Yuille & Kersten

Vision as Bayesian inference: analysis by synthesis?

- the study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

- the study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

- generalisation from **artificial stimuli** can lead to faulty generalisations

- the study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

- generalisation from **artificial stimuli** can lead to faulty generalisations

- **proposal for a theory** that can can deal with the complexity of natural images:

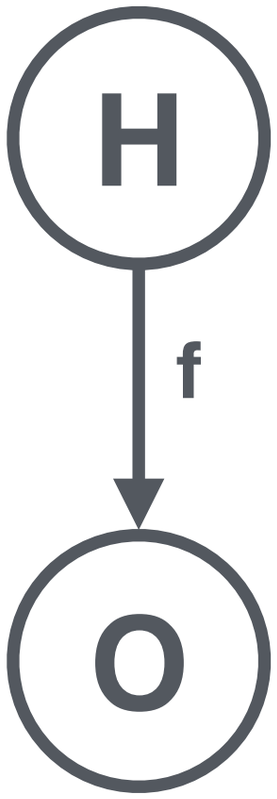
- the study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

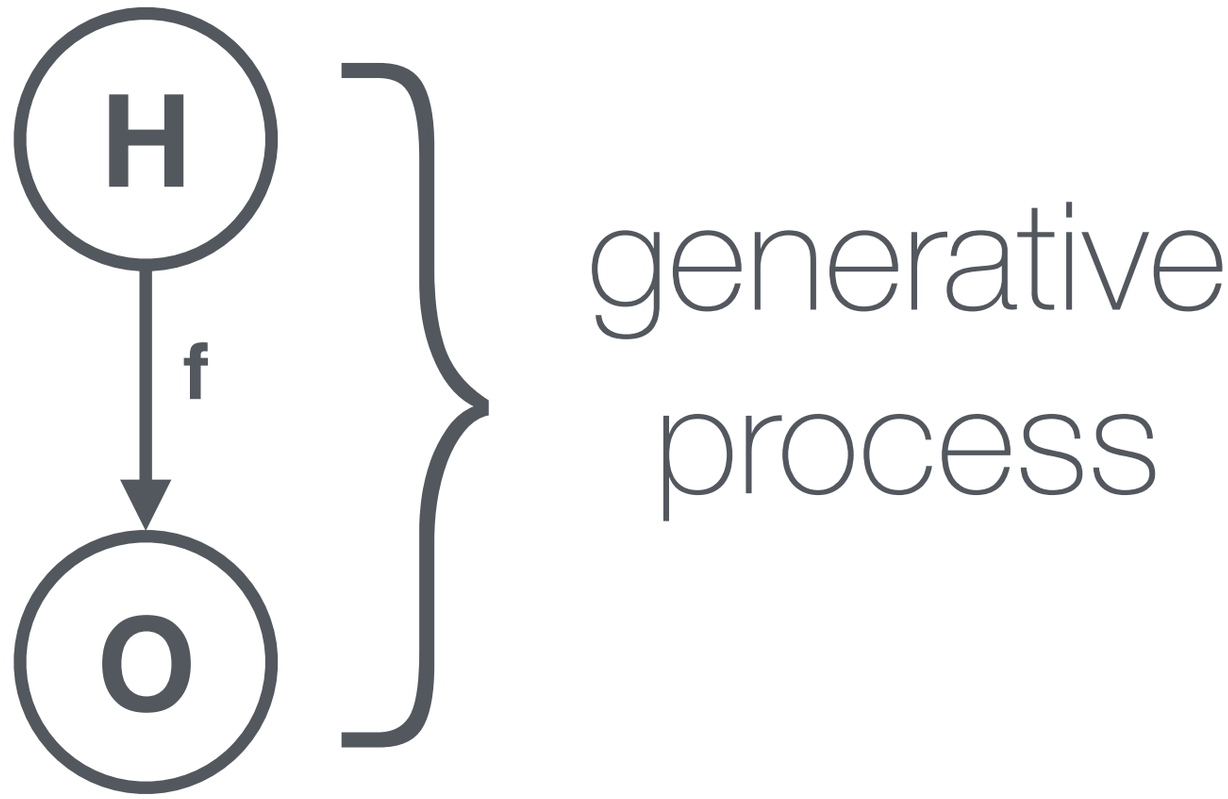
- generalisation from **artificial stimuli** can lead to faulty generalisations

