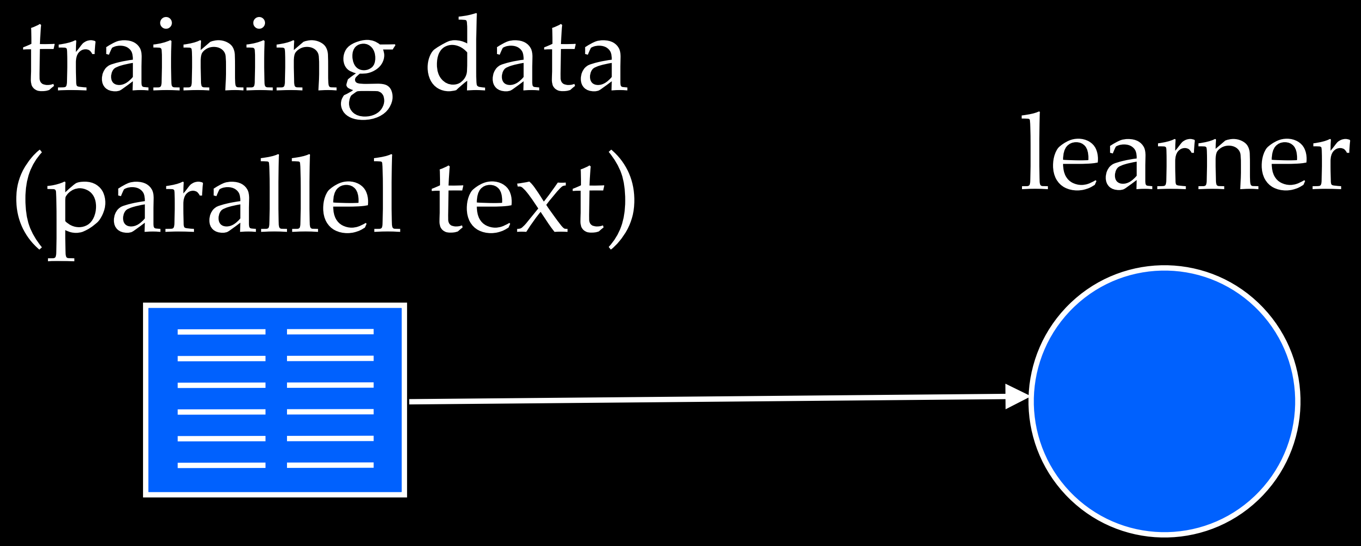
Quick Recap

Quick Recap

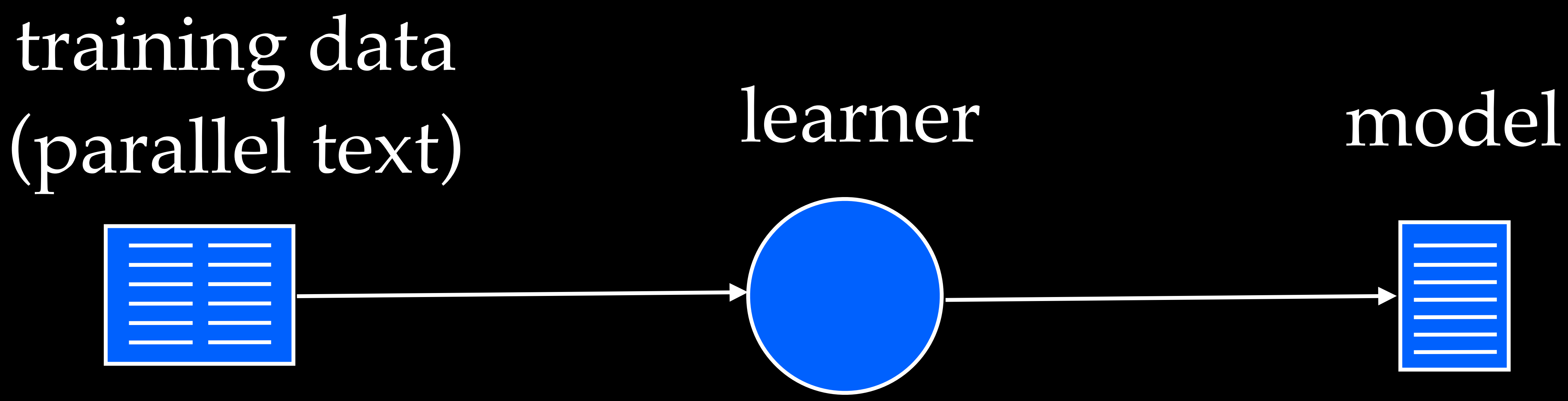
training data
(parallel text)



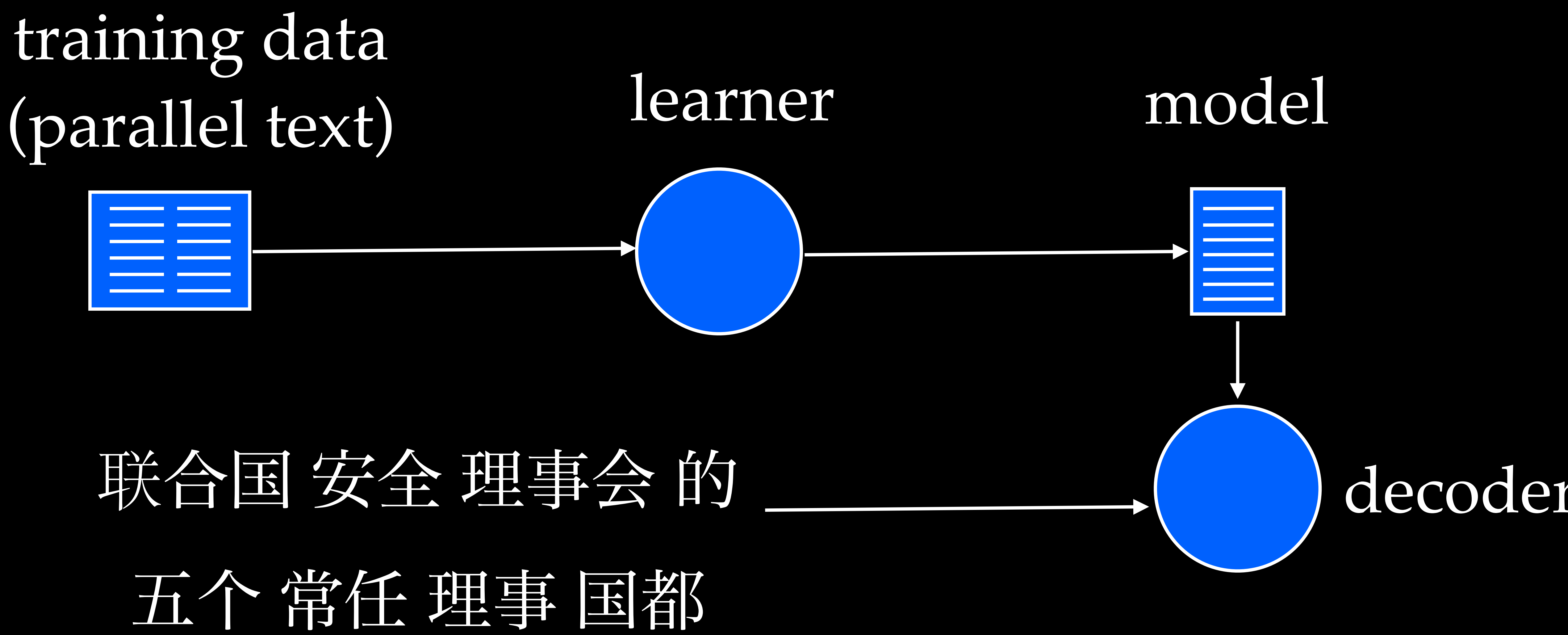
Quick Recap



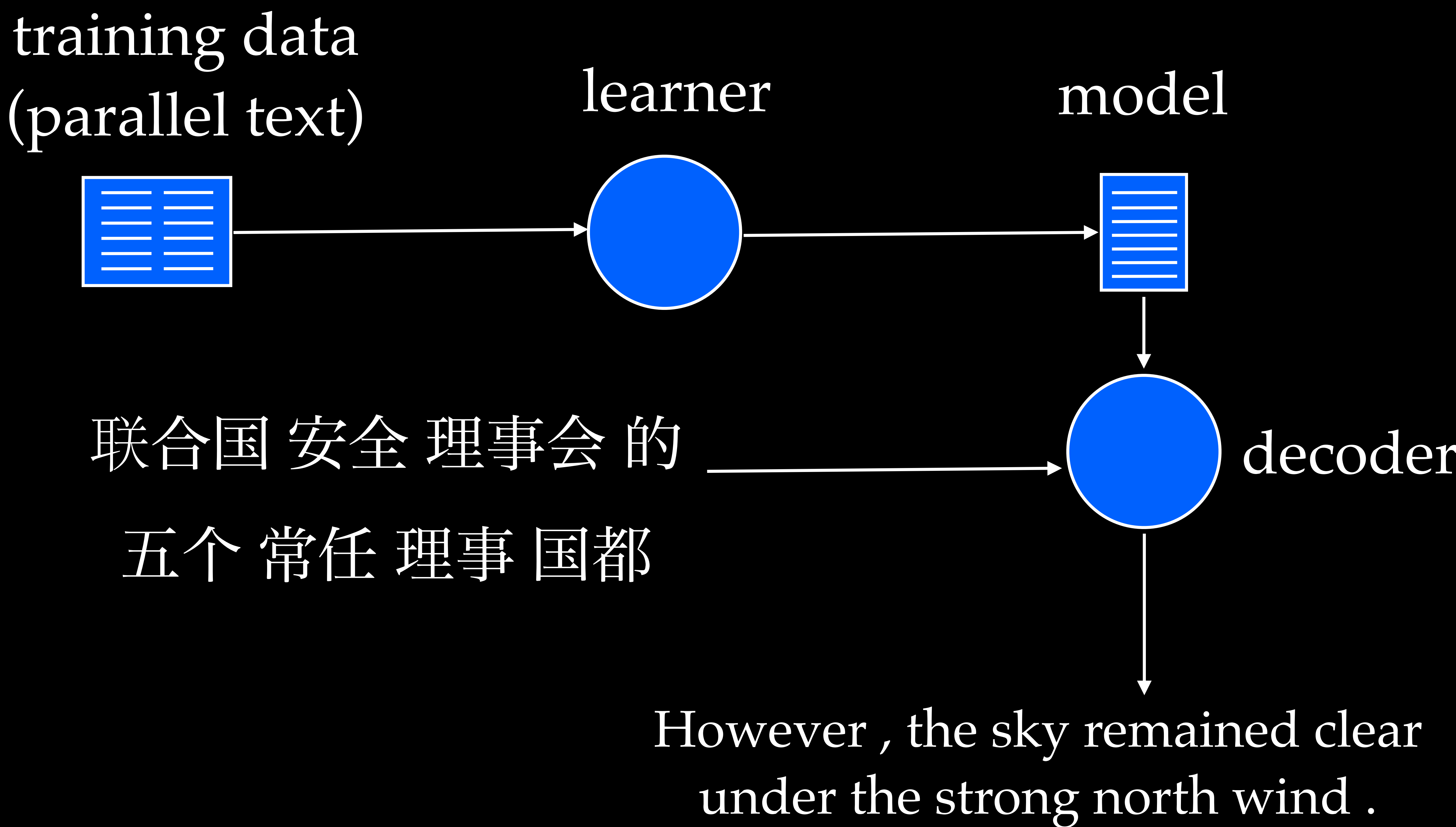
Quick Recap



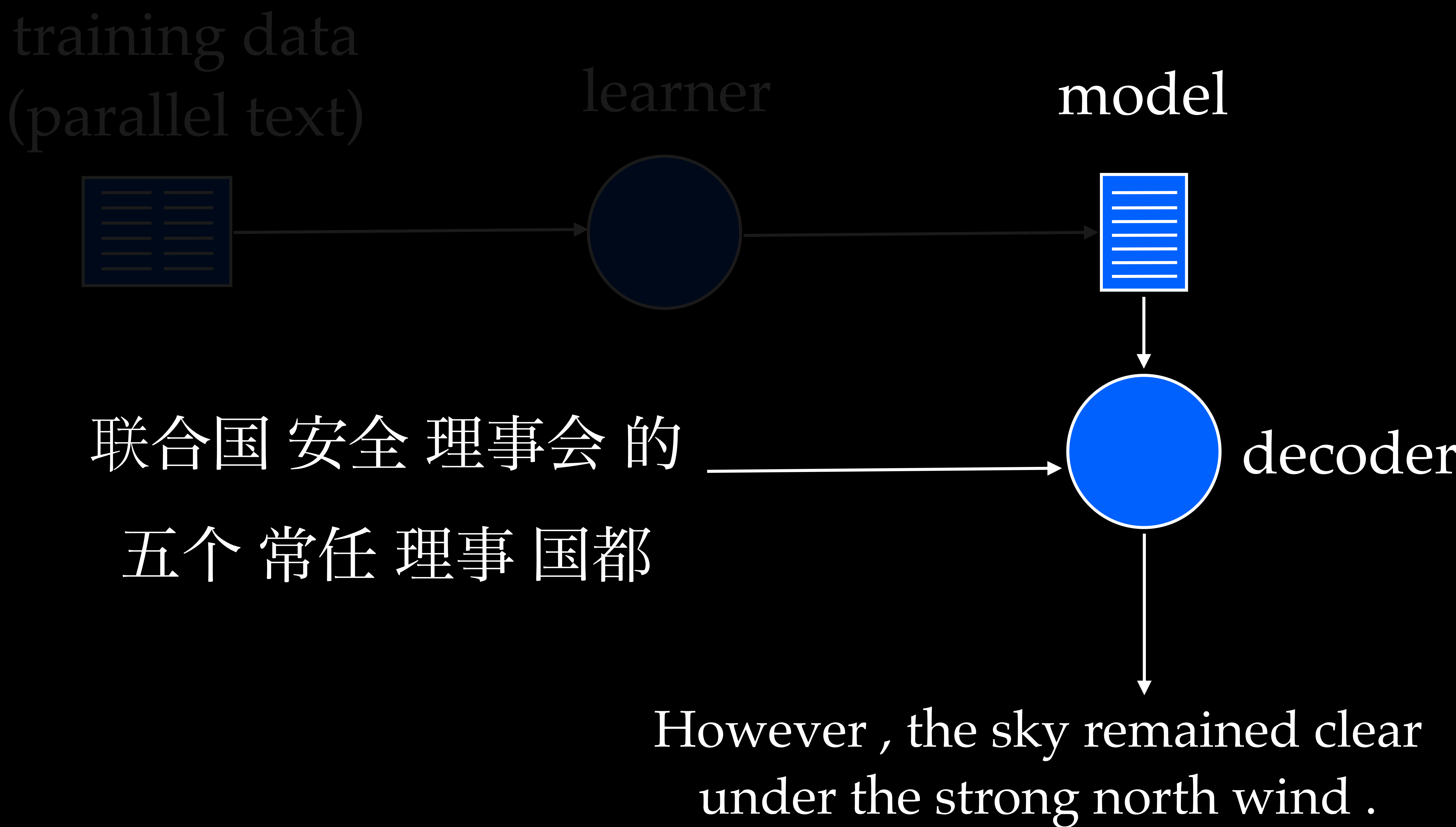
Quick Recap


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder

Quick Recap


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Quick Recap


training data
(parallel text)
learner
model
decoder
联合国 安全 理事会 的
五个 常任 理事 国都
However , the sky remained clear
under the strong north wind .

What’s a model?

What's a model?

For our purposes, a model will be
a probability distribution over sentence pairs.

What’s a model?

For our purposes, a model will be
a probability distribution over sentence pairs.

NOTE ASSUMPTION

Why Probability?

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for ***learning*** and ***prediction***.

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for ***learning*** and ***prediction***.
- Allows us to answer questions:

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for ***learning*** and ***prediction***.
- Allows us to answer questions:
  - What is the best explanation of observed data?

Why Probability?

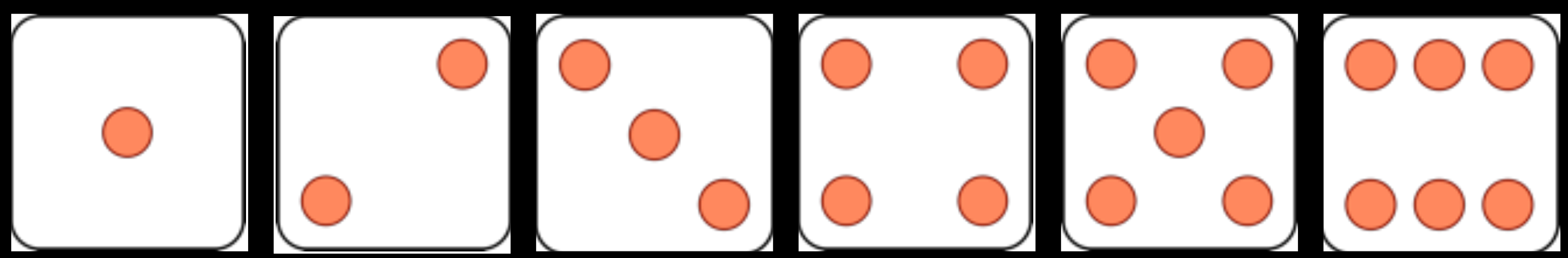
- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for ***learning*** and ***prediction***.
- Allows us to answer questions:
    - What is the best explanation of observed data?
    - Given some partially observed data (e.g. an input sentence), what is the most likely complete data (e.g. a sentence pair)?

Why Probability?

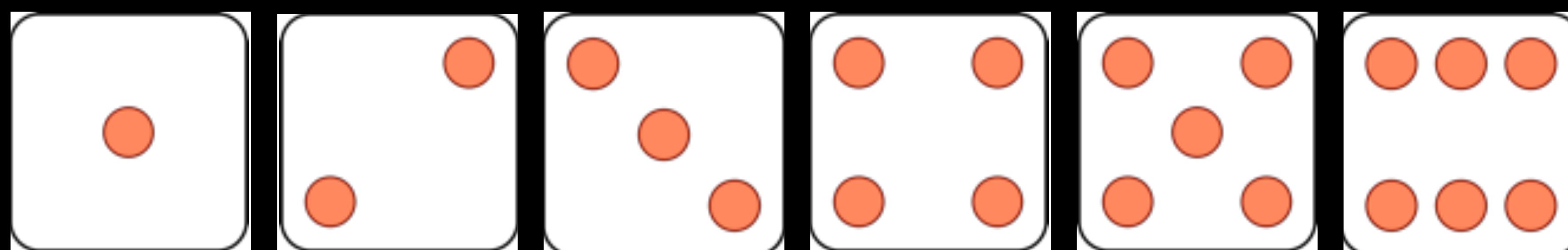
- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for ***learning*** and ***prediction***.
- Allows us to answer questions:
  - What is the best explanation of observed data?
  - Given some partially observed data (e.g. an input sentence), what is the most likely complete data (e.g. a sentence pair)?
- Common sense in mathematical form!

Probabilistic Primer

Probabilistic Primer

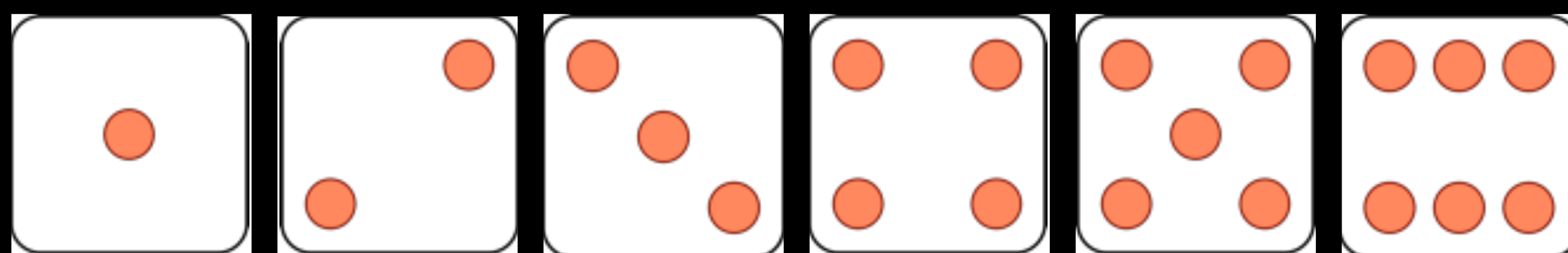


Probabilistic Primer



$\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

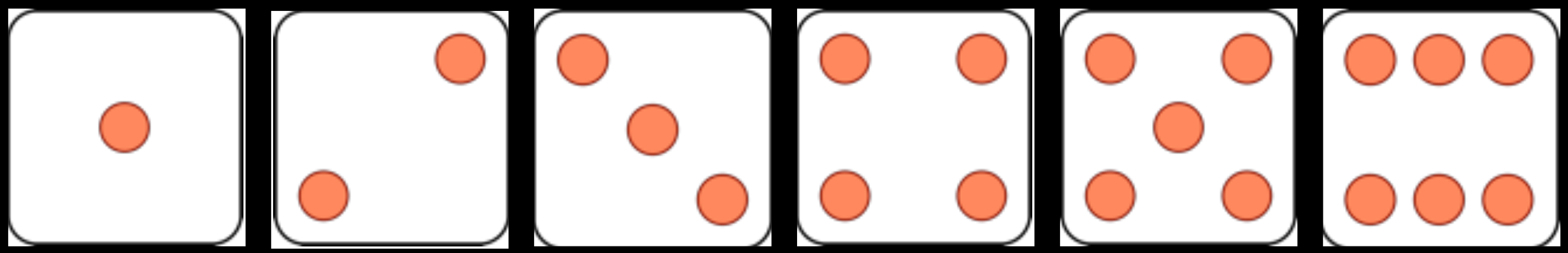
Probabilistic Primer



$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

The probabilities of all possible events must sum to 1.

Probabilistic Primer

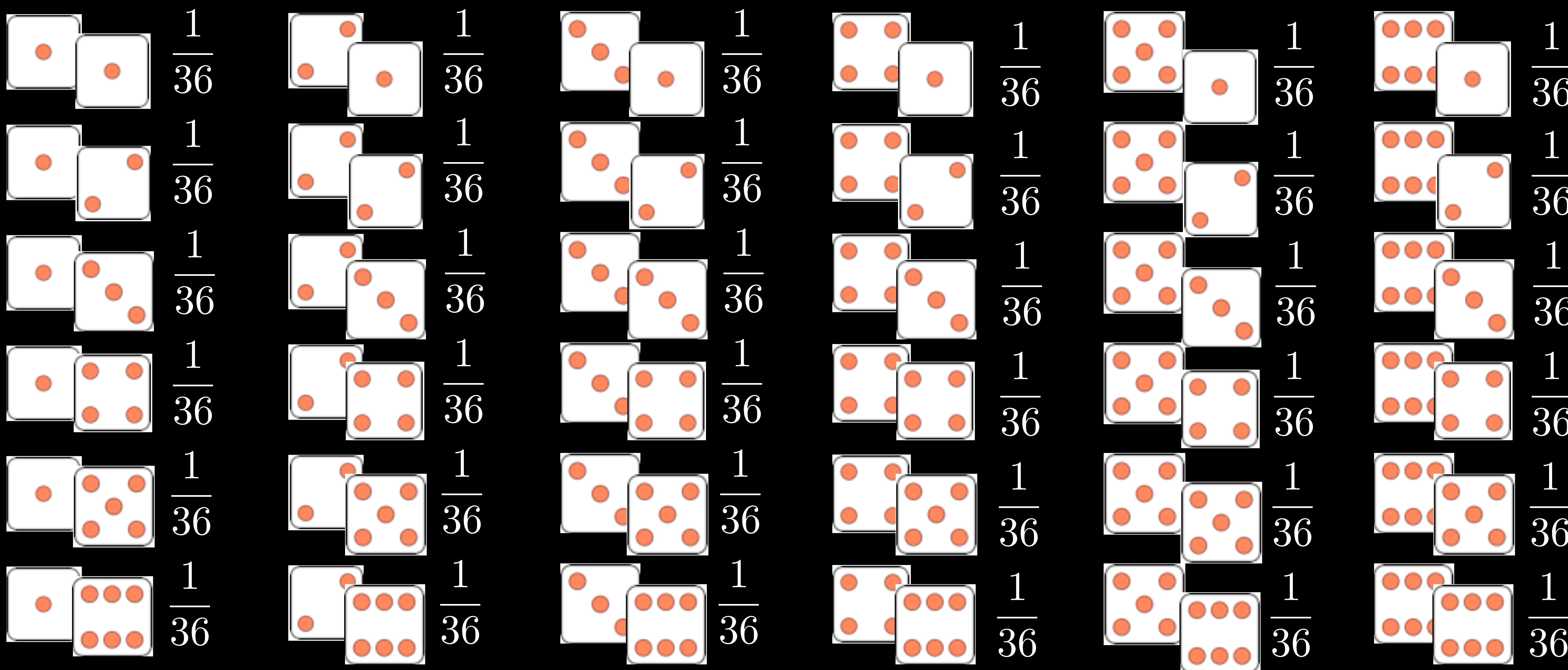


$\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

The probabilities of all possible events must sum to 1.

$$\sum_{e \in E} p(e) = 1$$

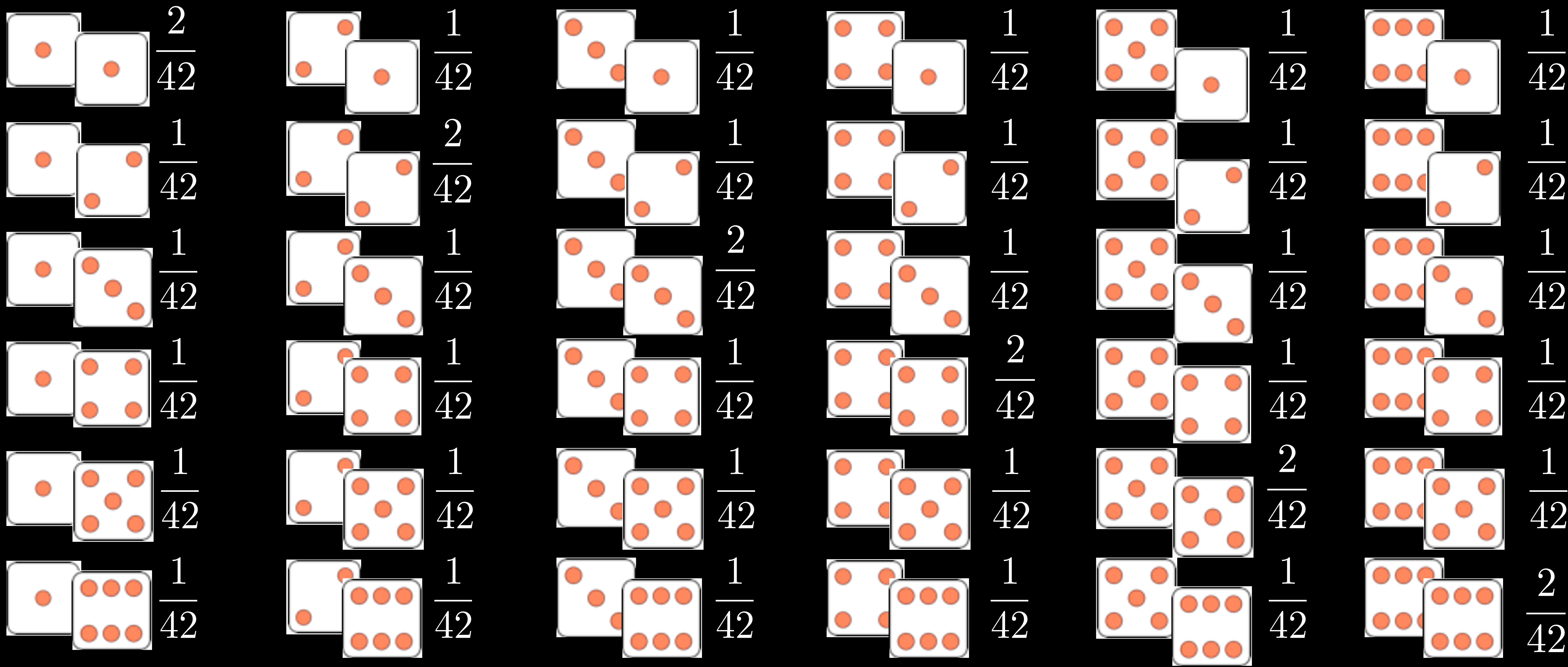
Probabilistic Primer



The probabilities of all possible events must sum to 1.

$$\sum_{e \in E} p(e) = 1$$

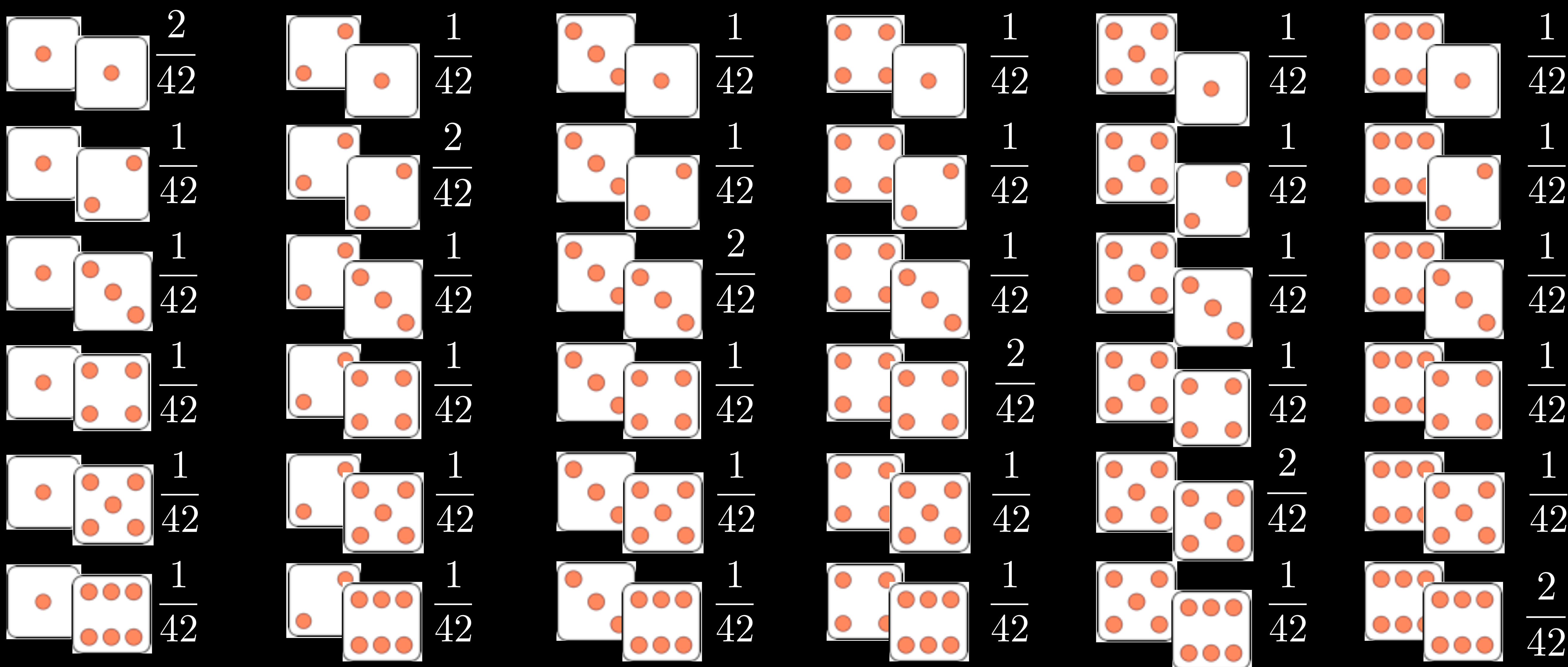
Probabilistic Primer



The probabilities of all possible events must sum to 1.

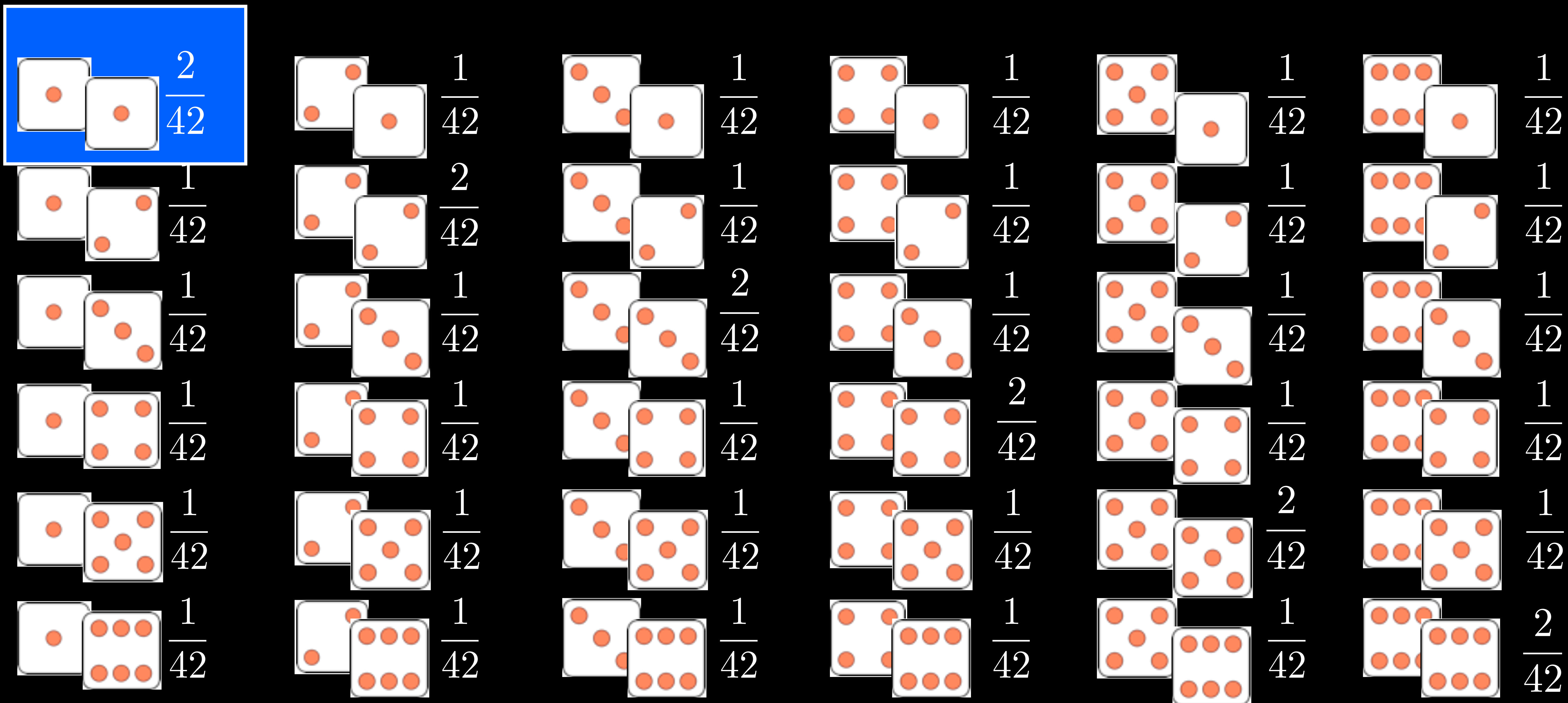
$$\sum_{e \in E} p(e) = 1$$

Probabilistic Primer



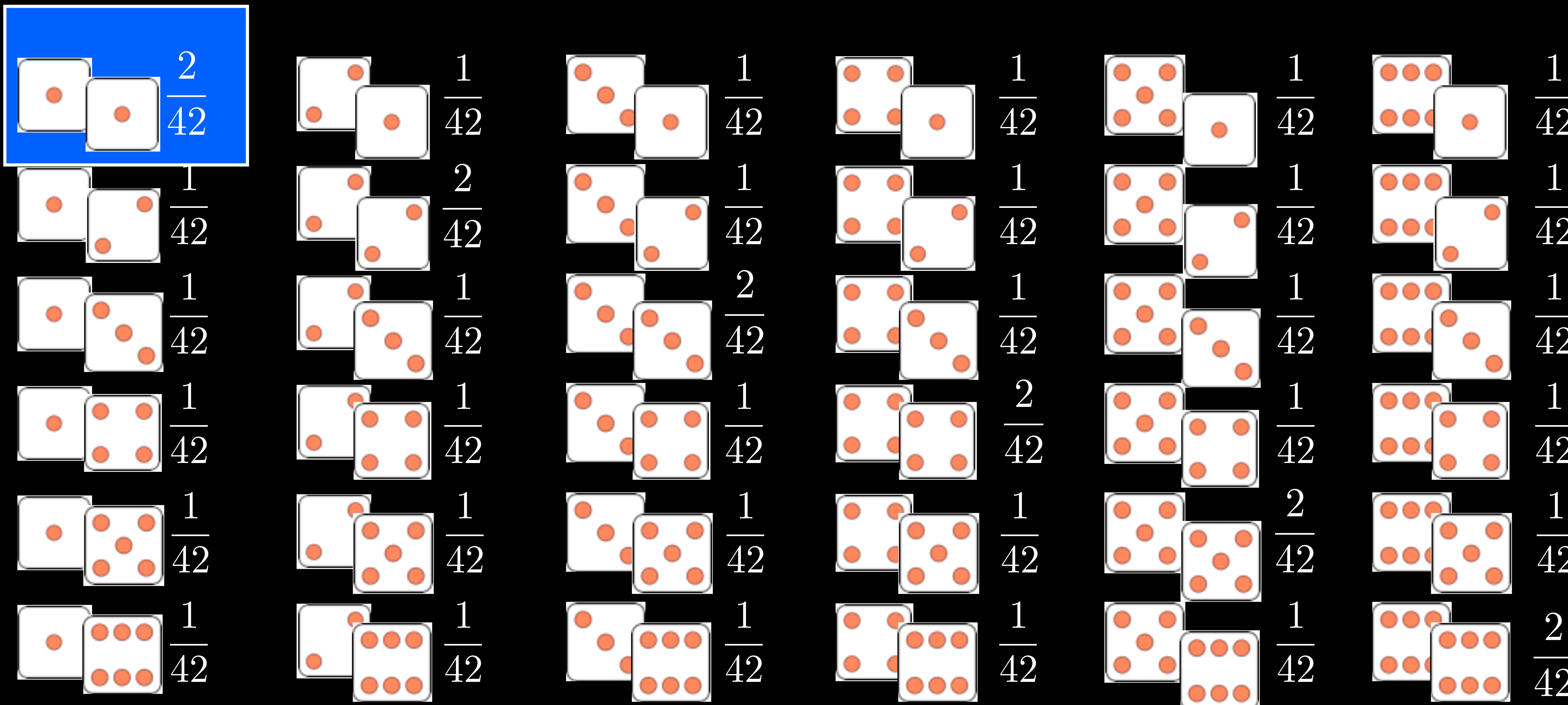
When an event consists of observations about more than one variable, it is a ***joint probability***.