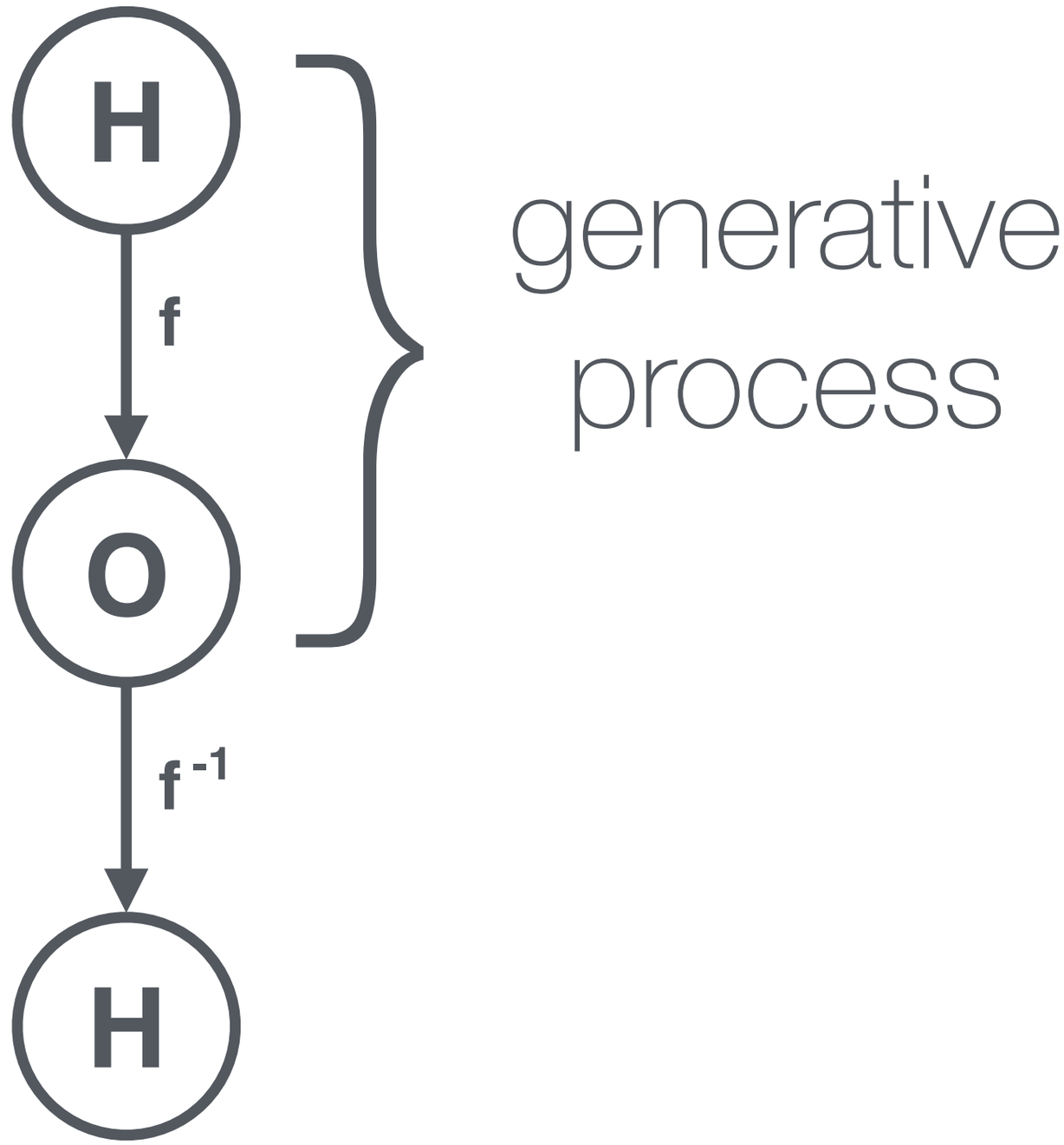
- **proposal for a theory** that can can deal with the complexity of natural images:

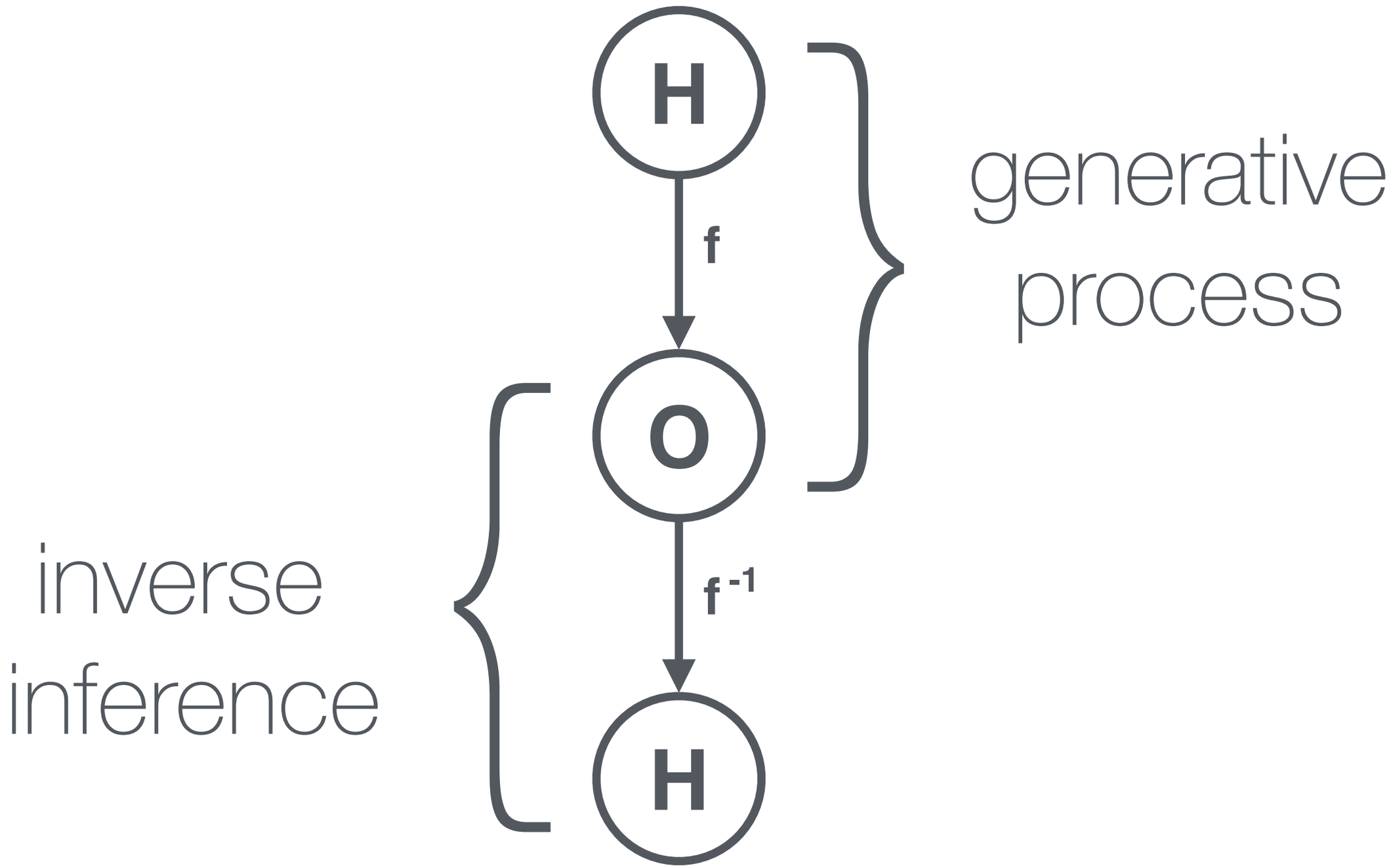
  - **perception as Bayesian inference**
  - **analysis by synthesis**

perception as inverse inference


H
f
O

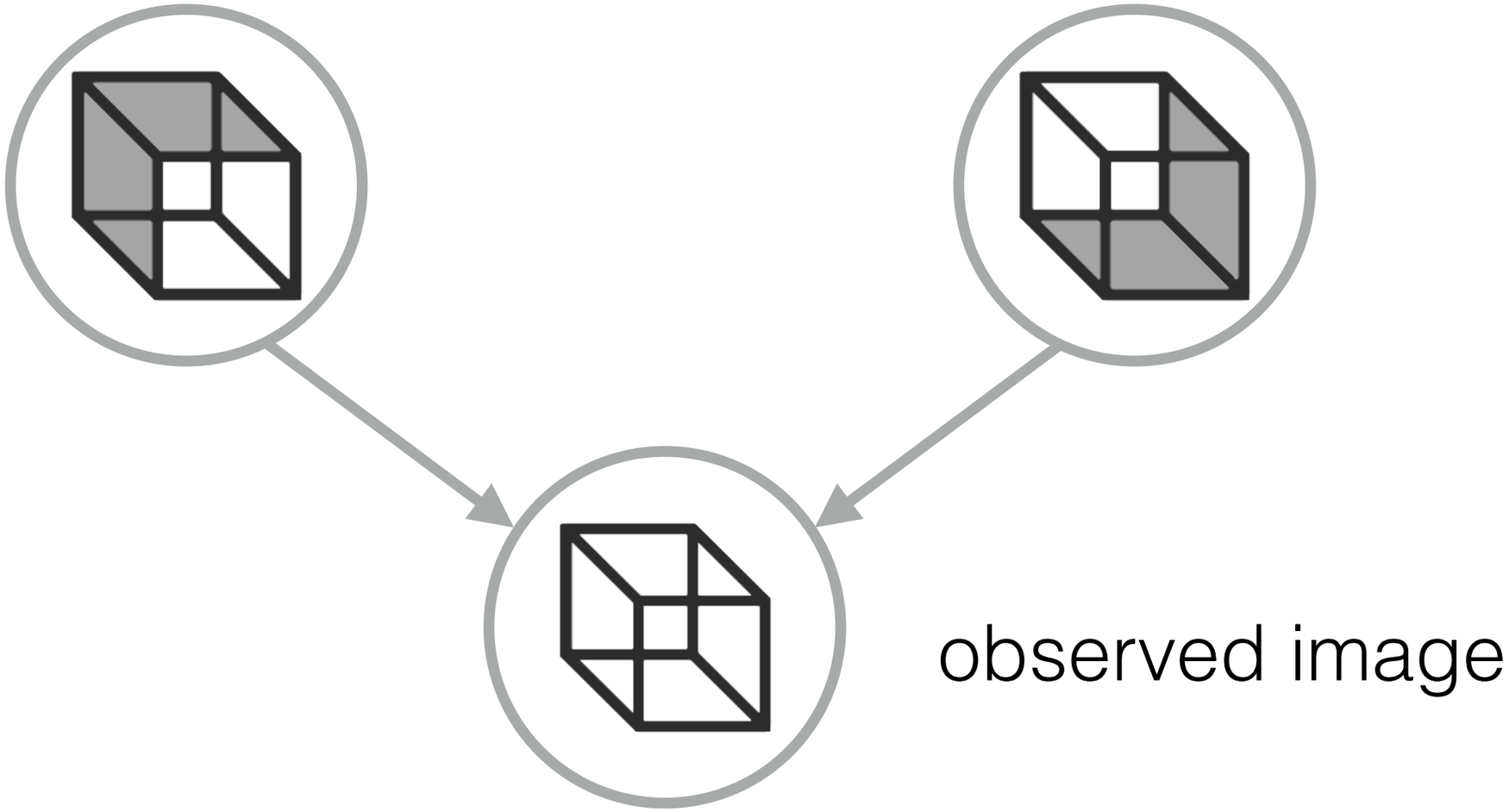

H
f
O
generative
process

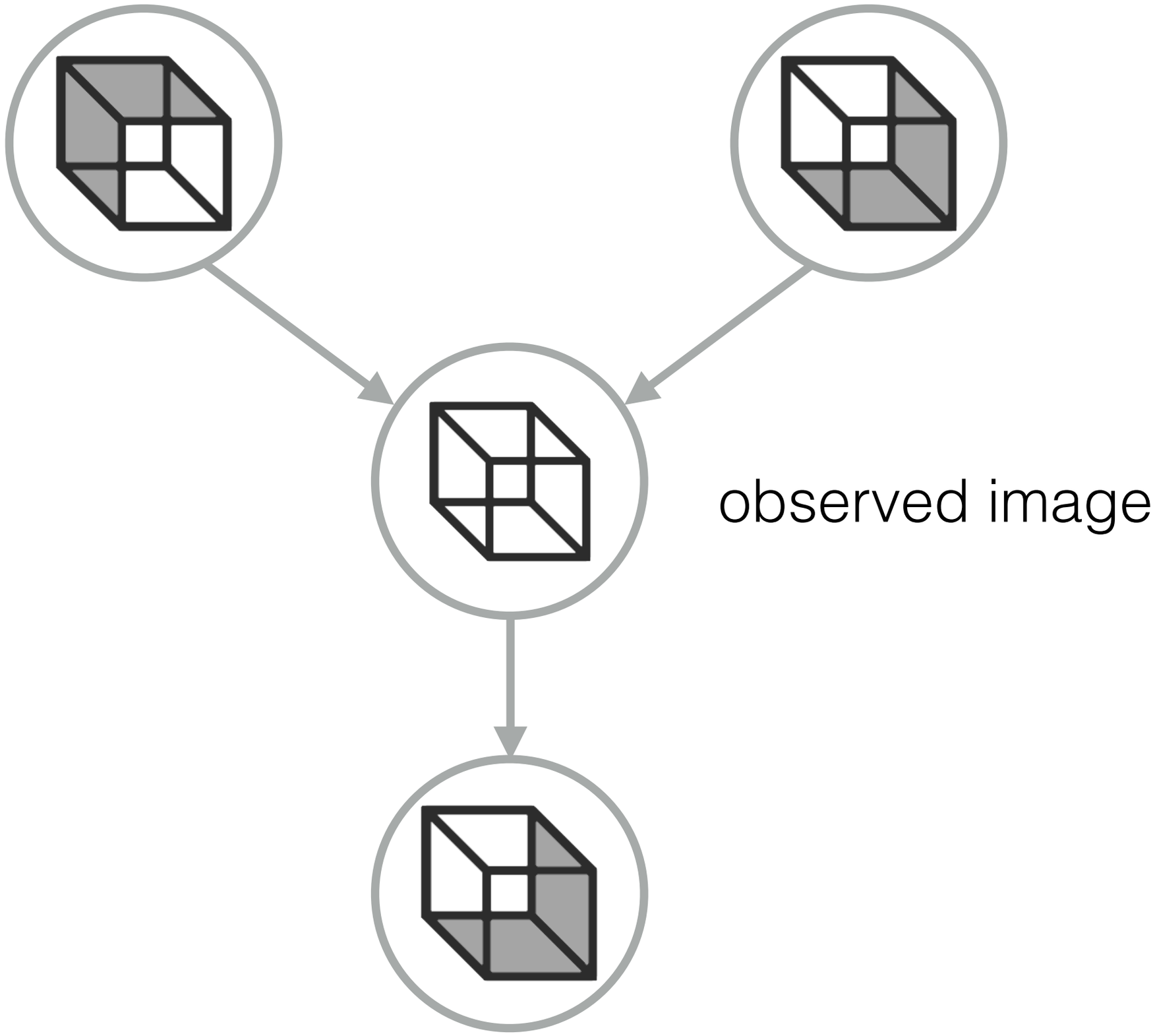

H
f
O
f -1
H
generative
process


H
f
O
f -1
H
generative
process
inverse
inference