Probabilistic Primer



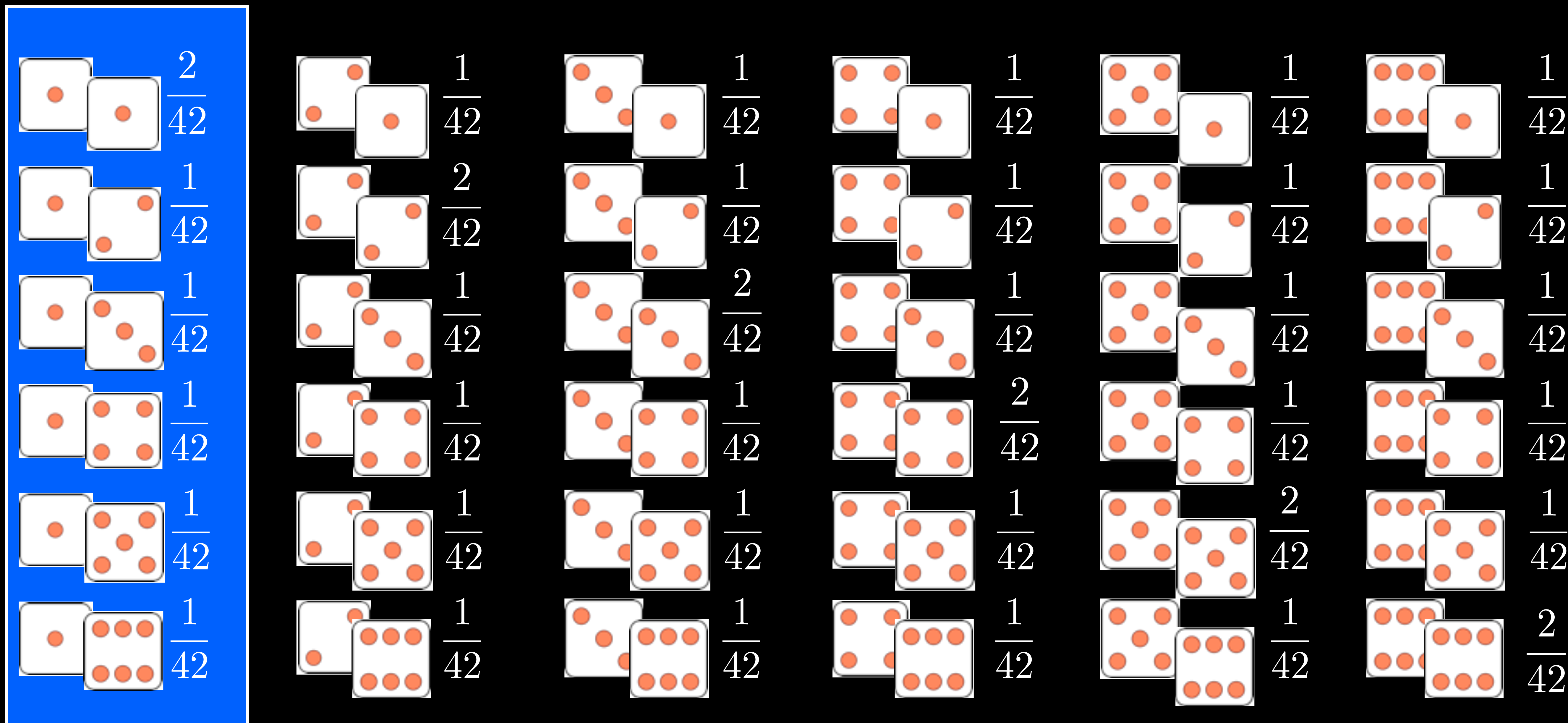
$$p(A = 1, B = 1) = \frac{2}{42}$$

Probabilistic Primer



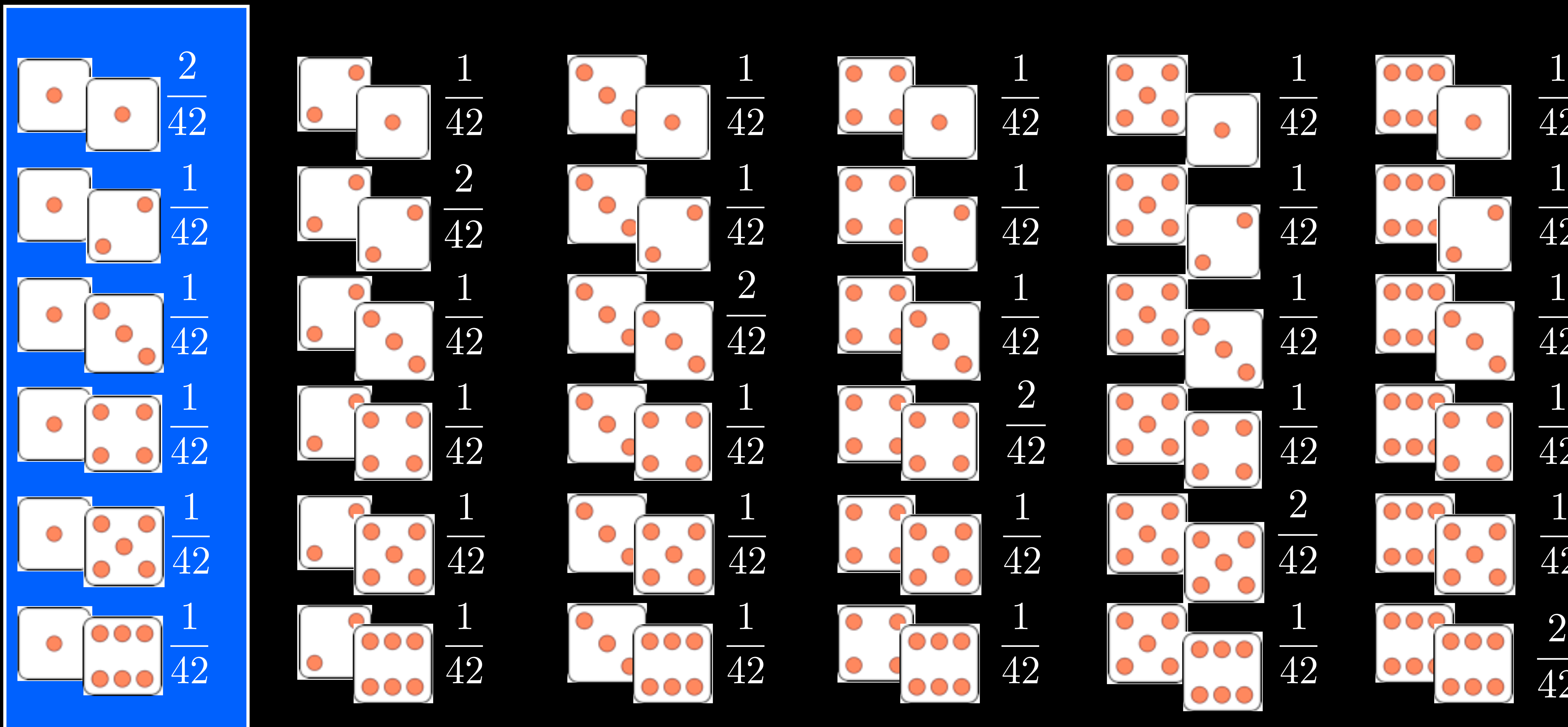
$$p(1,1) = \frac{2}{42}$$

Probabilistic Primer



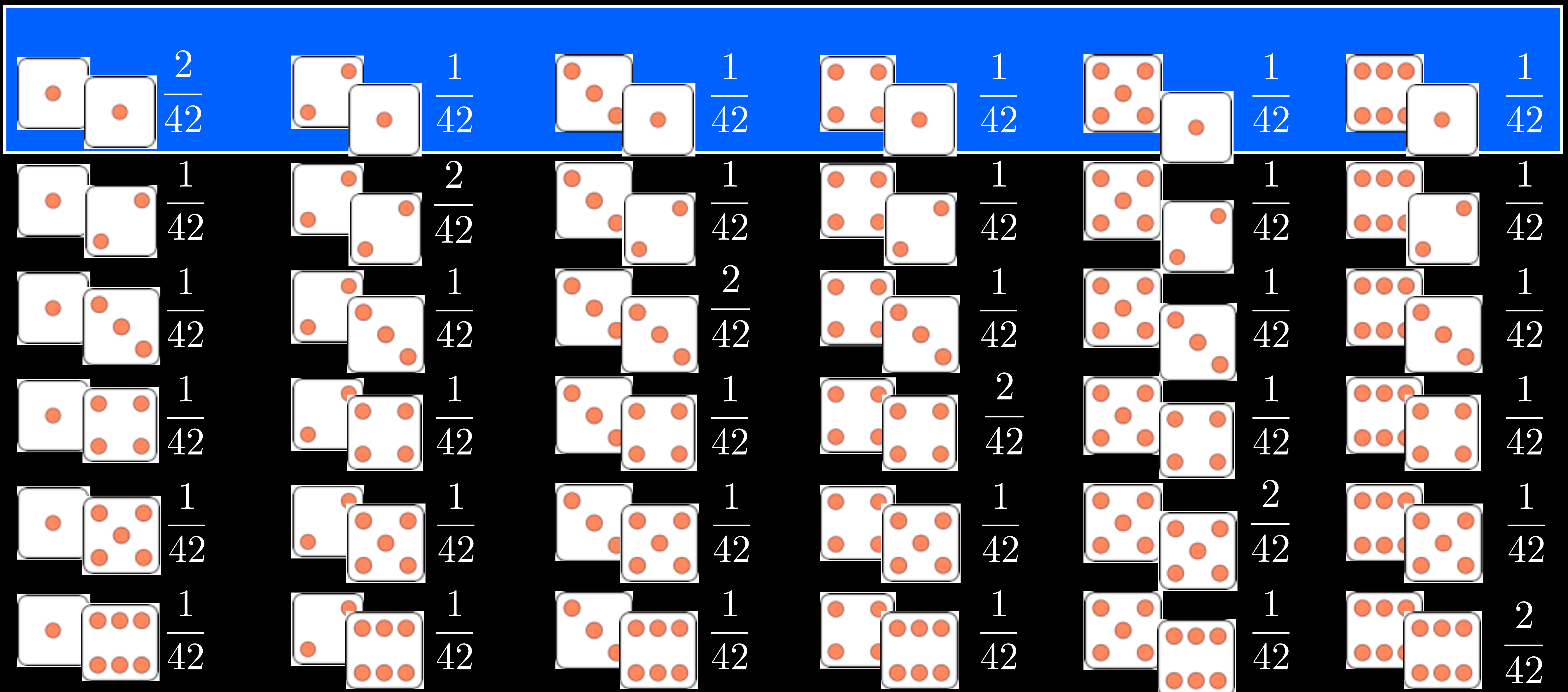
A probability distribution over a subset of variables is a *marginal probability*.

Probabilistic Primer



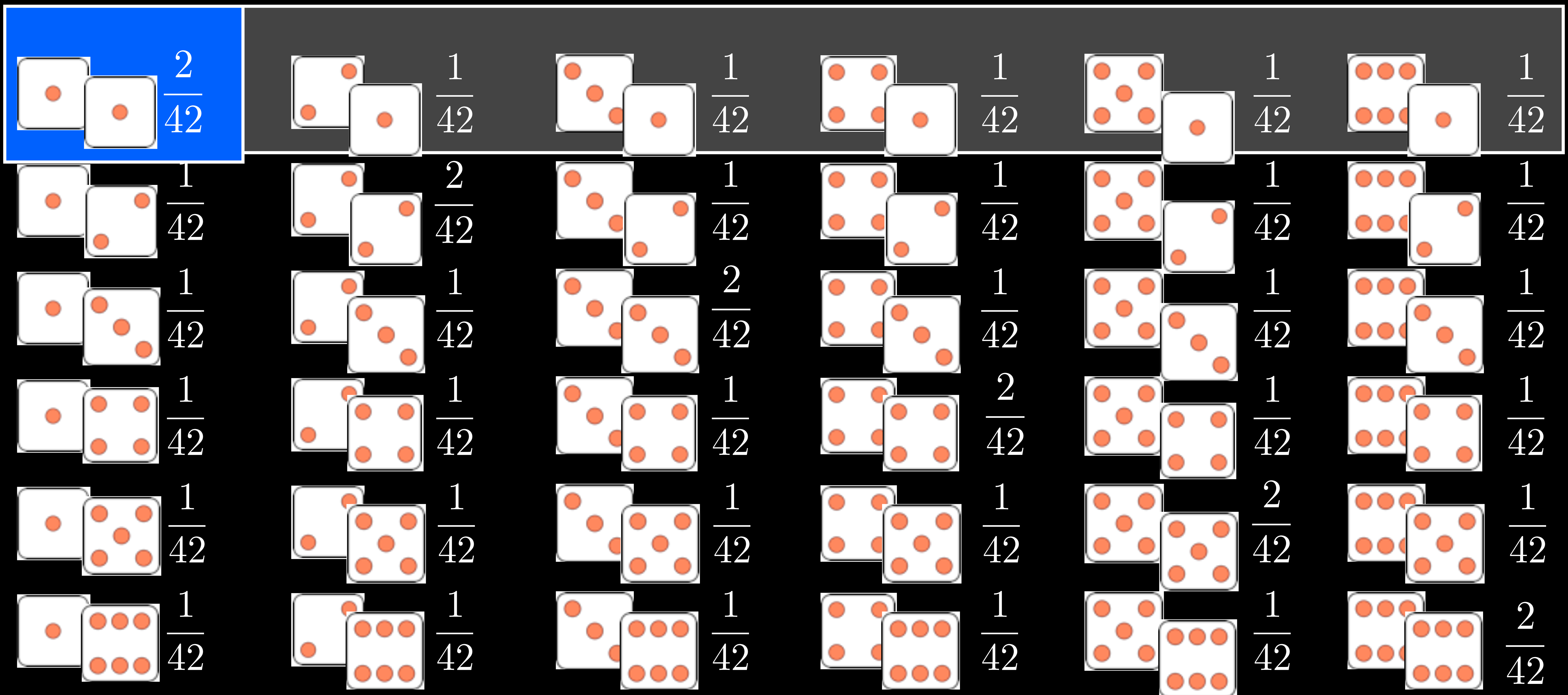
$$p(B=1) = p(\cdot, 1) = \sum_{a \in A} p(A=a, B=1) = \frac{1}{6}$$

Probabilistic Primer



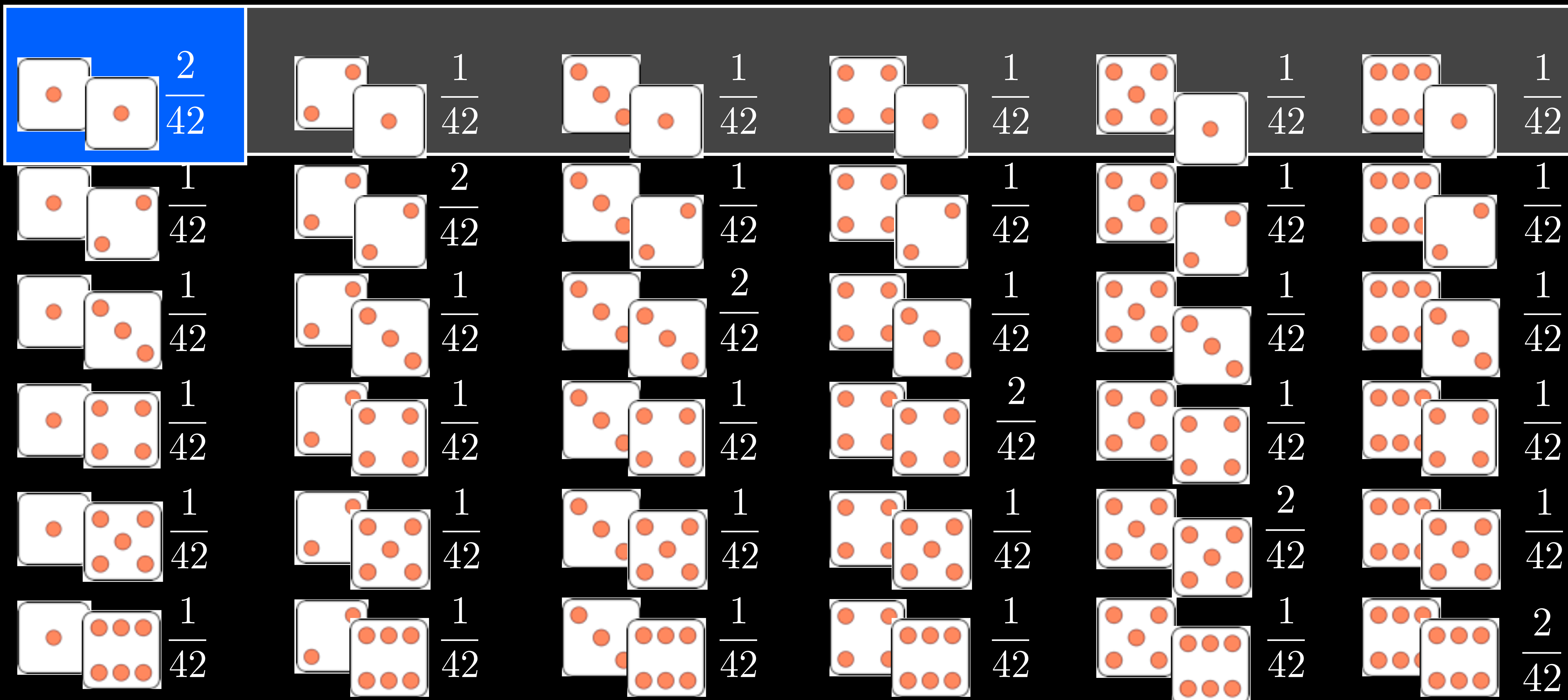
$$p(A=1) = p(1,\cdot) = \sum_{b \in B} p(A=1, B=b) = \frac{1}{6}$$

Probabilistic Primer



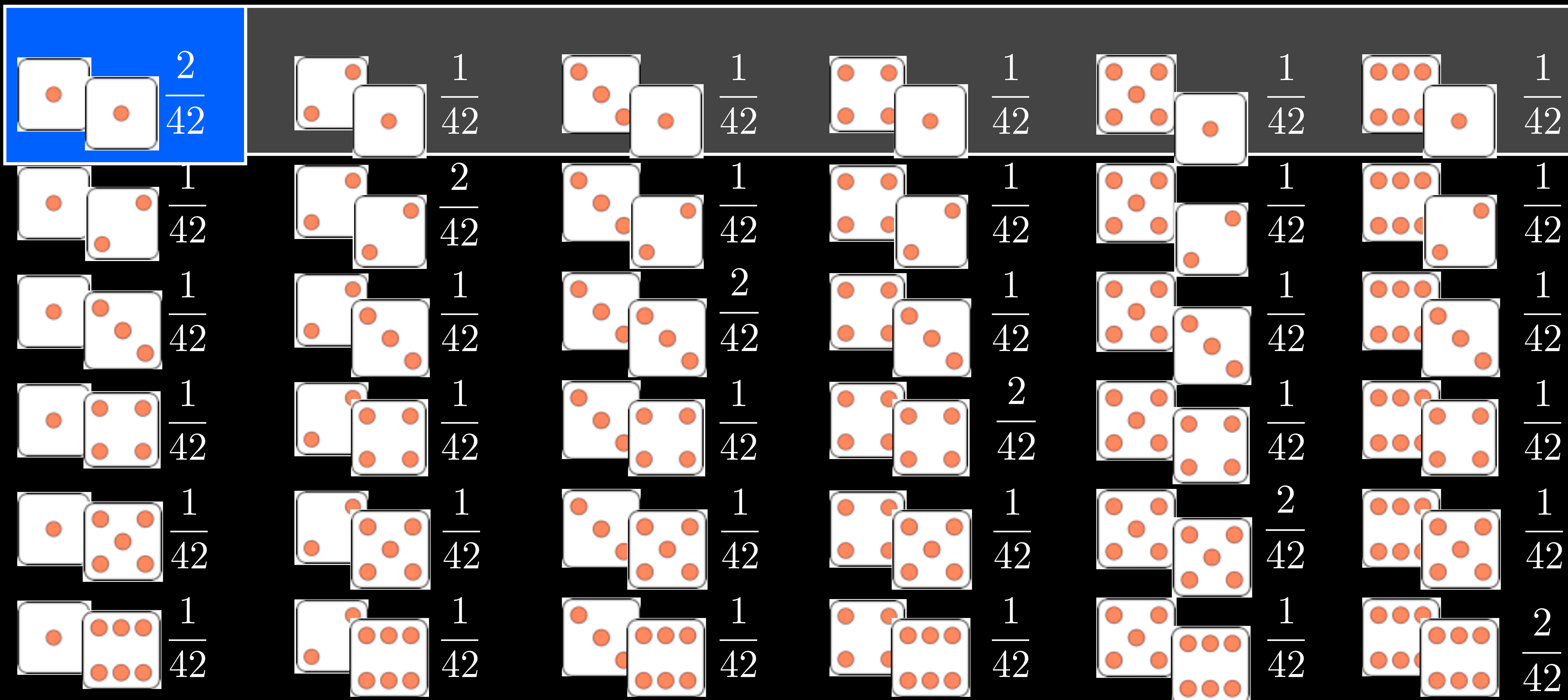
The probability of a variable under the condition that the other variables are fixed is the ***conditional probability***.