ambiguity

observed image

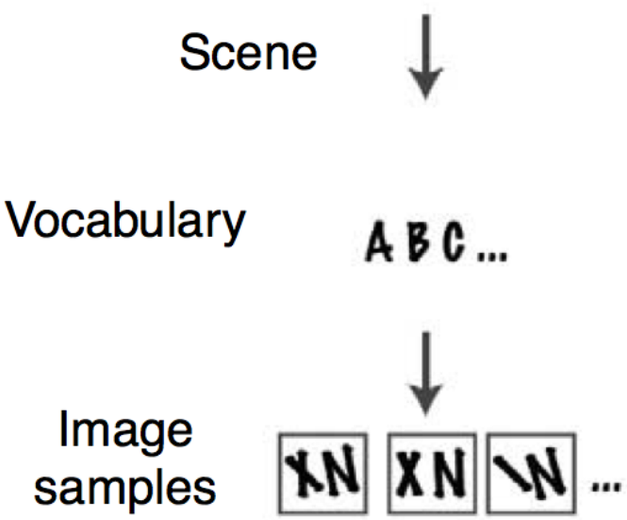

observed image

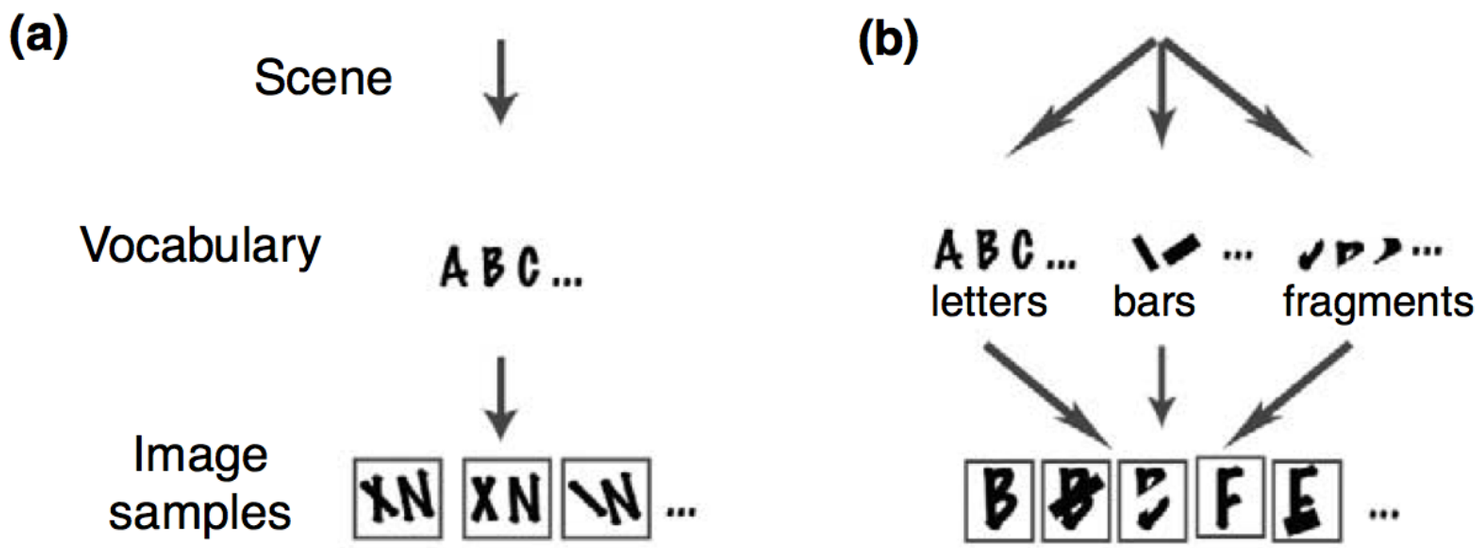

observed image

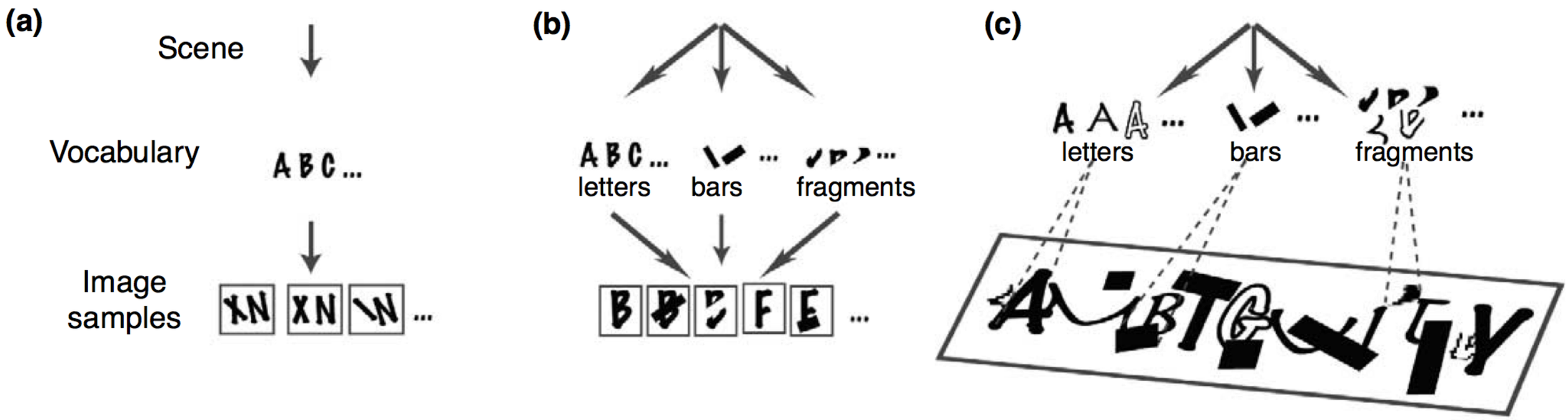
need for top down

synthetic images

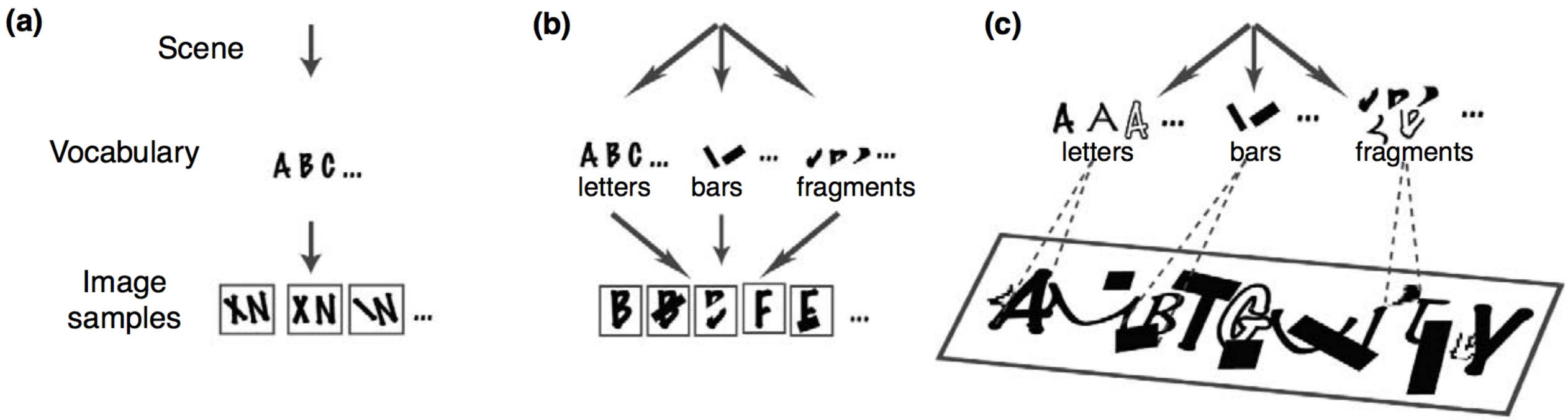