Probabilistic Primer



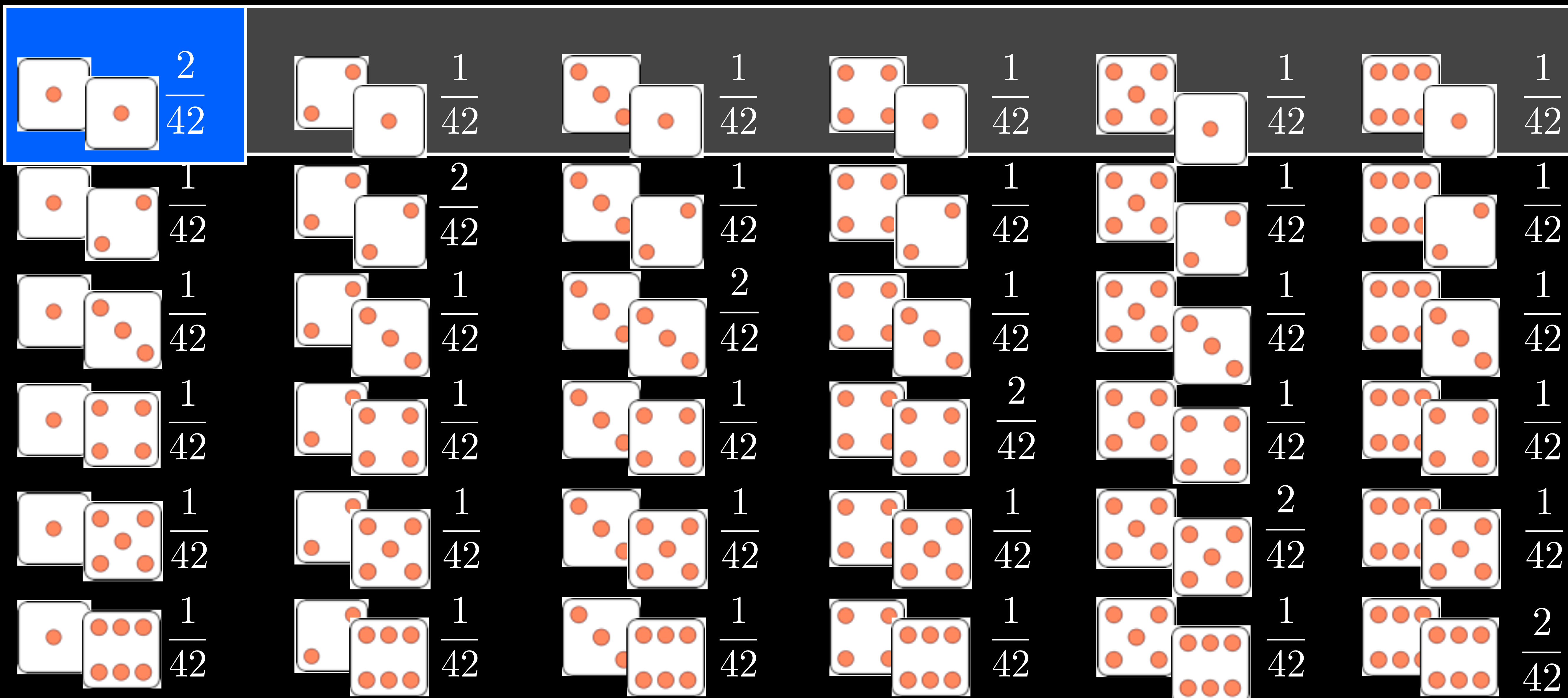
$$p(B=1|A=1) = \frac{p(A=1,B=1)}{\sum_{b\in B} p(A=1,B=b)} = \frac{2}{7}$$

Probabilistic Primer



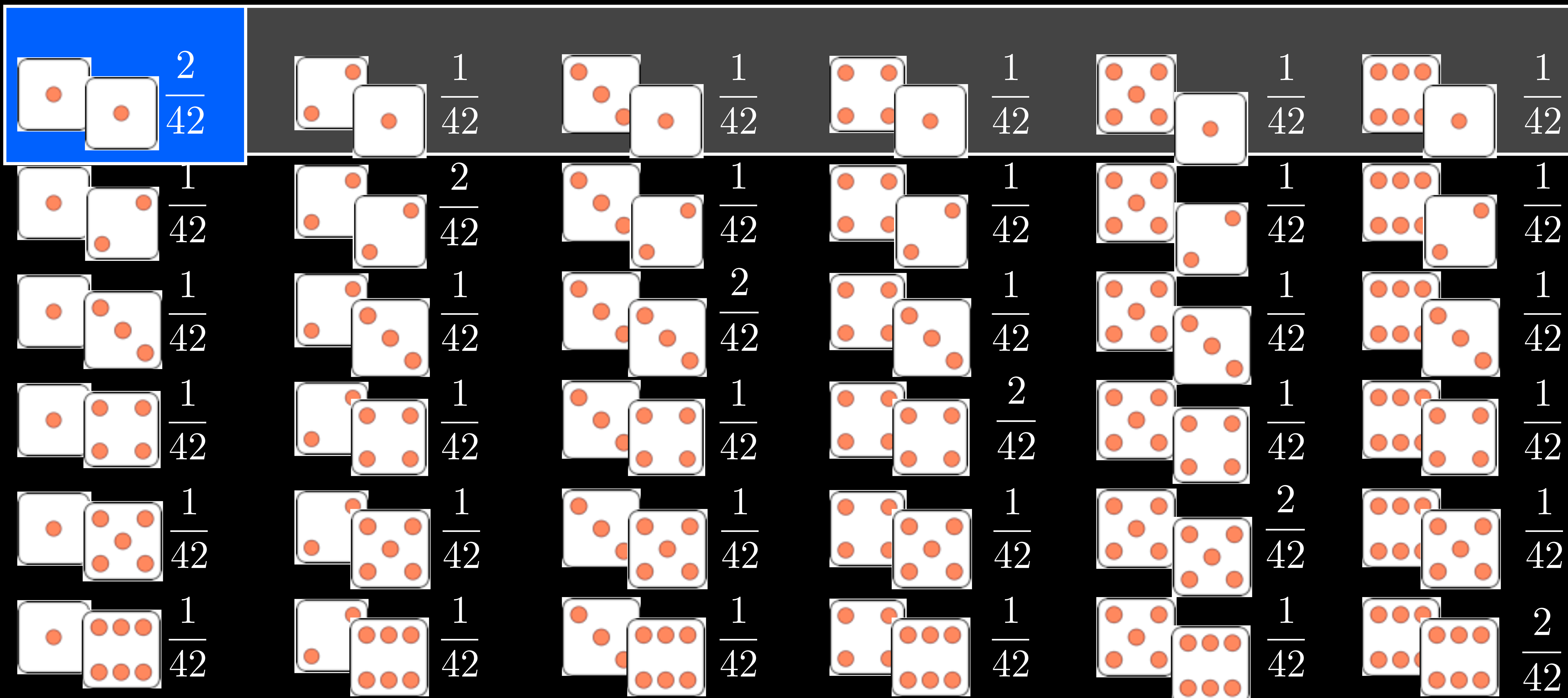
$$p(B=1|A=1) = \frac{p(A=1,B=1)}{\sum_{b\in B} p(A=1,B=b)} = \frac{2}{7} \qquad \frac{\text{joint}}{\text{marginal}}$$

Probabilistic Primer



A variable is ***conditionally independent*** of another iff its marginal probability = its conditional probability

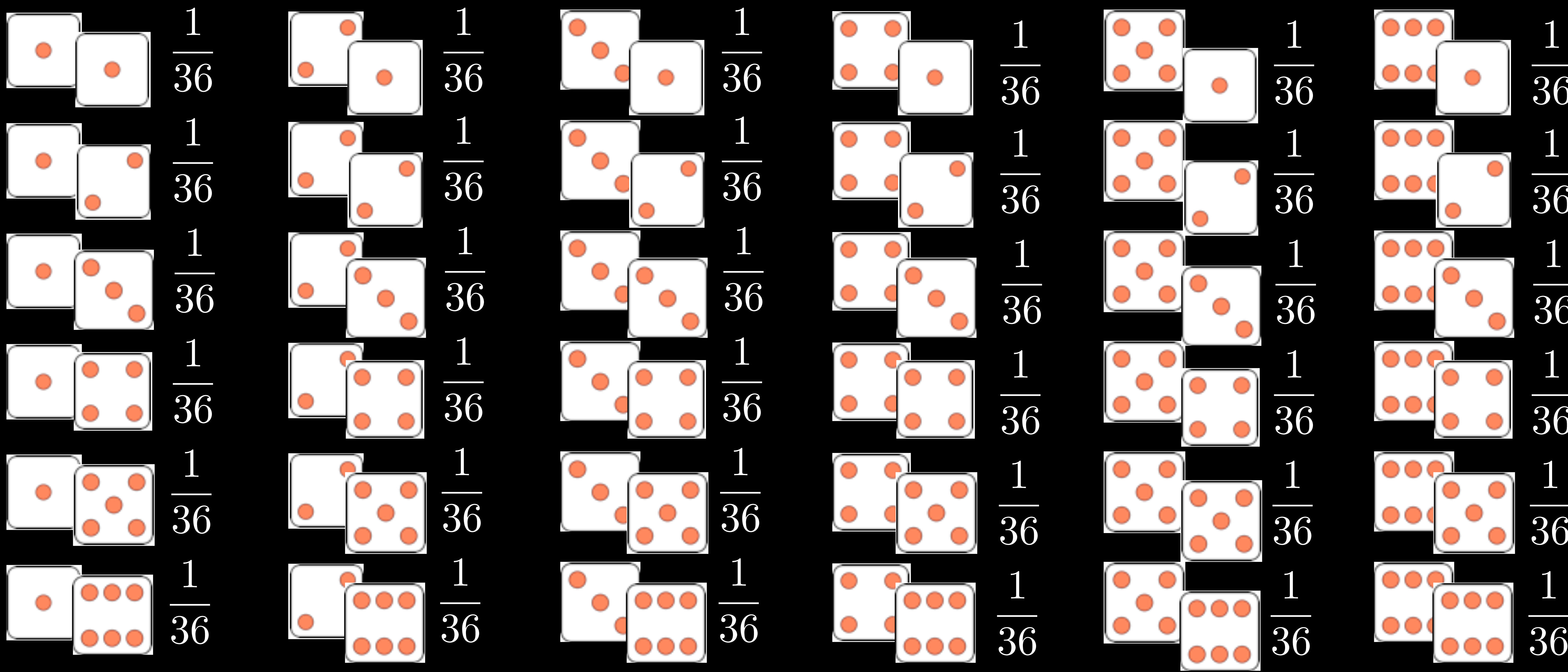
Probabilistic Primer



Under this distribution, B is ***not*** conditionally independent of A!

$$p(B = 1 | A = 1) = \frac{2}{7} \neq \frac{1}{6} = p(B = 1)$$

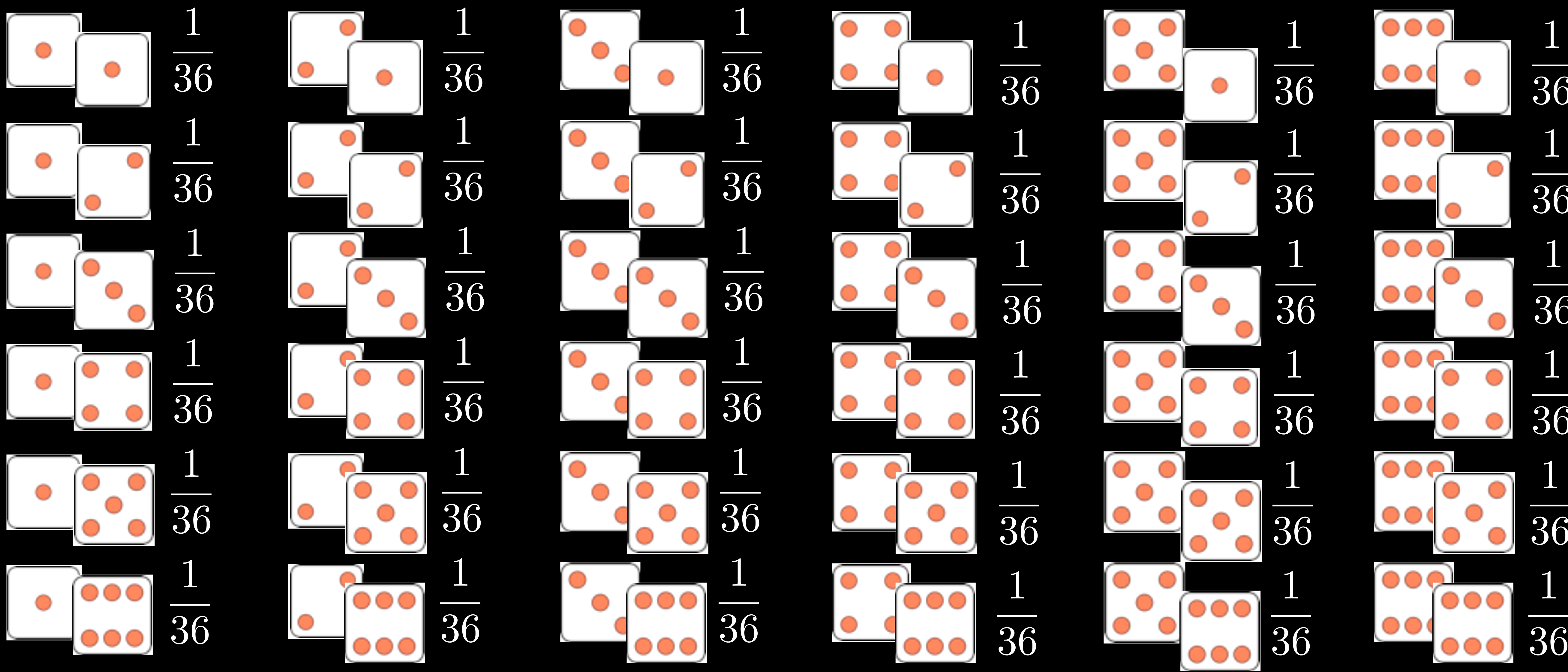
Probabilistic Primer



Knowing value of A does not change distribution over B.

$$p(B = 1 | A = 1) = \frac{1}{6} = \frac{1}{6} = p(B = 1)$$

Probabilistic Primer

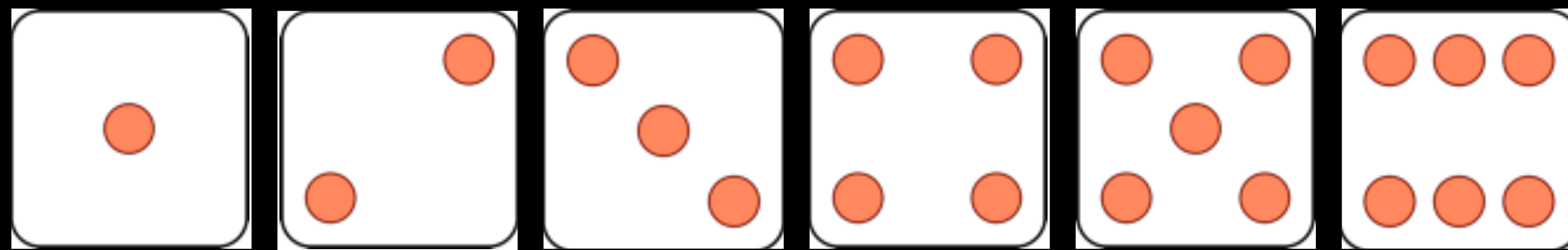


Under this distribution, B is independent of A.

$$p(B = 1|A = 1) = \frac{1}{6} = \frac{1}{6} = p(B = 1)$$

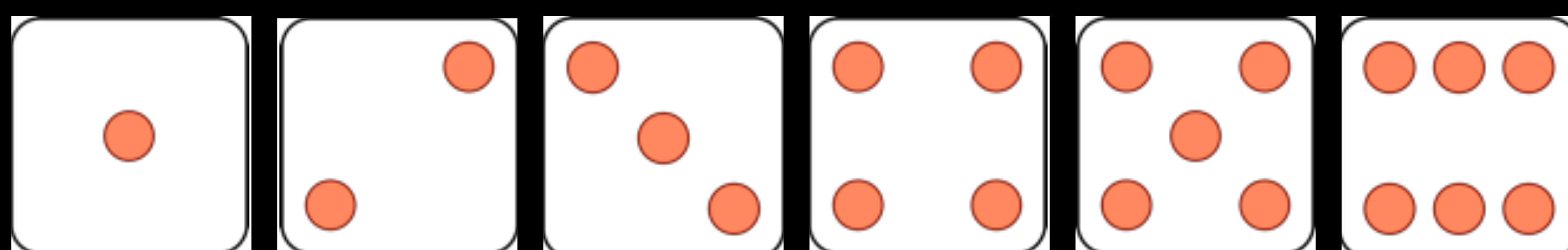
Probabilistic Primer

$p(A)$



$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

$p(B)$

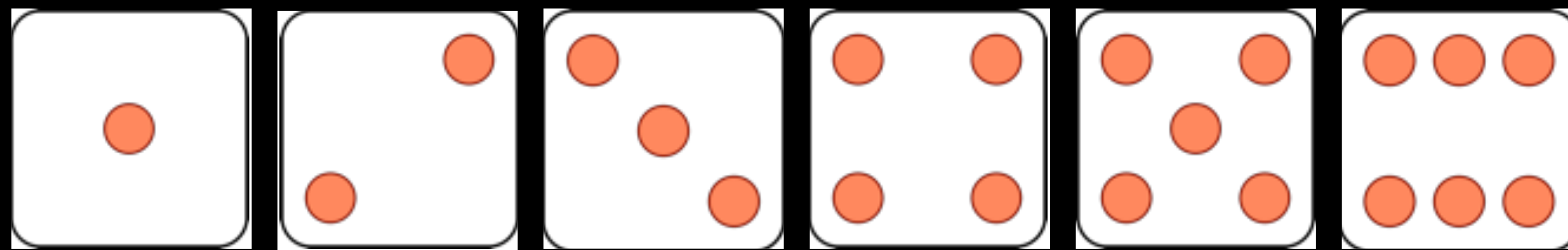


$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

Conditional independence means that the distributions that characterize your model are simpler.

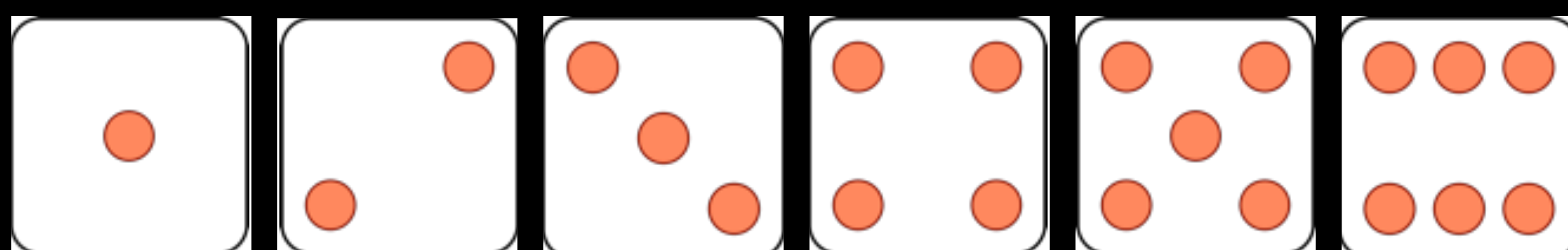
Probabilistic Primer

$p(A)$



$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

$p(B)$

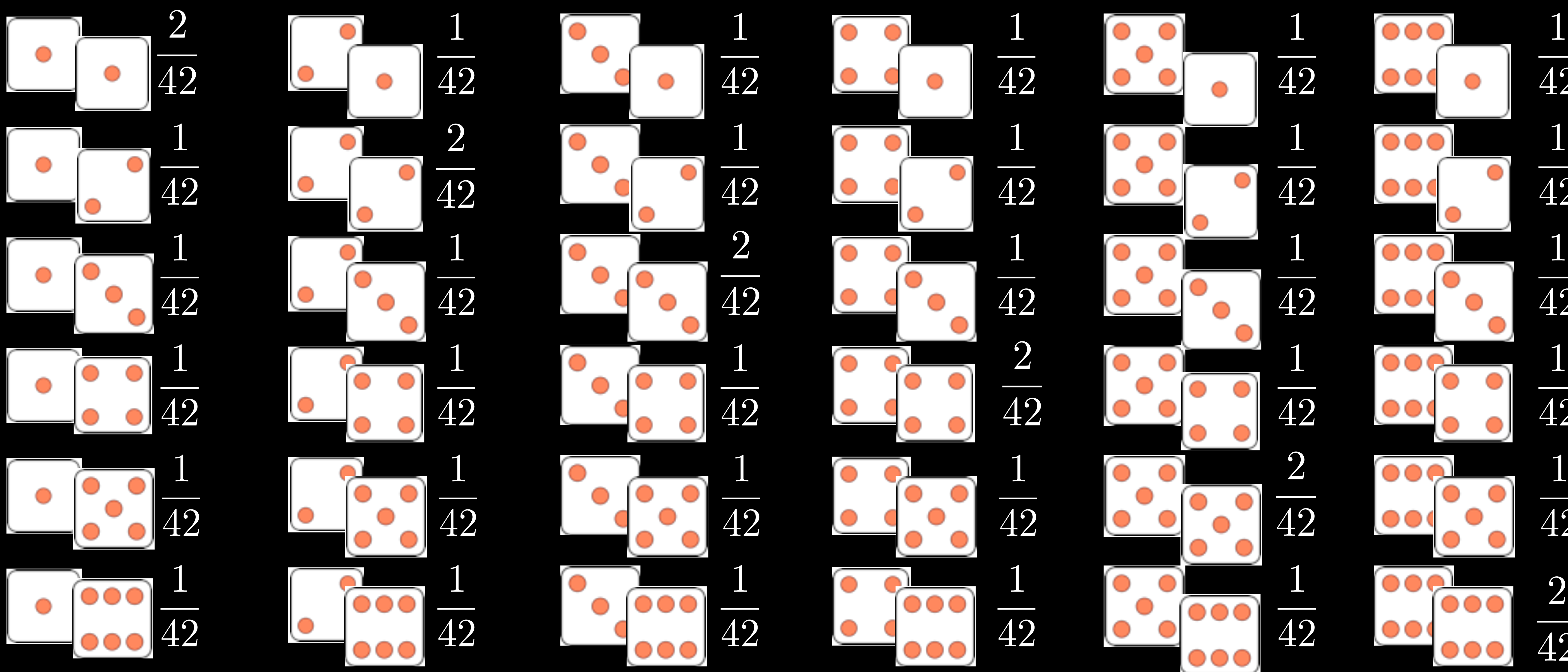


$$\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

It is easy to obtain the joint probability.