(a)



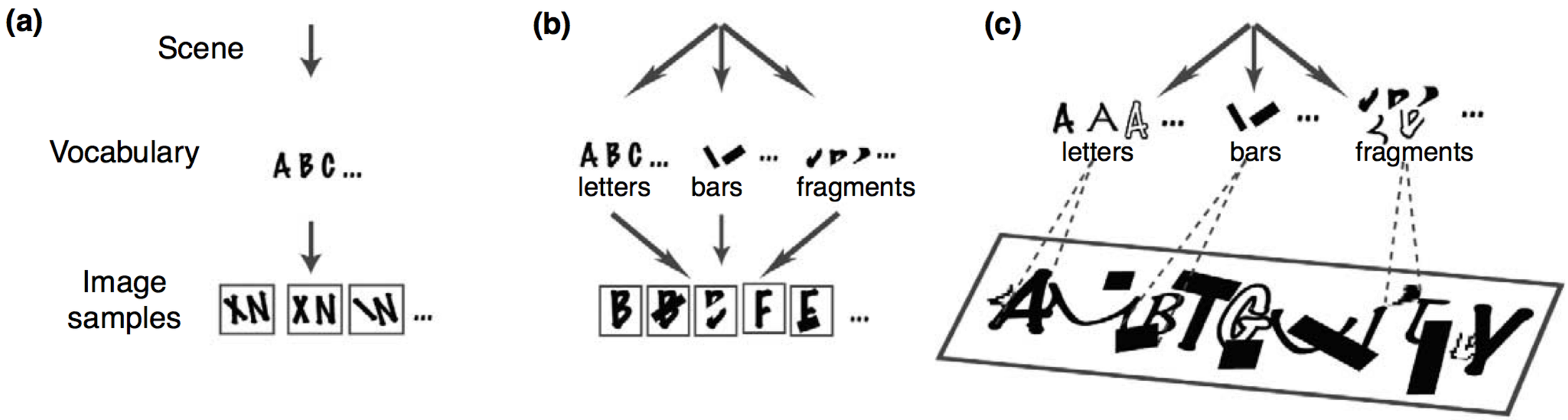

(a)
Scene
Vocabulary
A B C ...
Image samples
(b)
A B C ...
letters
bars
fragments


(a)
Scene
Vocabulary
Image samples
(b)
letters
bars
fragments
(c)
letters
bars
fragments

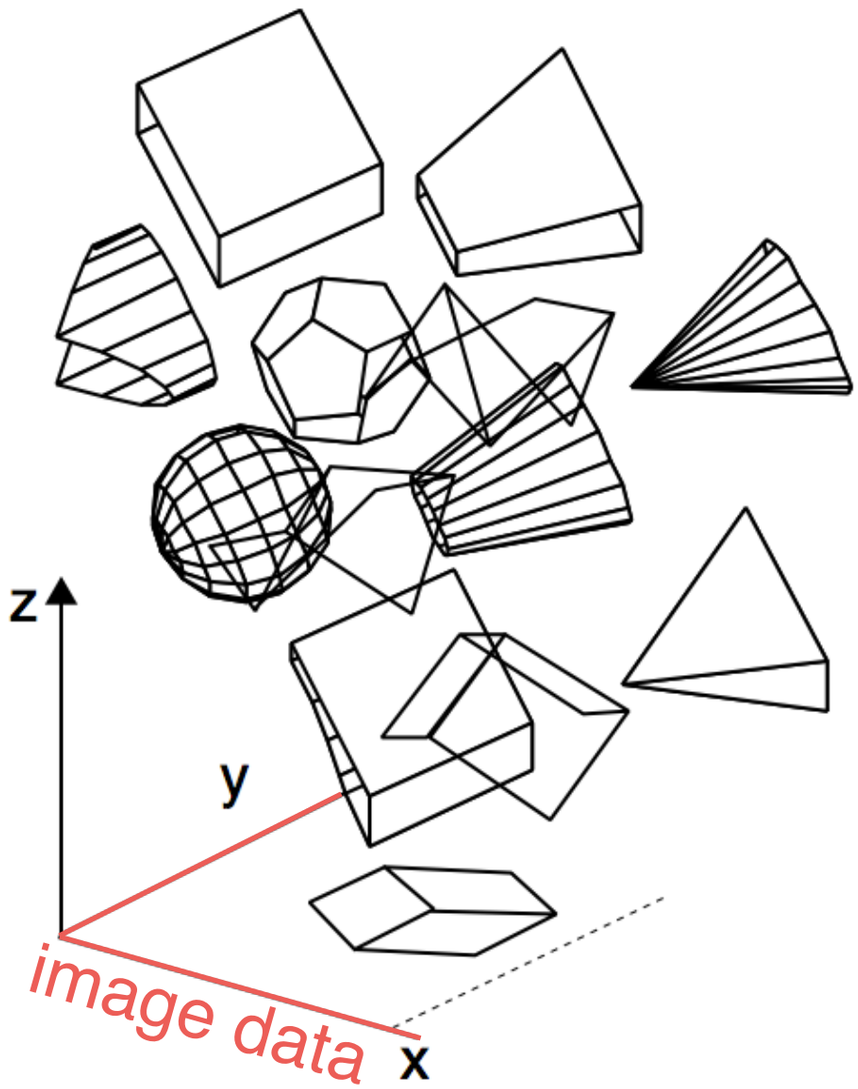
- ambiguity increases with model complexity



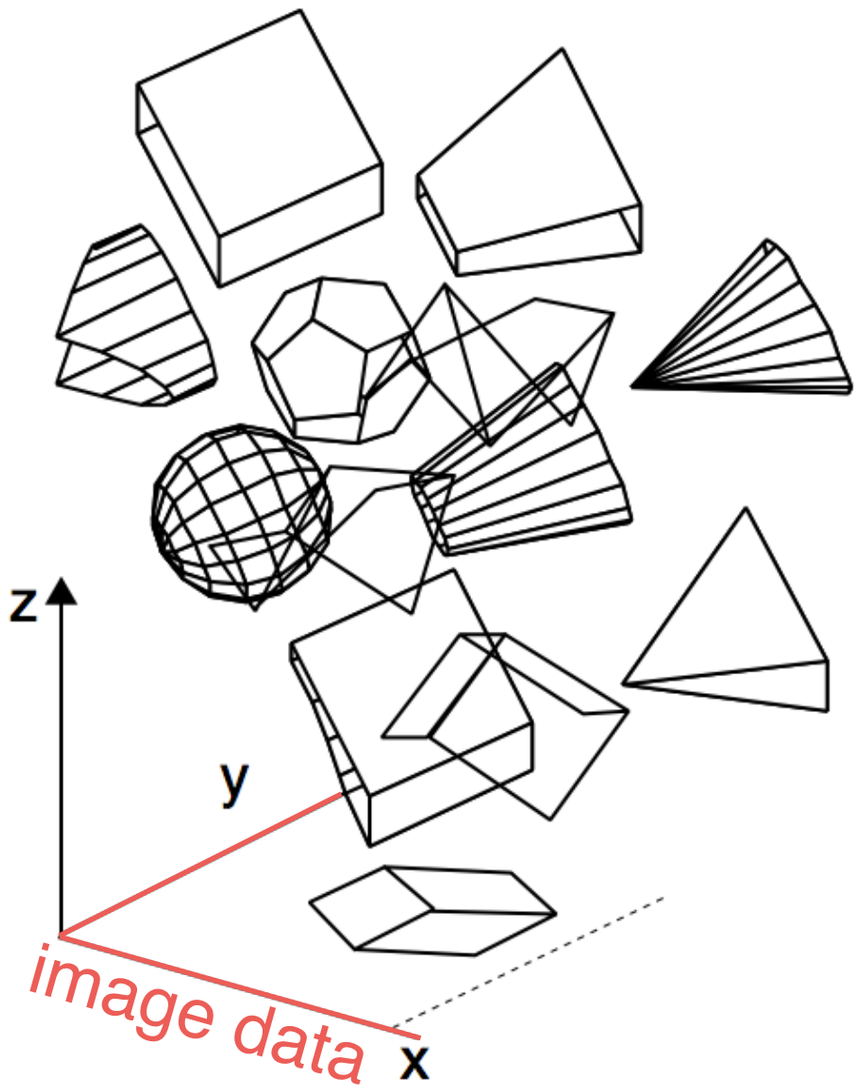
- ambiguity increases with model complexity