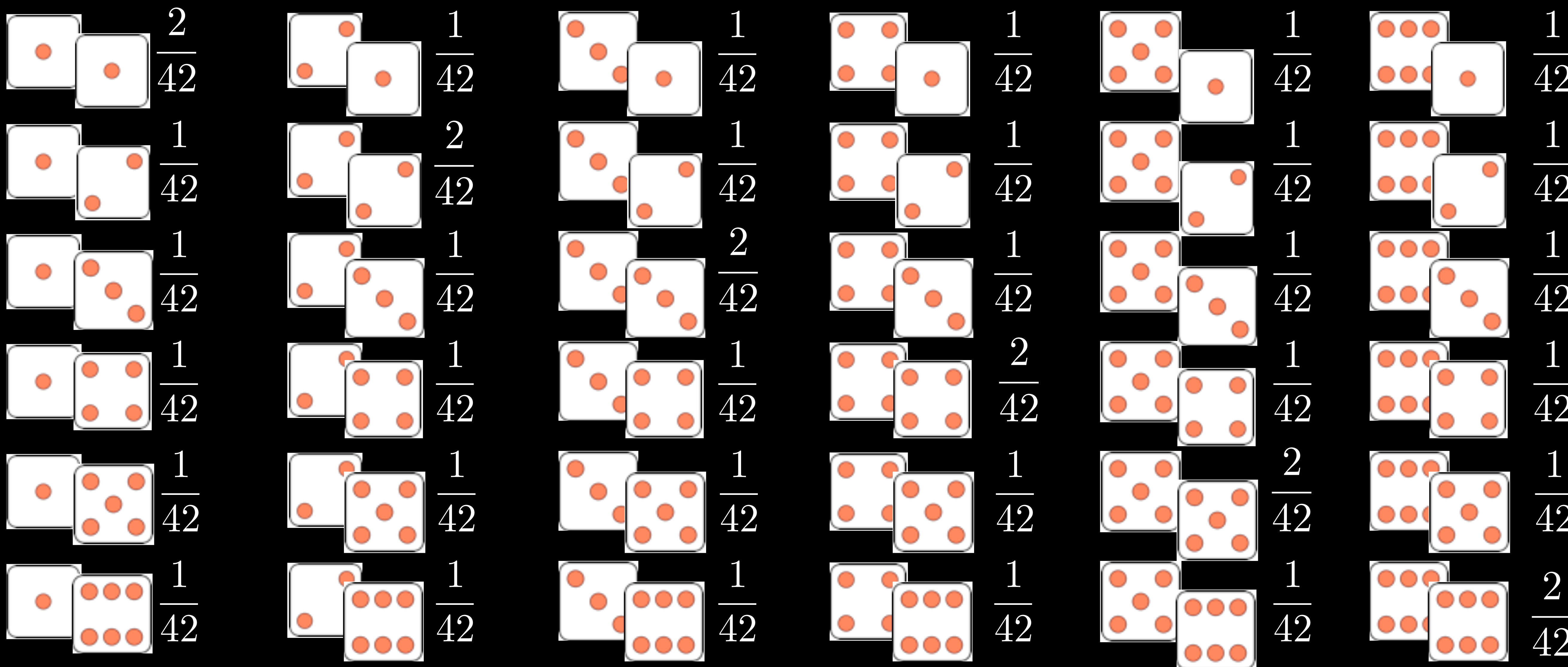
$$p(A = 1, B = 1) = p(A = 1) \cdot p(B = 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Probabilistic Primer



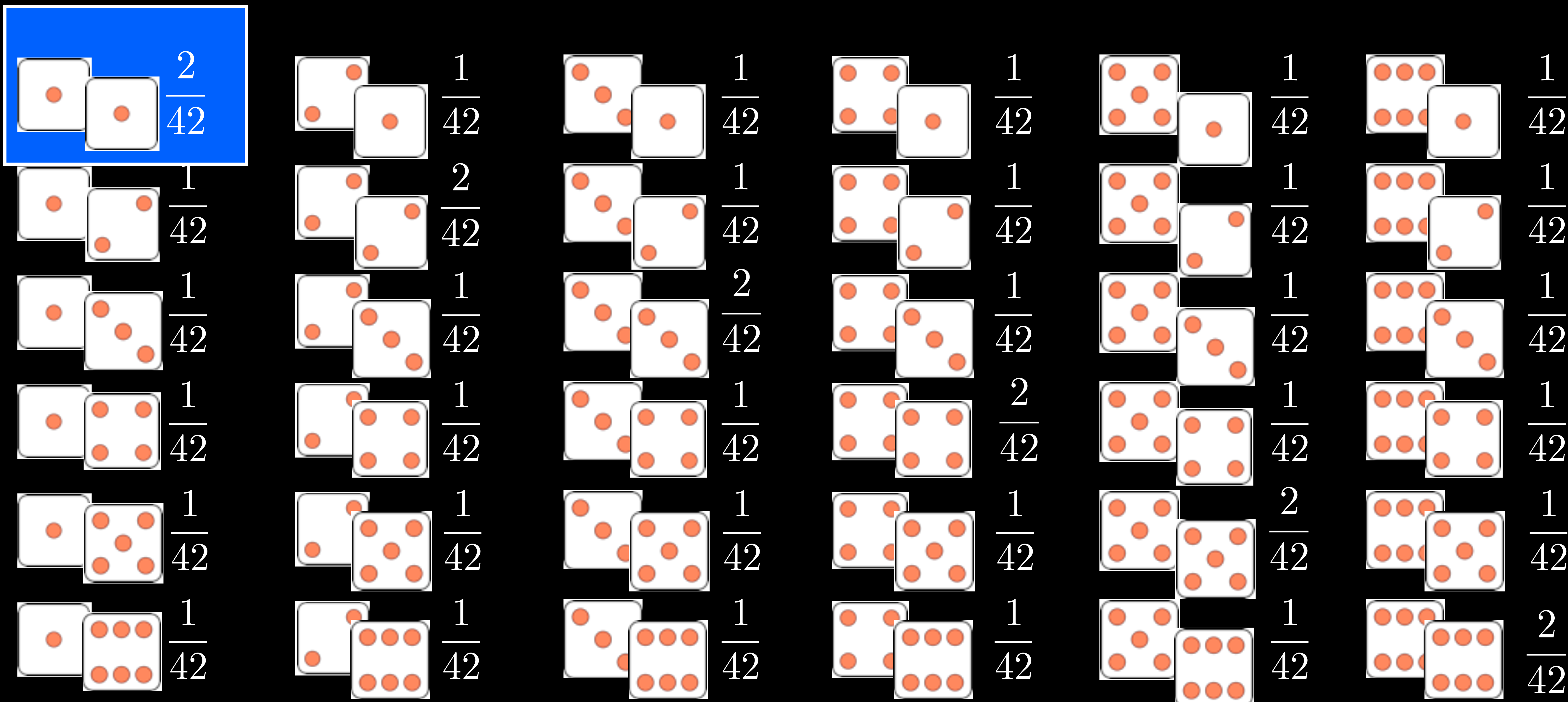
Caveat: if your data are not conditionally independent, the model will be a poor fit!

Probabilistic Primer



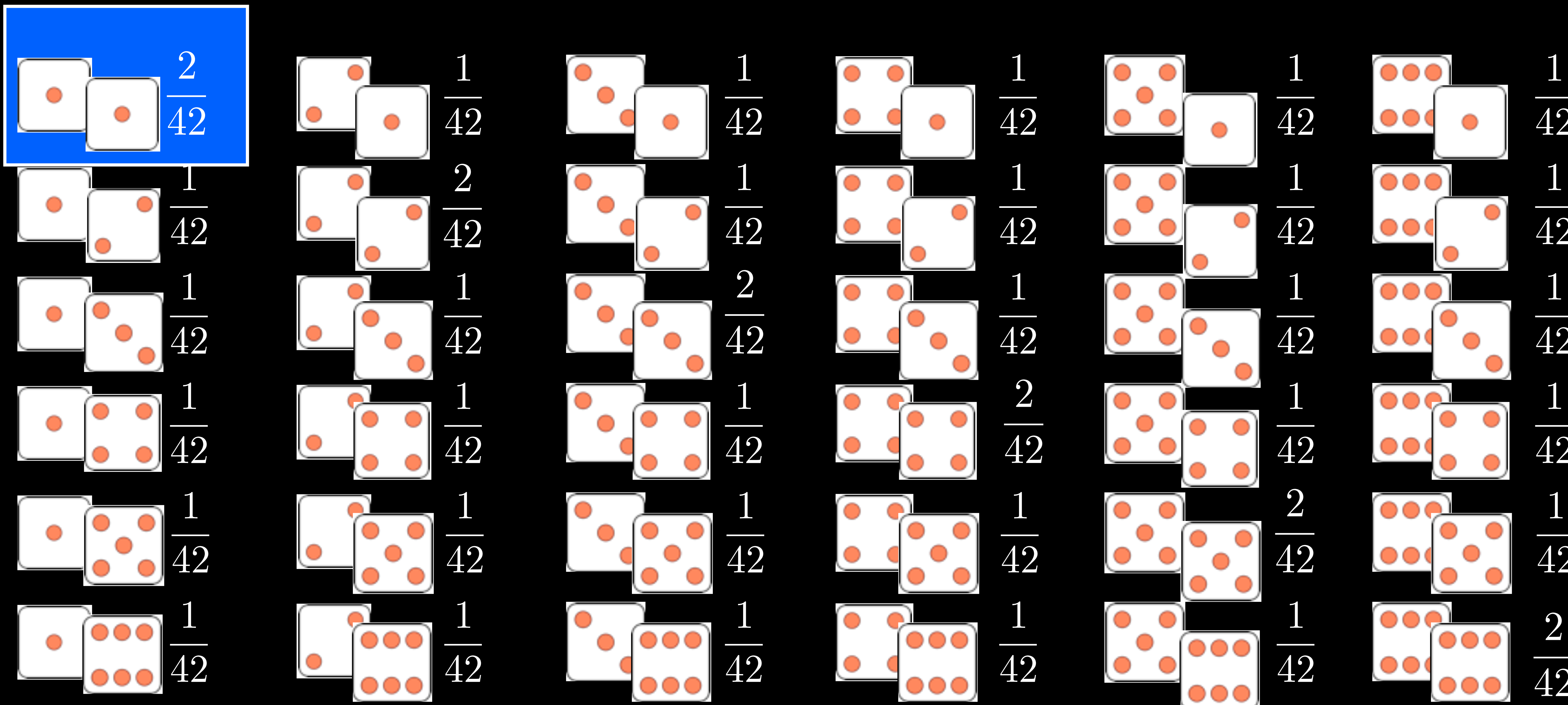
We can still represent the joint distribution as a product of other distributions.

Probabilistic Primer



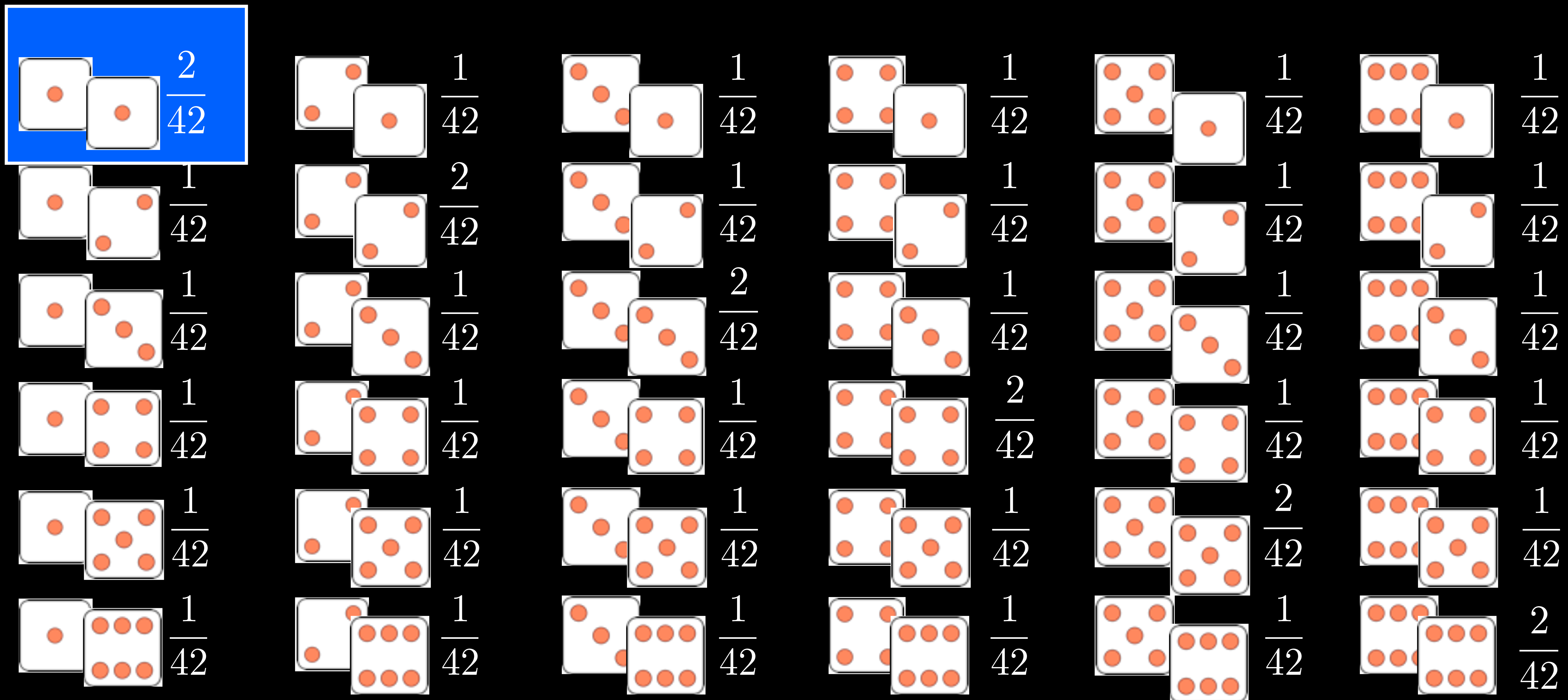
$$p(A = 1, B = 1) = p(A = 1, B = 1)$$

Probabilistic Primer



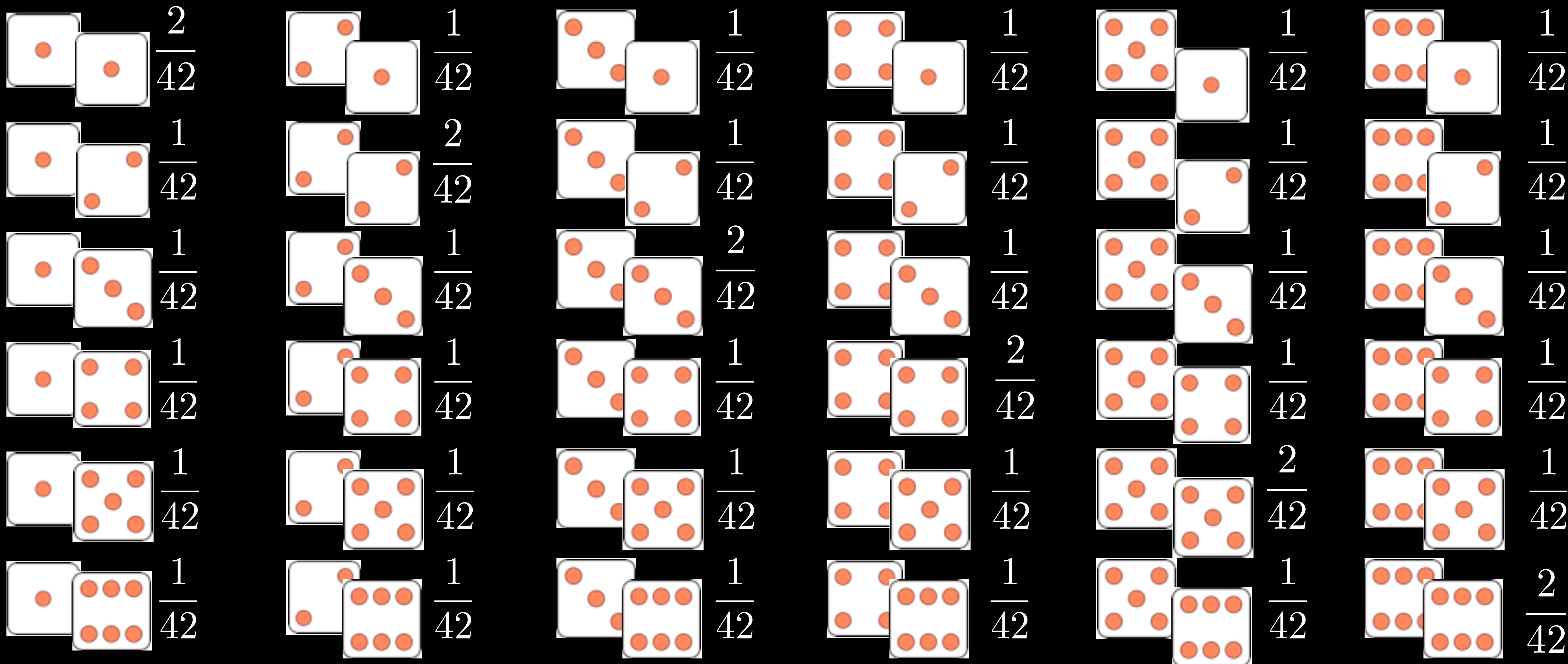
$$p(A=1,B=1)=\sum_{b\in B}p(A=1,B=b)\frac{p(A=1,B=1)}{\sum_{b\in B}p(A=1,B=b)}$$

Probabilistic Primer



$$p(A=1, B=1) = p(A=1) \cdot p(B=1|A=1)$$

Probabilistic Primer



$$p(A, B) = p(A) \cdot p(B|A)$$

Probabilistic Primer

$$p(A, B) = p(A) \cdot p(B|A)$$

Probabilistic Primer

$$p(A, B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

Probabilistic Primer

$$p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

Probabilistic Primer

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Probabilistic Primer

Bayes' Rule

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Probabilistic Primer

Bayes' Rule

posterior prior likelihood

$$\underbrace{p(B|A)}_{\text{posterior}} = \frac{\underbrace{p(B)}_{\text{prior}} \cdot \underbrace{p(A|B)}_{\text{likelihood}}}{p(A)}$$

...But the probability that an event has happened is the same as the probability I have to guess right if I guess it has happened. Wherefore the following proposition is evident: If there be two subsequent events, the probability of the 2d b/N and the probability both together P/N, and it being 1st discovered that the 2d event has also happened, the probability I am right is P/b.

Thomas Bayes

...But the probability that an event has happened is the same as the probability I have to guess right if I guess it has happened. Wherefore the following proposition is evident: If there be two subsequent events, the probability of the 2d b/N and the probability both together P/N, and it being 1st discovered that the 2d event has also happened, the probability I am right is P/b.

Thomas Bayes

(image by Chris Dyer)

Bayes' Rule

$$p(English)$$

Bayes' Rule

$$p(English)$$

configuration

Bayes' Rule

$$p(English)$$

configuration

$$p(image|English)$$

Bayes' Rule

$$p(English)$$

configuration

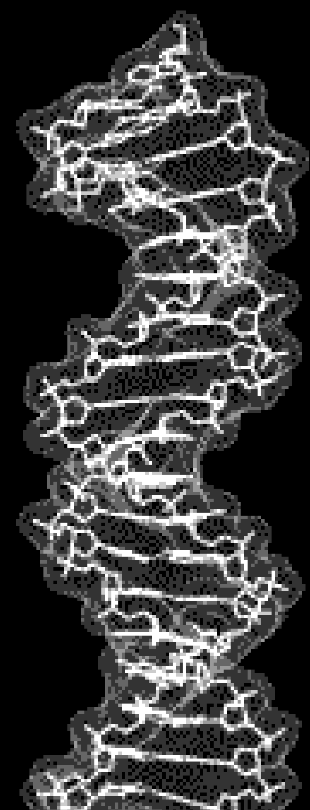
$$p(image|English)$$

Bayes' Rule

$p(DNA)$

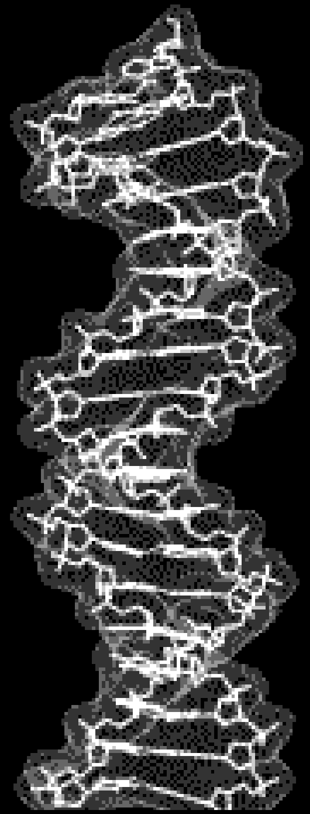
Bayes' Rule

$p(DNA)$



Bayes' Rule

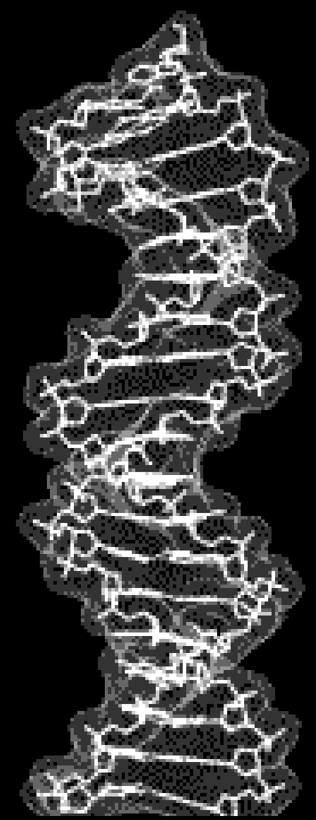
$p(DNA)$



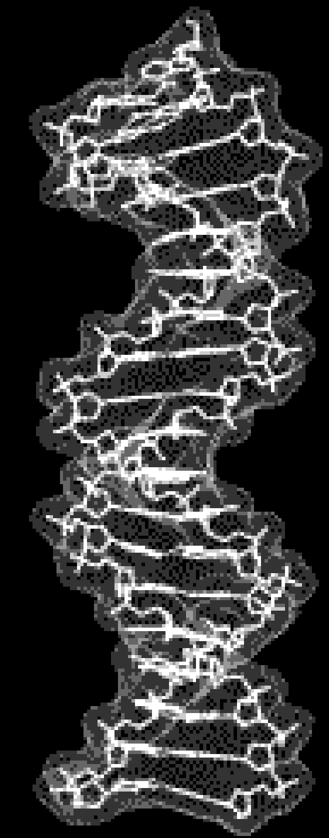
$p(mutation|DNA)$

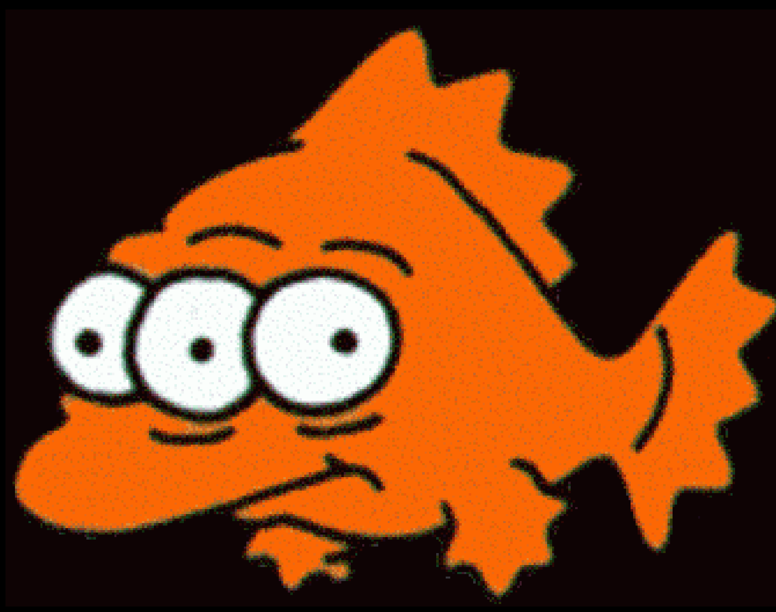
Bayes' Rule

$p(DNA)$



$p(mutation|DNA)$





Bayes' Rule

$$p(English)$$

Bayes' Rule

$$p(English)$$

However, the sky remained clear under the strong north wind .

Bayes' Rule

$$p(English)$$

However, the sky remained clear under the strong north wind .

$$p(Chinese|English)$$

Bayes' Rule

$$p(English)$$

However, the sky remained clear under the strong north wind .

$$p(Chinese|English)$$

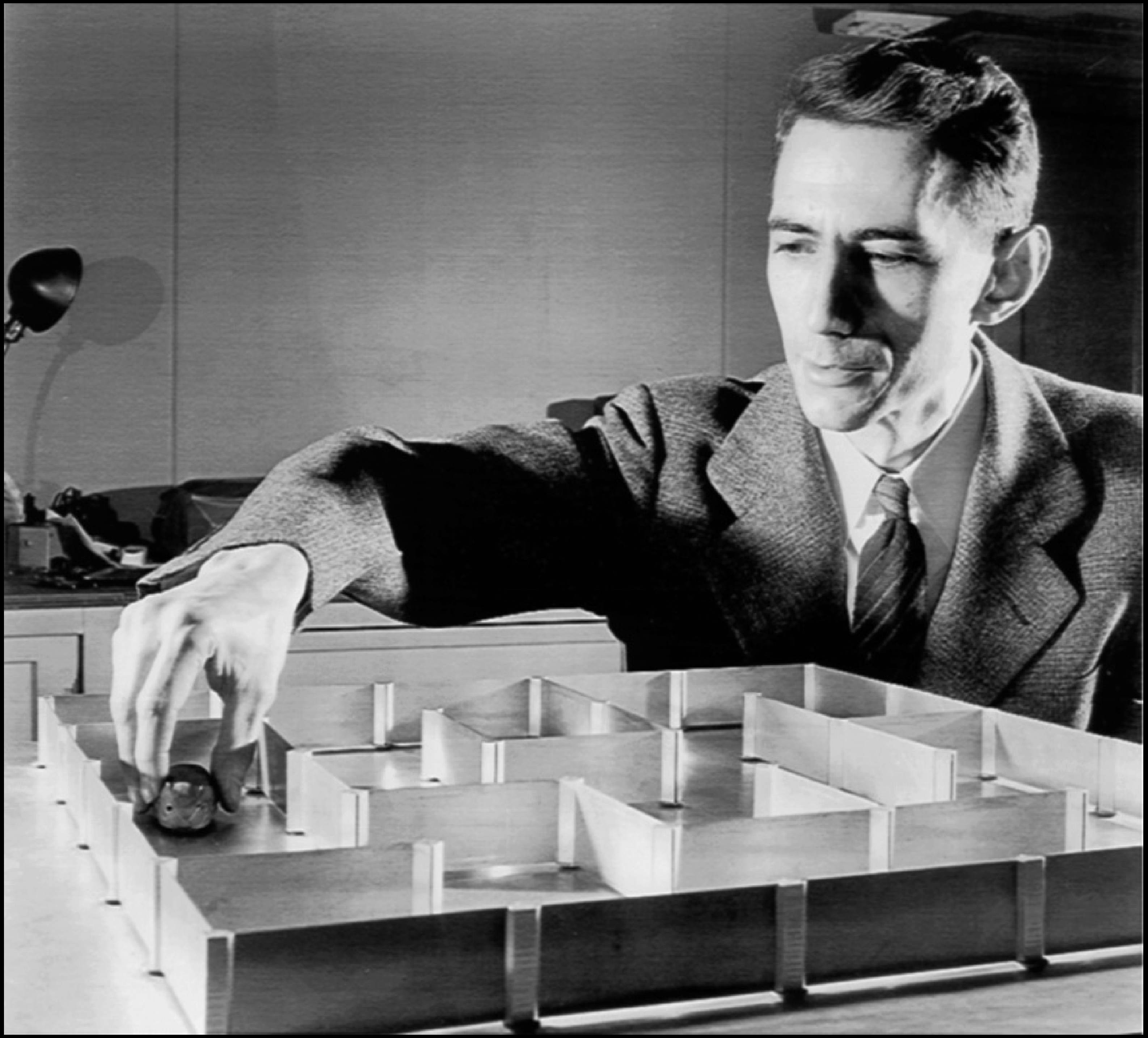
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Warren Weaver (1949)

THE MATHEMATICAL THEORY OF COMMUNICATION

by Claude E. Shannon and Warren Weaver



Claude Shannon

Bayes' Rule

$$p(English|Chinese) =$$

$$\frac{\underbrace{p(English)}_{\text{prior}} \times \underbrace{p(Chinese|English)}_{\text{likelihood}}}{p(Chinese)}$$

prior

likelihood

normalization term (ensures we're working with valid probabilities).

Noisy Channel

$$p(English|Chinese) = \frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

signal model

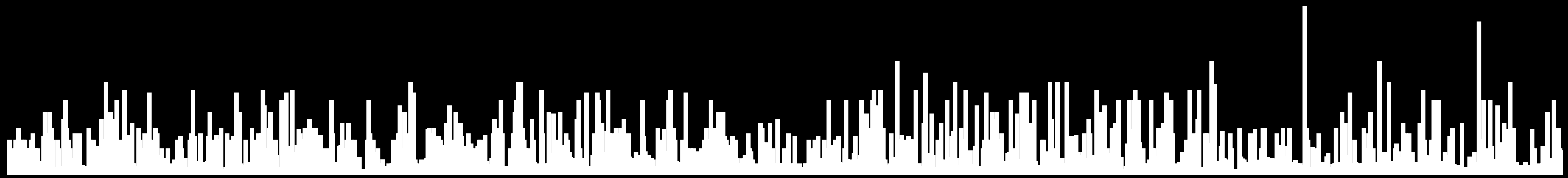
channel model

normalization term (ensures we're working with valid probabilities).

Machine Translation

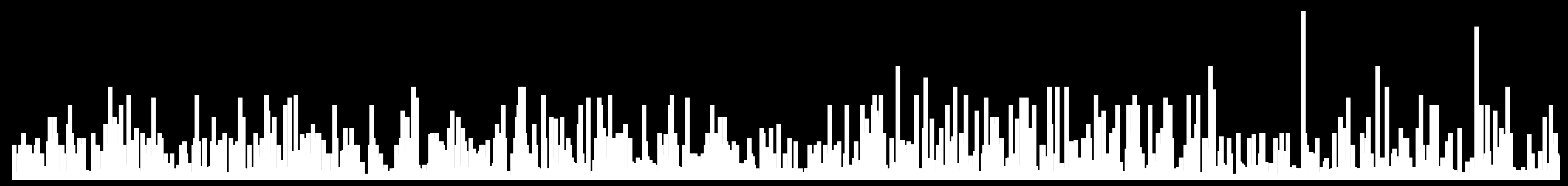
$$p(English|Chinese) = \frac{\underbrace{p(English)}_{\text{language model}} \times \underbrace{p(Chinese|English)}_{\text{translation model}}}{\underbrace{p(Chinese)}_{\text{normalization term (ensures we're working with valid probabilities).}}}$$

$p(Chinese|English)$

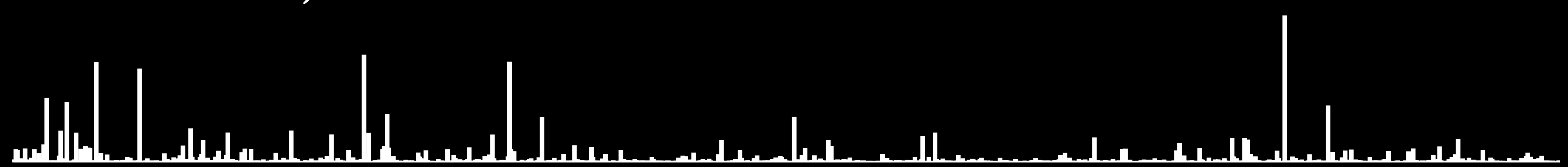


$English$

$$p(Chinese|English)$$

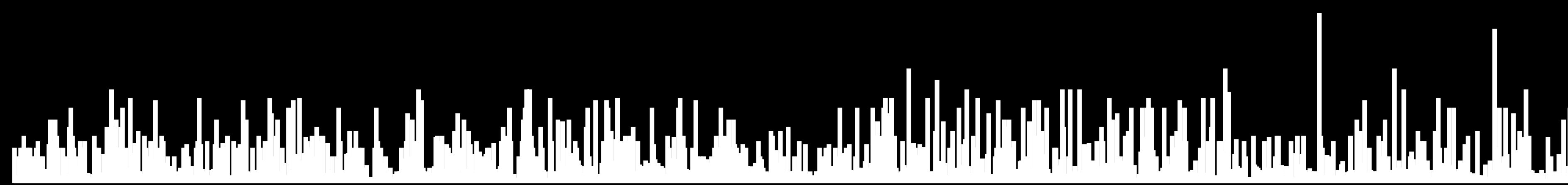


$$\times\ p(English)$$

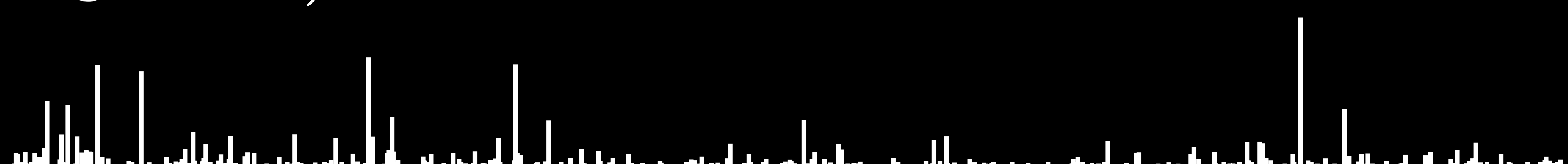


$English$

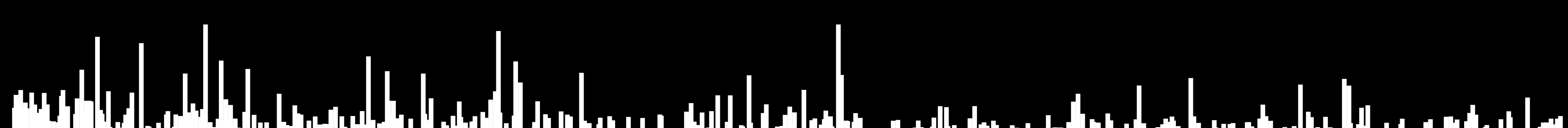
$$p(Chinese|English)$$



$$\times\ p(English)$$



$$\sim p(English|Chinese)$$



$English$

Machine Translation

$$p(English|Chinese) =$$

$$\frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

language model

translation model

normalization term

(remember: probabilities must sum to 1).

Machine Translation

$$p(English|Chinese) \sim$$

$$p(English) \times p(Chinese|English)$$

Machine Translation

$$p(English|Chinese) \sim$$

$$p(English) \times p(Chinese|English)$$

What is the probability of an English sentence?

Machine Translation

$$p(English|Chinese) \sim$$

$$p(English) \times p(Chinese|English)$$

What is the probability of an English sentence?

What is the probability of a Chinese sentence, given a particular English sentence?

Language Models

Our language model must assign a probability to *every possible English sentence*.

Language Models

Our language model must assign a probability to *every possible English sentence*.

Q: What should this model look like?

Language Models

Our language model must assign a probability to *every possible English sentence*.

Q: What should this model look like?

A: What is the dumbest thing you can think of?

Language Models

Every sequence of English words receives a non-zero probability.

Language Models

Every sequence of English words receives a non-zero probability.

Problem 1: there are an infinite number of such sequences.

Language Models

Every sequence of English words receives a non-zero probability.

Problem 1: there are an infinite number of such sequences.

Problem 2: it would be hard to estimate.

Language Models

Every sequenc[illegible]English [illegible]s receives a non-[illegible]pro[illegible]ty.

Problem 1: there are [illegible]nite number of such

Problem 2: [illegible]uld b[illegible] to estimate.

Language Models

Idea: since the language model is a joint model over all words in a sentence, make words depend on words earlier in the sentence.

Language Models

$$p(However|START)$$

Language Models

$$p(However|START)$$

A number between 0 and 1.

Language Models

$$p(However|START)$$

A number between 0 and 1.

$$\sum_x p(x|START) = 1$$

Language Models

However

$$p(However|START)$$

Language Models

However ,

$$p(,|However)$$

Language Models

However , the

$$p(the|,)$$

Language Models

However , the sky

$$p(sky|the)$$

Language Models

However , the sky remained

$$p(remained|sky)$$

Language Models

However , the sky remained clear

$$p(clear|remained)$$

Language Models

However , the sky remained clear ... wind .

... $p(STOP|.)$

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i | word_{i-1})$$

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i | word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i | word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

This model explains every word in the English sentence.

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i | word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

This model explains every word in the English sentence.

But it makes very strong conditional independence assumptions!

Language Models

Question: where do these numbers come from?

$$p(sky|the)$$

$$p(clear|remained)$$

$$p(remained|sky)$$

Language Models

This is just a model that we can train on data.

... in the night sky as it orbits earth ...

... said that the sky would fall if ...

... falling dollar , sky high interest rates ...

However , the sky remained clear ...

$$p(remained|sky) = ???$$


1998

5 KRONER
DANMARK

$$p(heads)$$

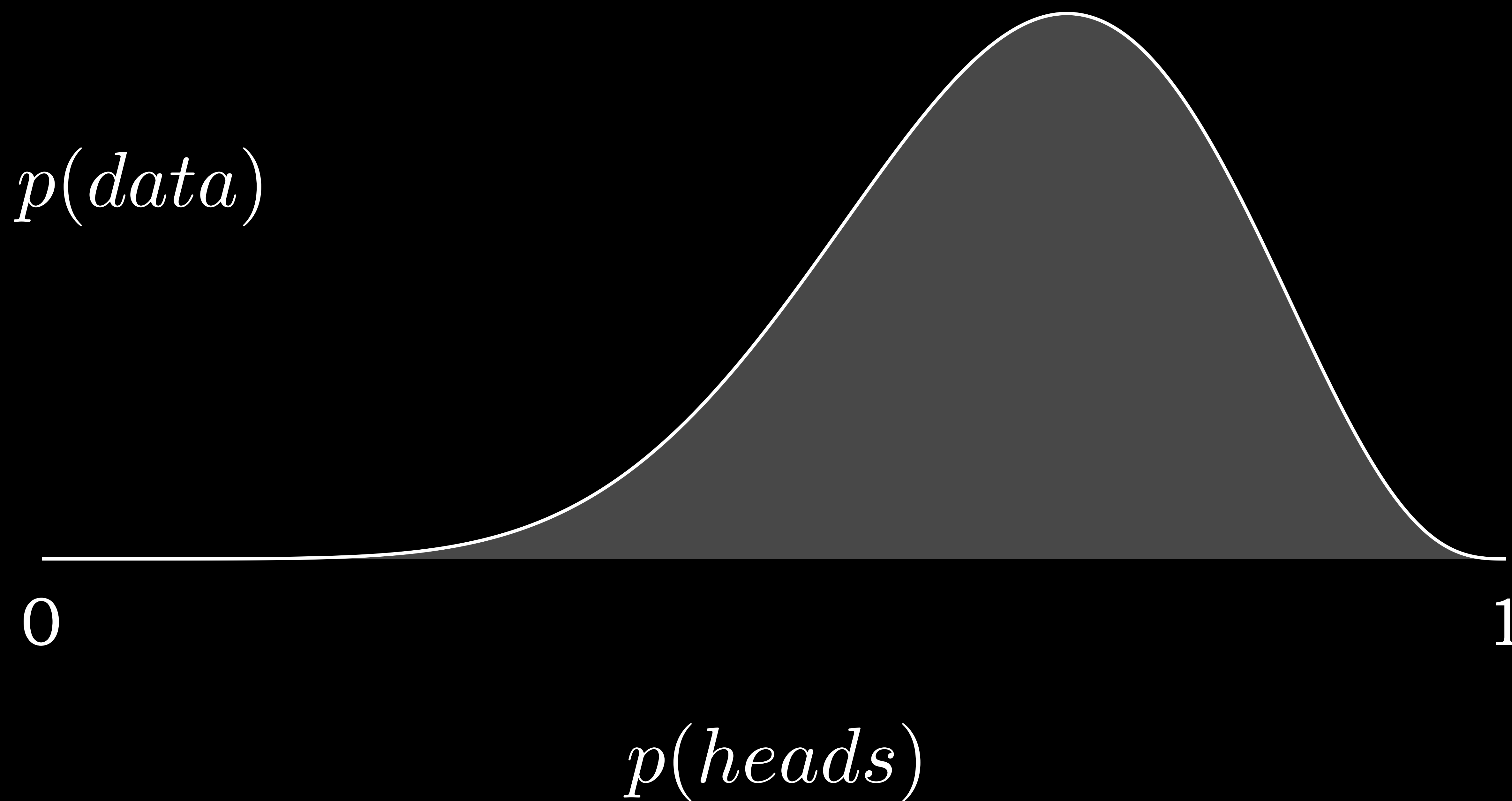
$p(heads)$ $1 - p(heads)$


1998
5 KRONER
DANMARK
1998
1998
1998
5 KRONER
DANMARK
1998
1998
5 KRONER
DANMARK
1998

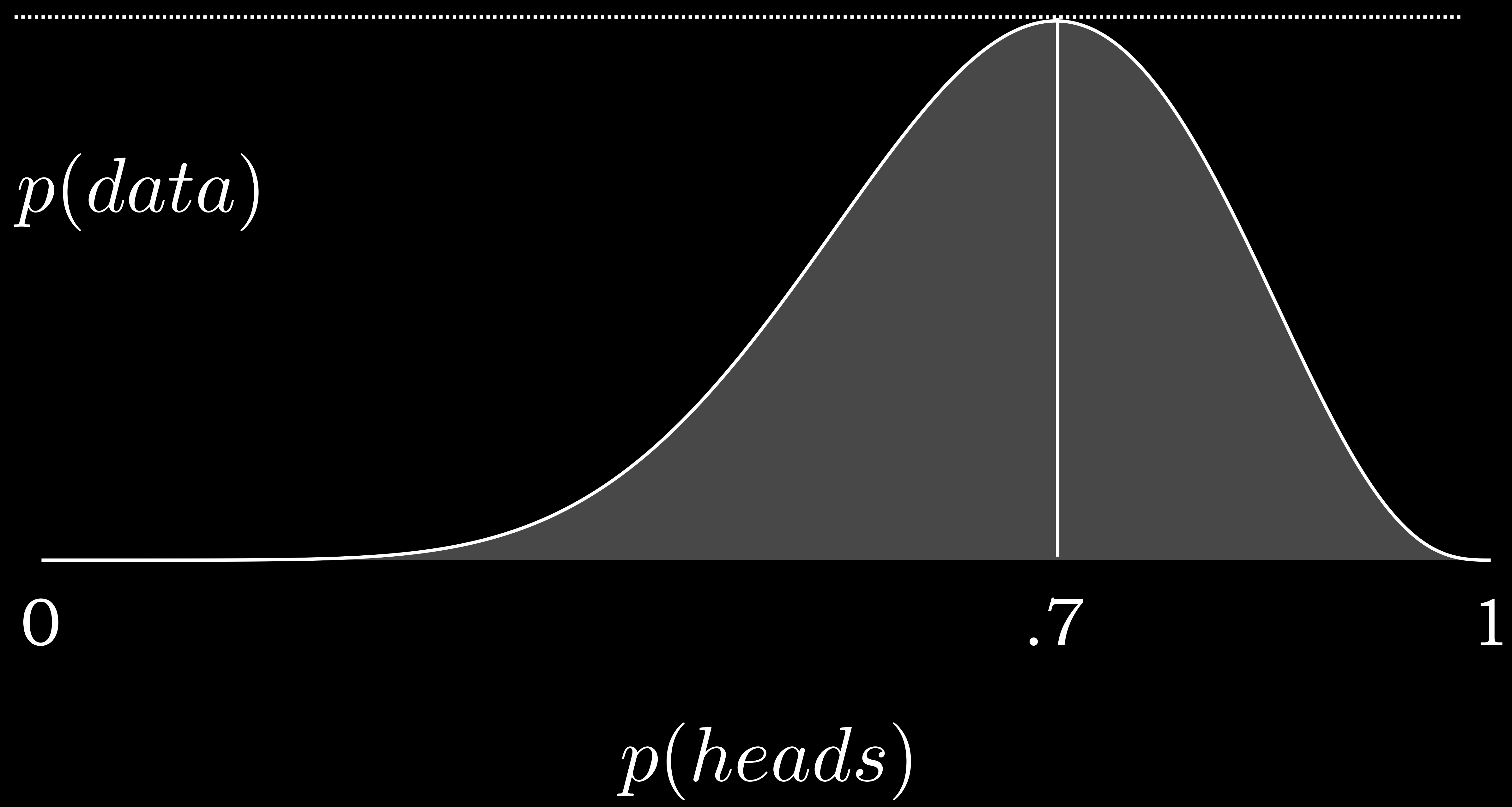
$p(heads)$?