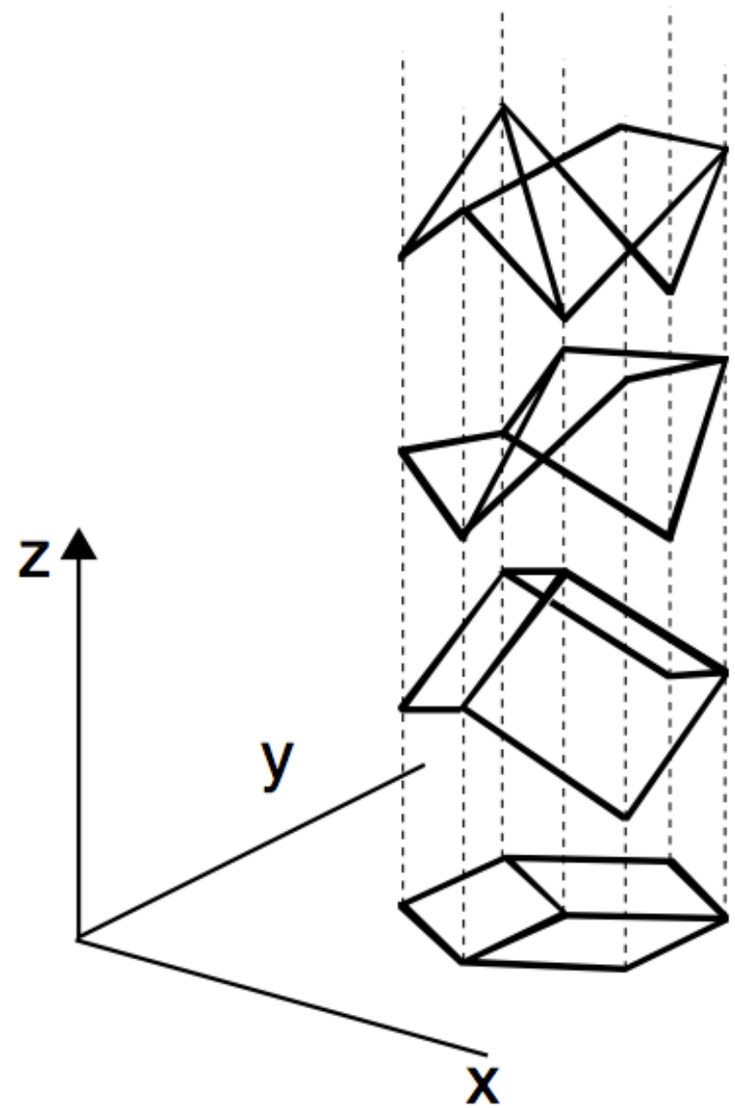
- but it can be resolved using Bayesian inference