$$p(data) = p(heads)^7 \times p(tails)^3$$

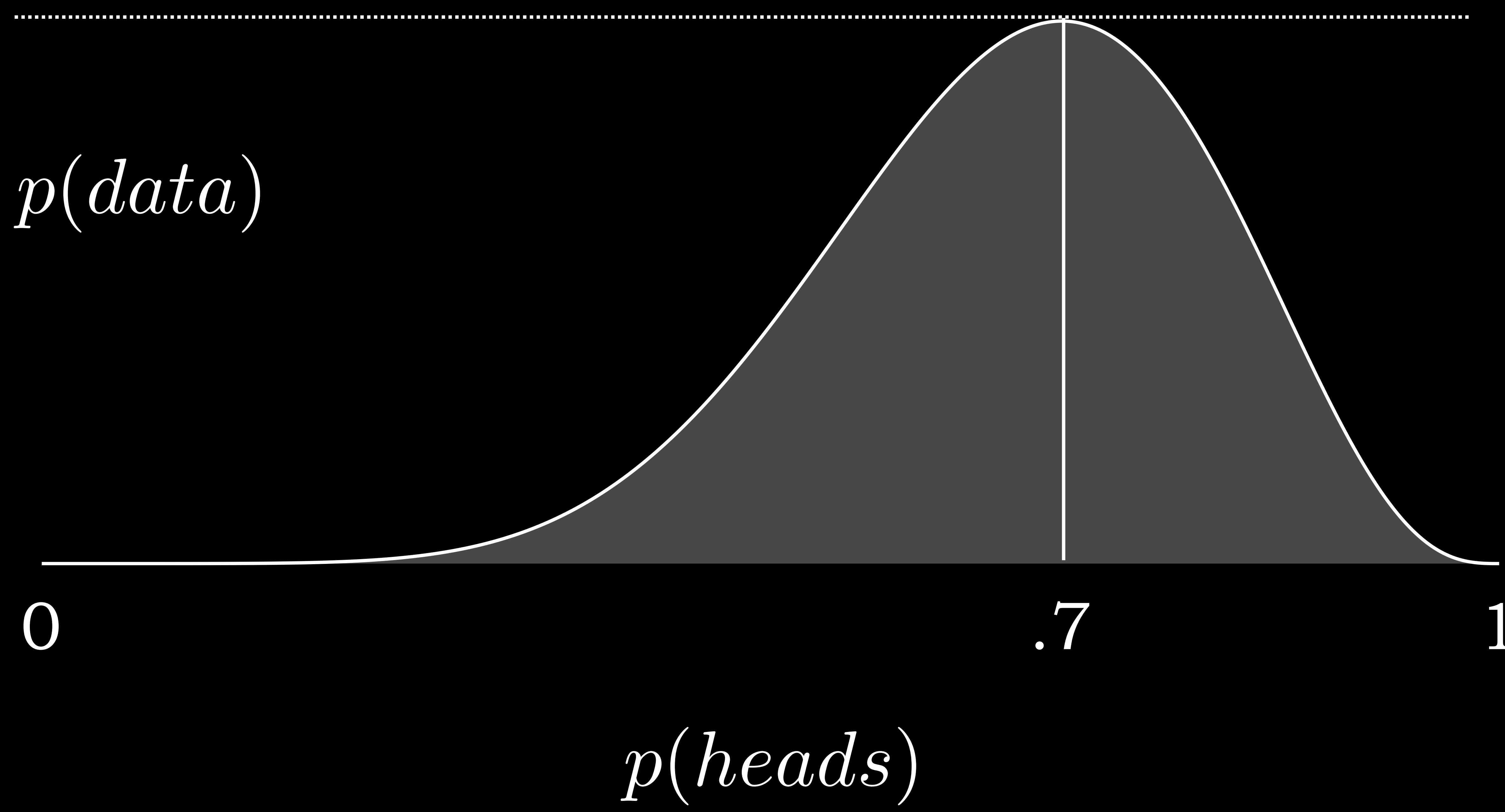
$$p(data) = p(heads)^7 \times [1 - p(heads)]^3$$



$p(data)$
0
1
$p(heads)$


$p(data)$
0
.7
1
$p(heads)$

can be derived analytically using Lagrange multipliers




THE WØRD

- Optimization

Language Models

$$p(remained|sky) = \frac{\text{\# of times I saw ``sky remained''}}{\text{\# of times I saw ``sky''}}$$

Language Models

This is a pretty old trick.

Language Models

This is a pretty old trick.

http://twitter.com/markov_bible

Language Models

This is a pretty old trick.

http://twitter.com/markov_bible

Jesus shall raise up children unto the way of the spices.

And some of them that do evil.

Language Models

This is a pretty old trick.

http://twitter.com/markov_bible

Jesus shall raise up children unto the way of the spices.

And some of them that do evil.

But be careful! What if we haven't seen some word sequences?

Language Models

This is a pretty old trick.

http://twitter.com/markov_bible

Jesus shall raise up children unto the way of the spices. And some of them that do evil.

But be careful! What if we haven't seen some word sequences?

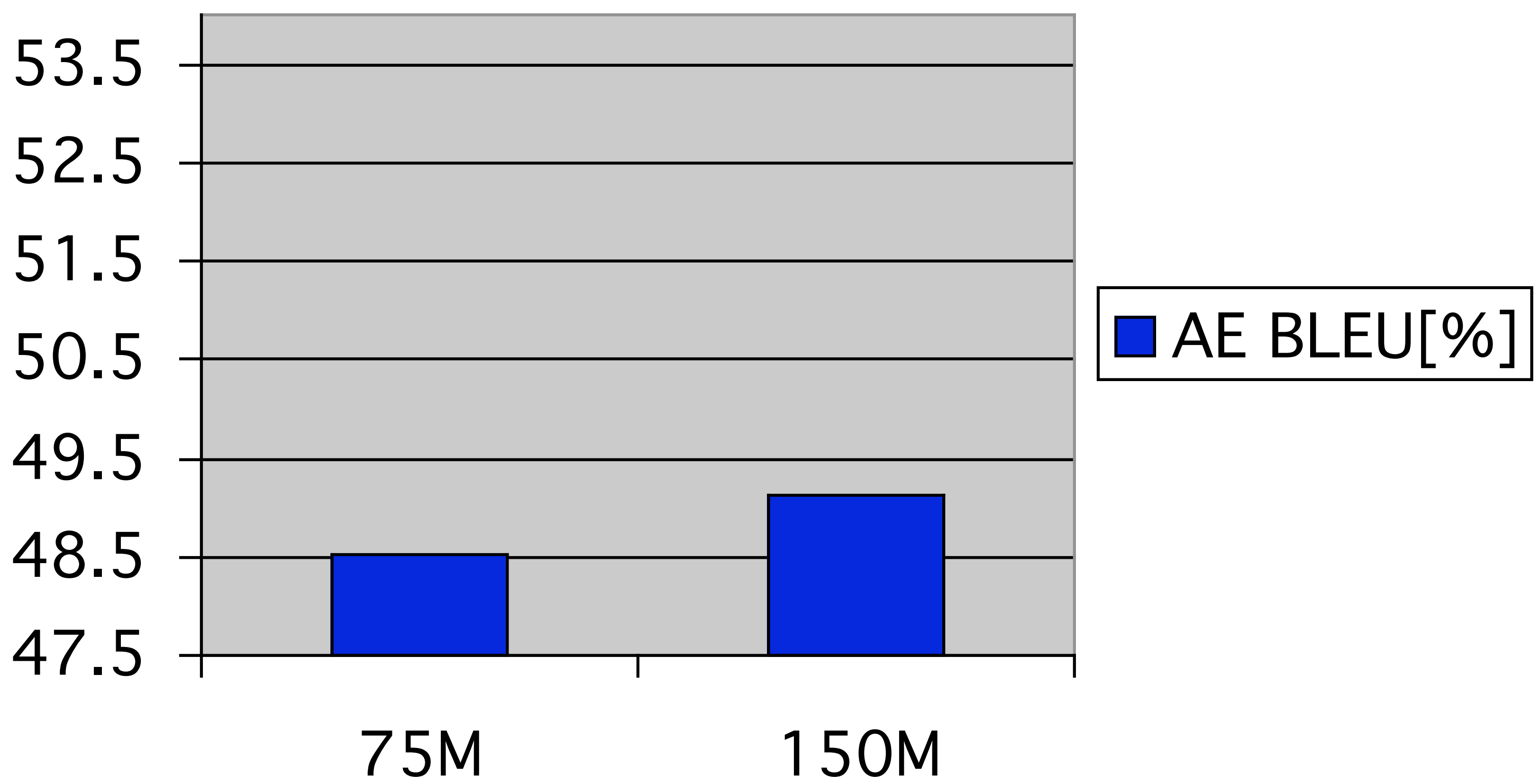
Won't cover this too much, but keyword is *smoothing*.

Language Models

- The language model does not depend in any way on parallel data.
- How much English data should we train it on?

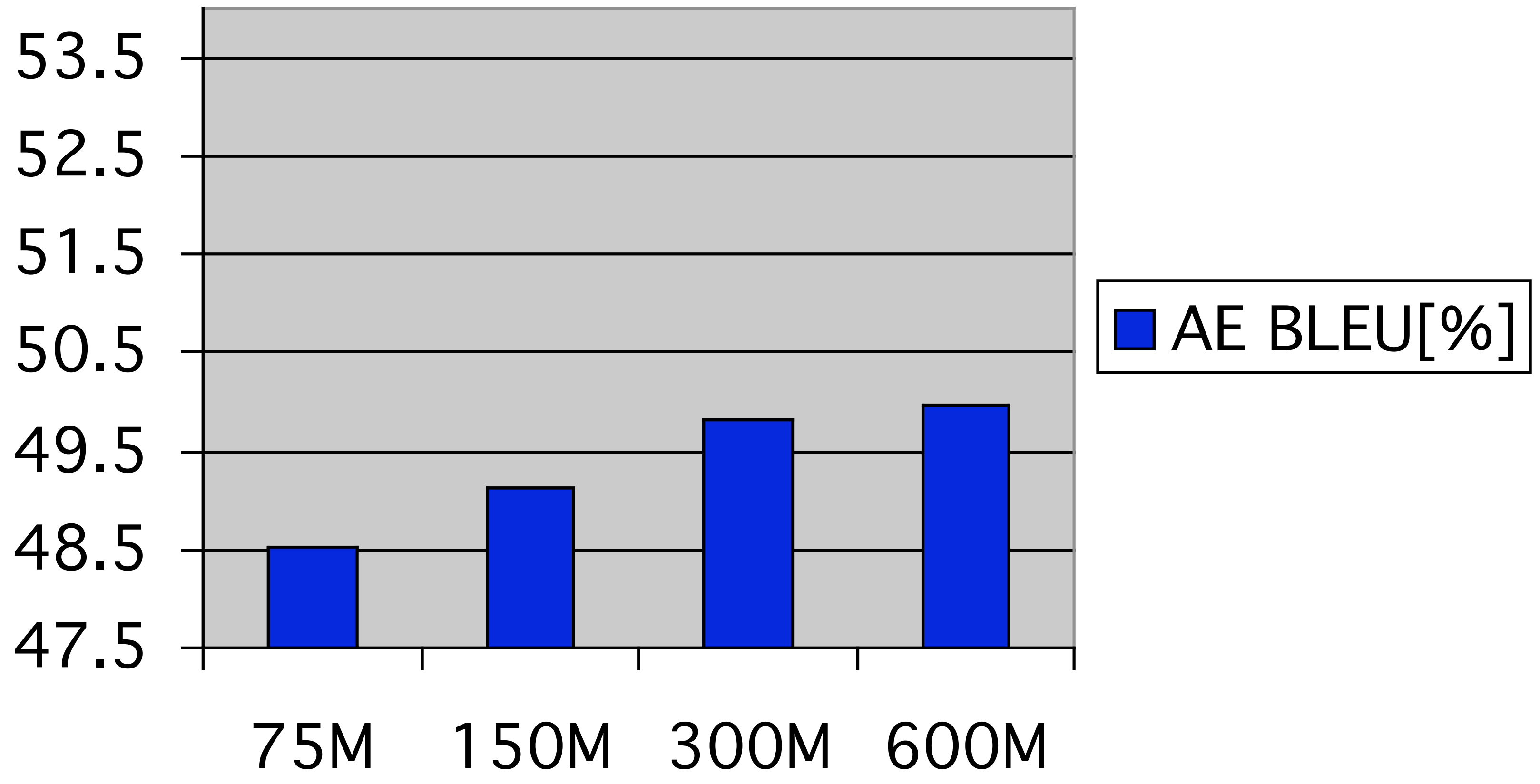
Language Models

Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system (NIST test data)

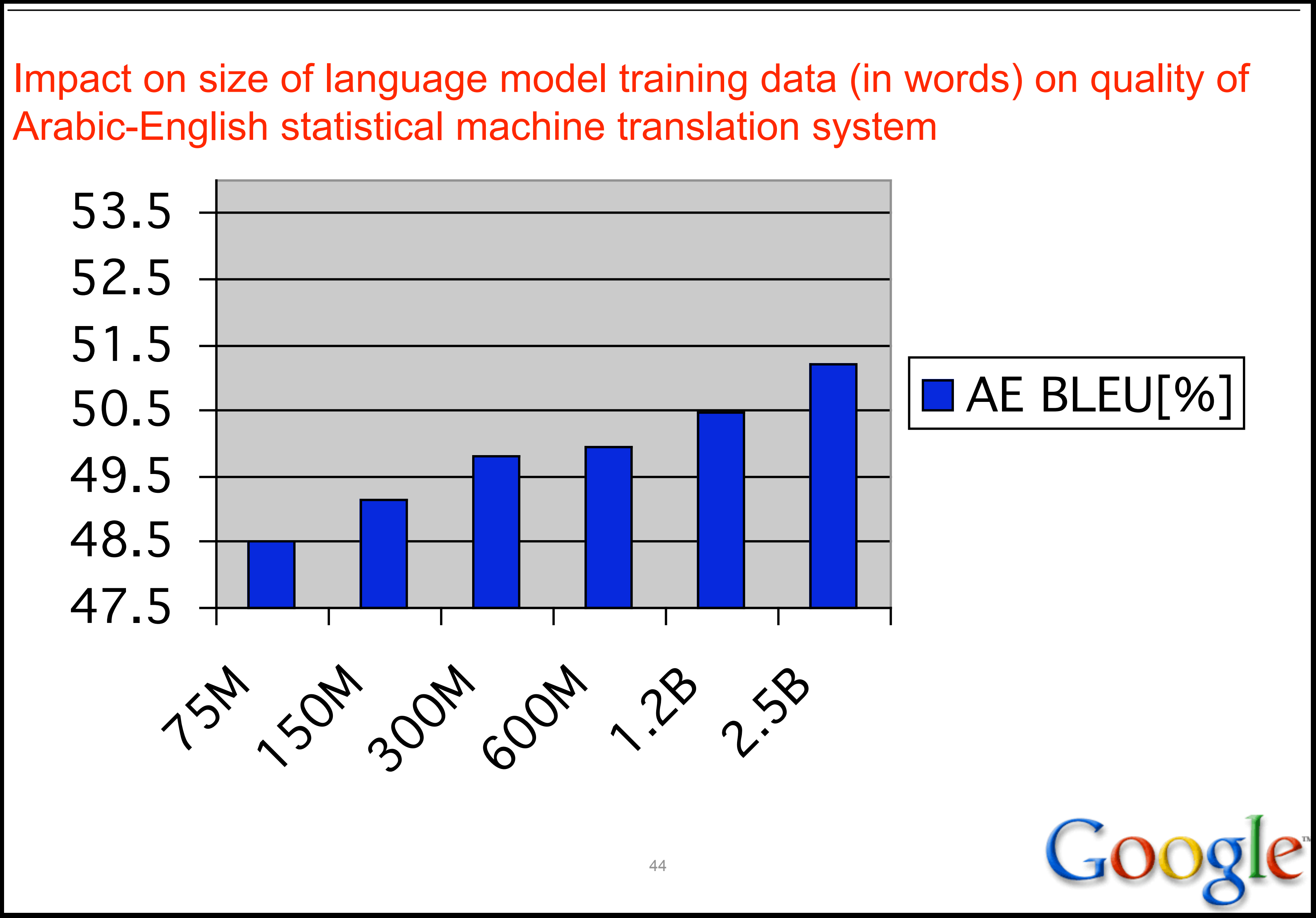


Language Models

Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system

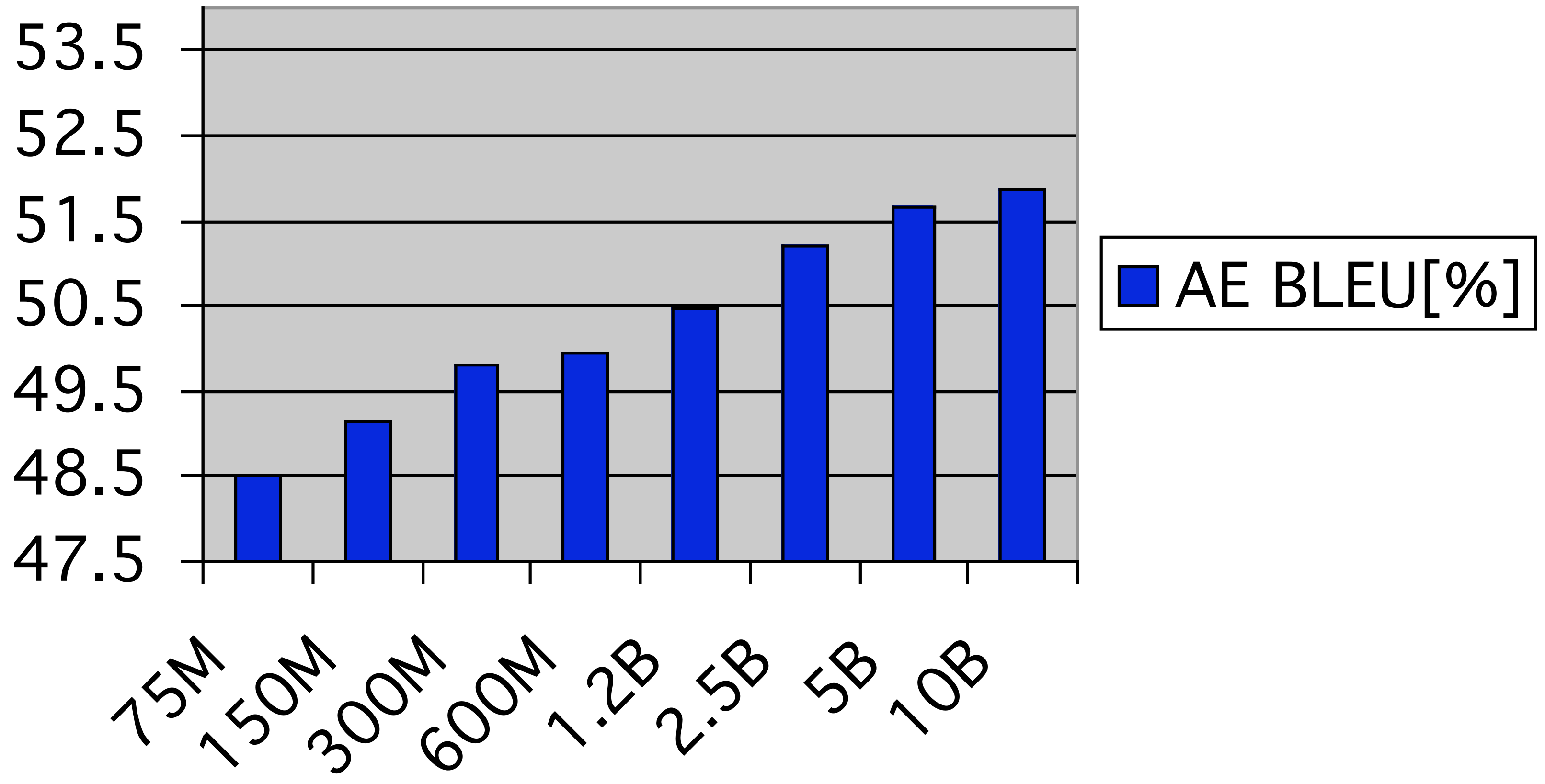


Language Models



Language Models

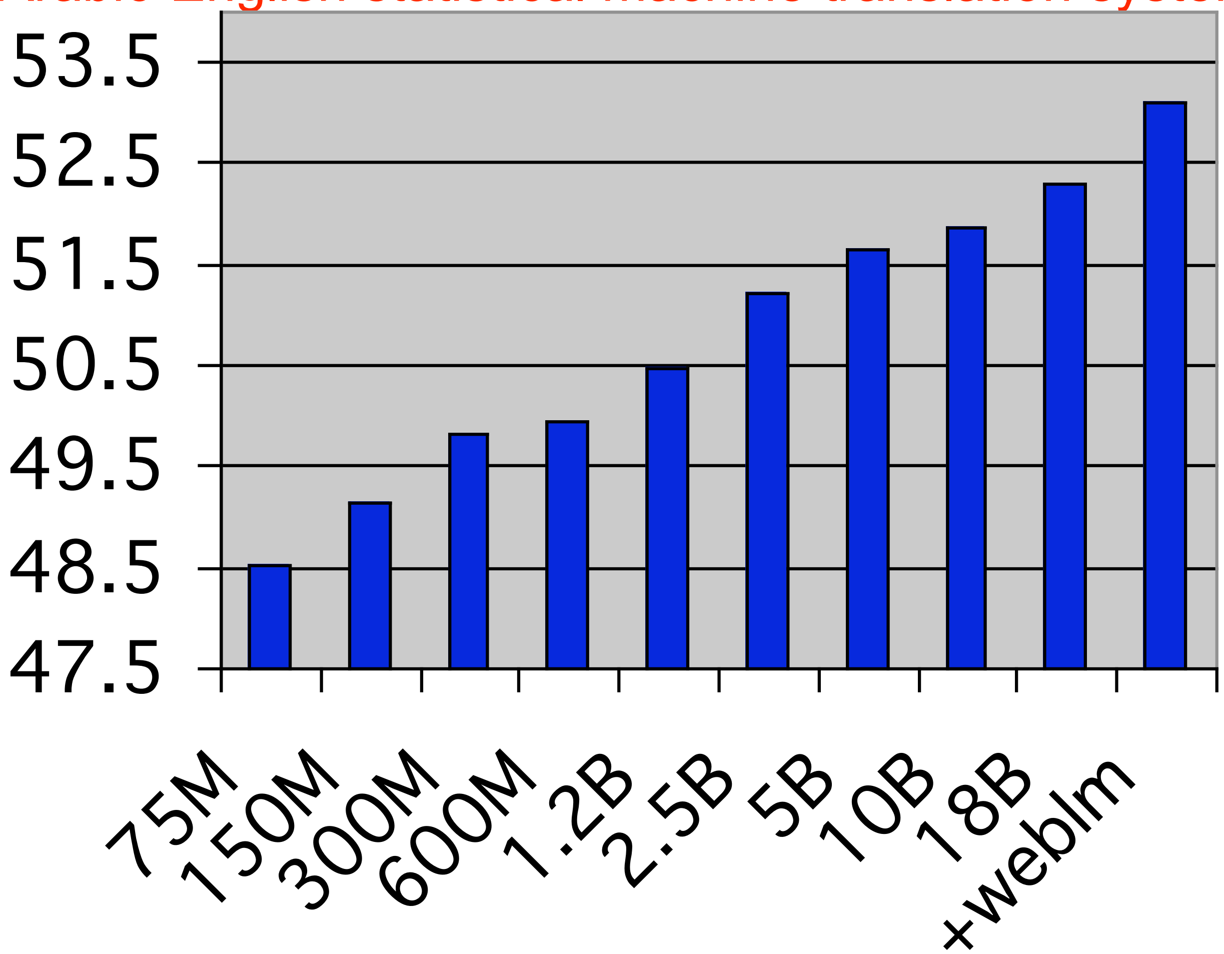
Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



46

Language Models

Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



+weblm =
LM trained on
219B words of
web data

Popular Implementations

- SRI-LM -- www.speech.sri.com/projects/srilm
- KenLM -- http://kheafield.com/code/kenlm/
- BerkeleyLM -- http://code.google.com/p/berkeleylm/

Language Models

- There’s no data like more data.
- Language models serve a similar function in speech recognition, optical character recognition, and other probabilistic models of text data.

Translation Models

What is a good story about how a Chinese sentence came into being, given that we already have an English sentence?

Translation Models

What is a good story about how a Chinese sentence came into being, given that we already have an English sentence?

Note: in this example I'll show you an English sentence, conditioned on a Chinese sentence. Note that we can apply the same technique in either direction.

Translation Models

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

$$p(English|Chinese)$$

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

$$p(English|Chinese)$$

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(English|Chinese)$$

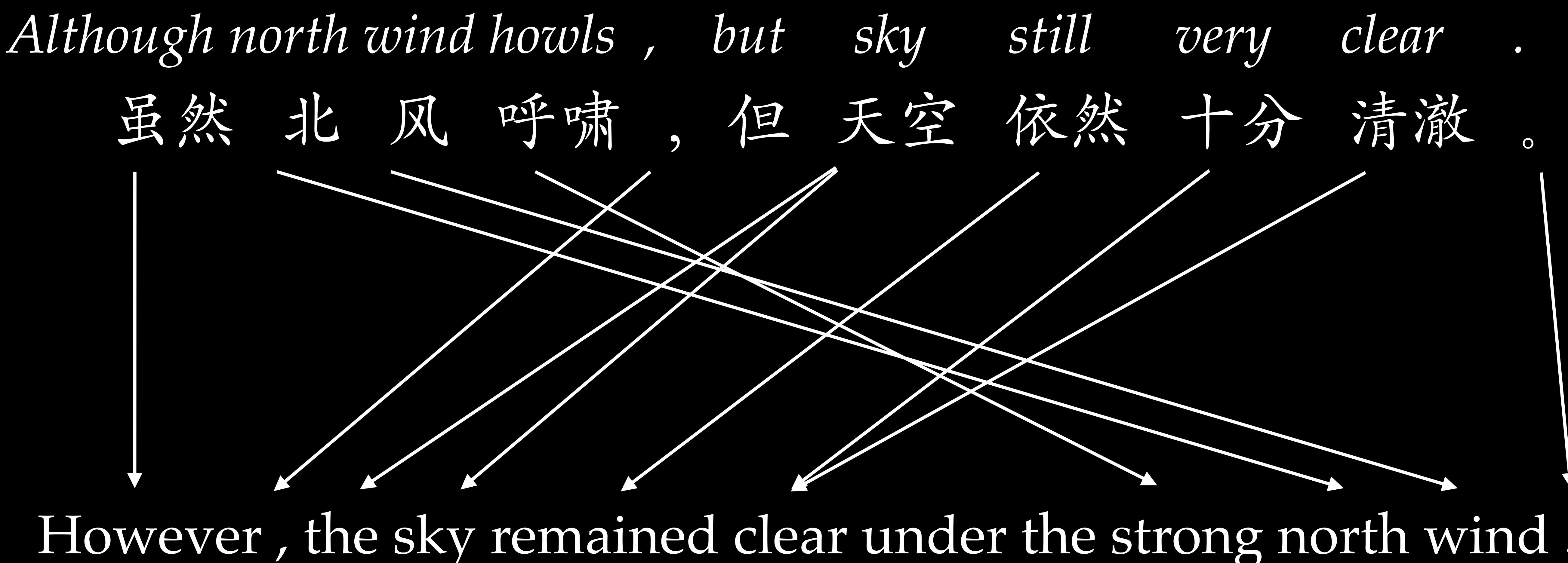
Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Translation Models



$$p(English|Chinese)?$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

$$p(English\ length|Chinese\ length)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

$\_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_ \quad \_\_\_$

$$p(English\ length|Chinese\ length)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

$$p(Chinese\ word\ position)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

$$p(English\ word|Chinese\ word)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However ,

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However ,

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

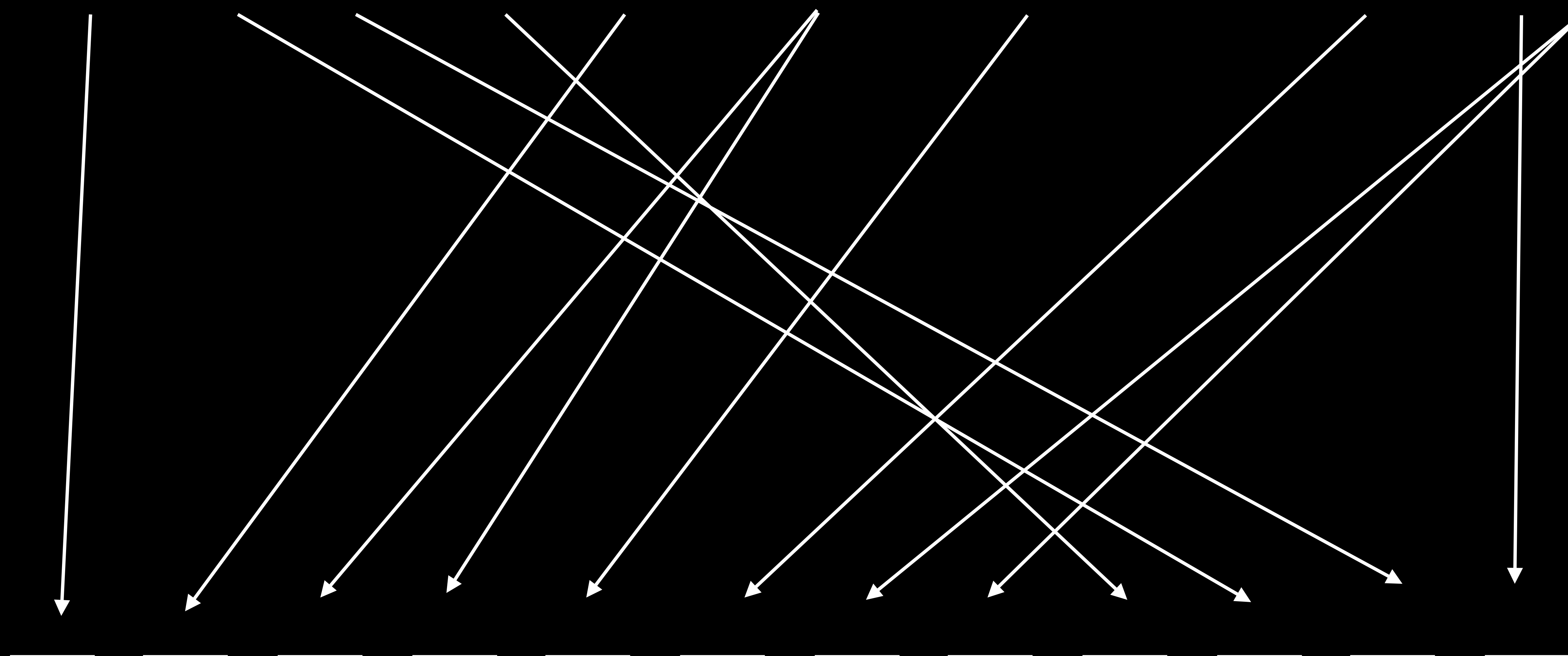
___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However , the

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε



However , the sky remained clear under the strong north wind .

IBM Model 1

IBM Model 1

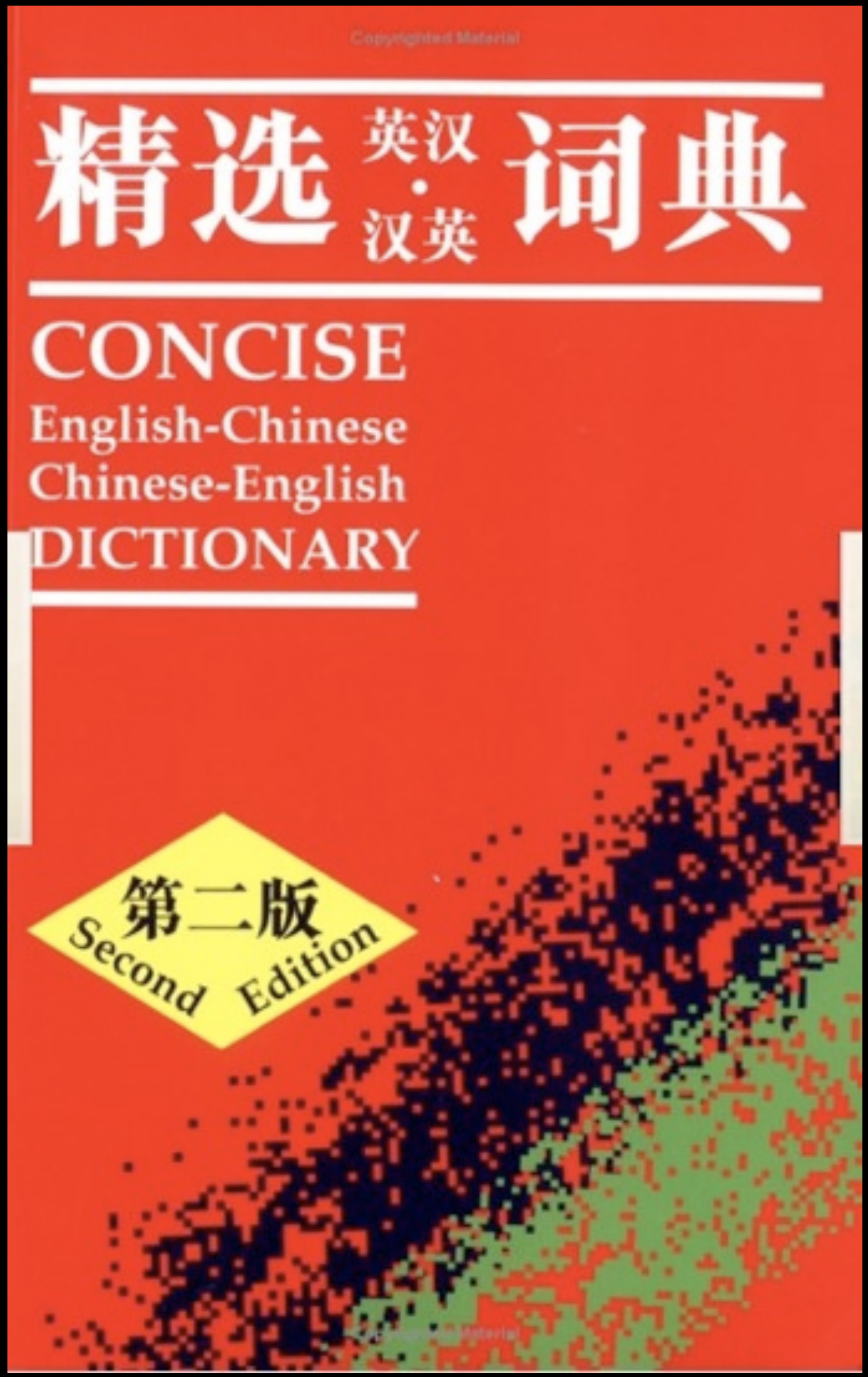
- Word translation probabilities.

IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.

IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.



IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.

IBM Model 1

$$p(despite|\text{虽然})$$

$$p(however|\text{虽然})$$

$$p(although|\text{虽然})$$

$$p(northern|\ \text{北}\ )$$

$$p(north|\ \text{北}\ )$$

IBM Model 1

$$p(despite|\text{虽然}) \qquad ???$$

$$p(however|\text{虽然}) \qquad ???$$

$$p(although|\text{虽然}) \qquad ???$$

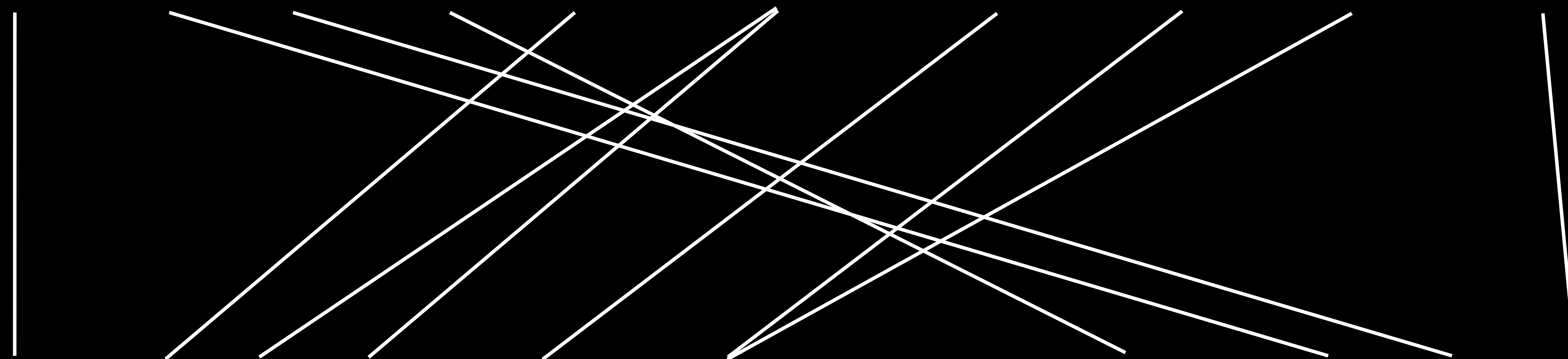
$$p(northern|\text{北}) \qquad ???$$

$$p(north|\text{北}) \qquad ???$$

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

$$p(however | \text{虽然}) = \frac{\text{\# of times 虽然 aligns to However}}{\text{\# of times 虽然 occurs}}$$

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(however \mid \text{虽然}) = \frac{\text{\# of times 虽然 aligns to However}}{\text{\# of times 虽然 occurs}}$$

Up Next:

- Word alignment.
- Assignment 0: by tomorrow!
- Assignment 1 (word alignment) posted soon.
- Research lectures on machine translation!
    - Chris Dyer (CMU), Monday 10am, Stieff
    - Shankar Kumar (Google), Tuesday noon, B17

Leaderboard

This page contains the leaderboards for all as
date is downloaded according the base URL

	Assignments					
Handle	**#0**	**#1**	**#2**	**#3**	**#4**	**All**
obzk	63	-	-	-	-	63.00
rlk	47	-	-	-	-	47.00
NathanStark	42	-	-	-	-	42.00
thrax	14	-	-	-	-	14.00
SI	7	-	-	-	-	7.00
Lakie	7	-	-	-	-	7.00
TangDou	5	-	-	-	-	5.00
Shibboleth	4	-	-	-	-	4.00
PandaPirate	1	-	-	-	-	1.00

Up Next:

- Word alignment.
- Assignment 0: by tomorrow!
- Assignment 1 (word alignment) posted soon.
- Research lectures on machine translation!
  - Chris Dyer (CMU), Monday 10am, Stieff
  - Shankar Kumar (Google), Tuesday noon, B17

Leaderboard

This page contains the leaderboards for all as
date is downloaded according the base URL

	Assignments					
Handle	**#0**	**#1**	**#2**	**#3**	**#4**	**All**
obzk	63	-	-	-	-	63.00
rlk	47	-	-	-	-	47.00
NathanStark	42	-	-	-	-	42.00
thrax	14	-	-	-	-	14.00
SI	7	-	-	-	-	7.00
Lakie	7	-	-	-	-	7.00
TangDou	5	-	-	-	-	5.00
Shibboleth	4	-	-	-	-	4.00
PandaPirate	1	-	-	-	-	1.00