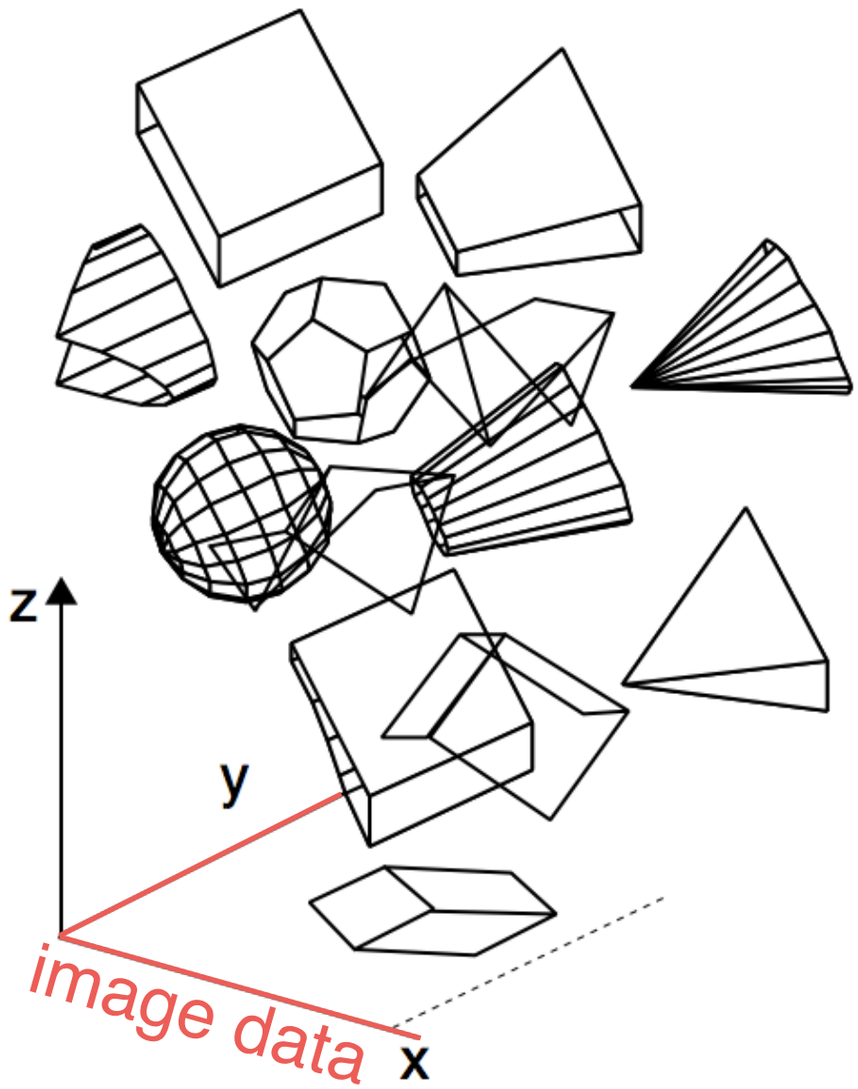
hypothesis space



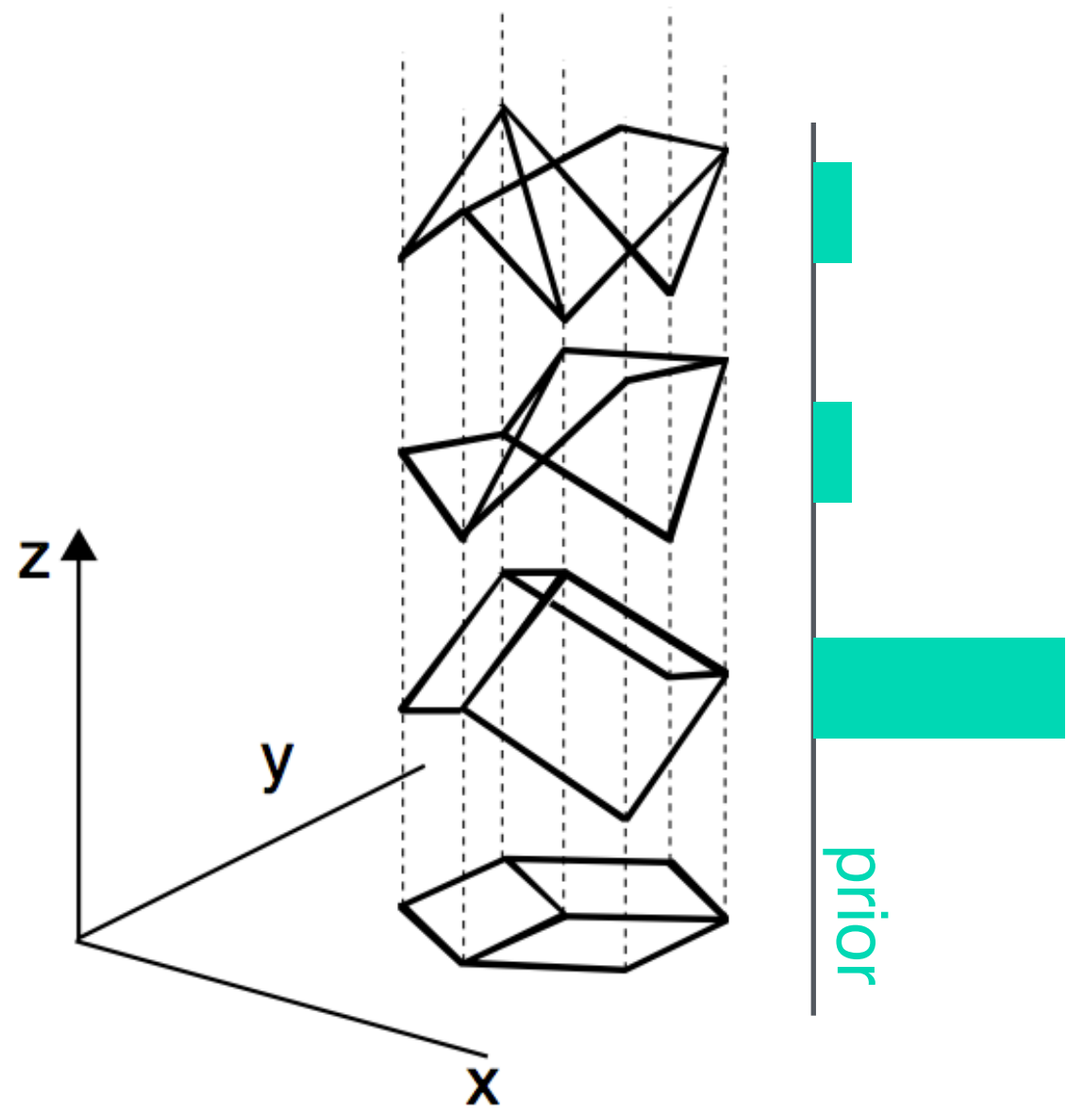
hypothesis space



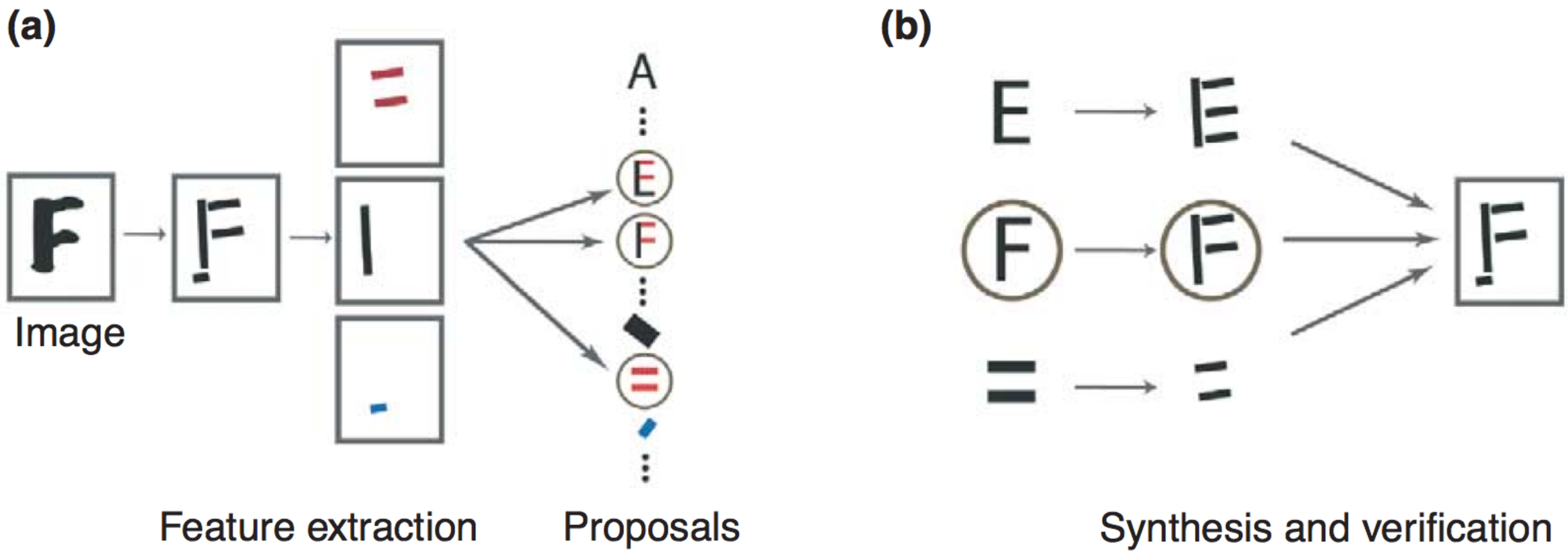
hypotheses that explain data well



hypothesis space

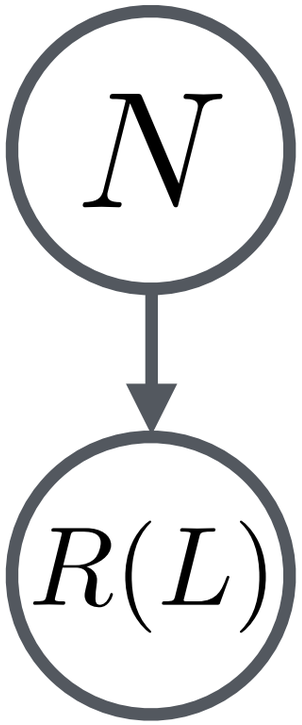


hypotheses that explain data well


(a)
Image
A
Feature extraction
Proposals
(b)
Synthesis and verification

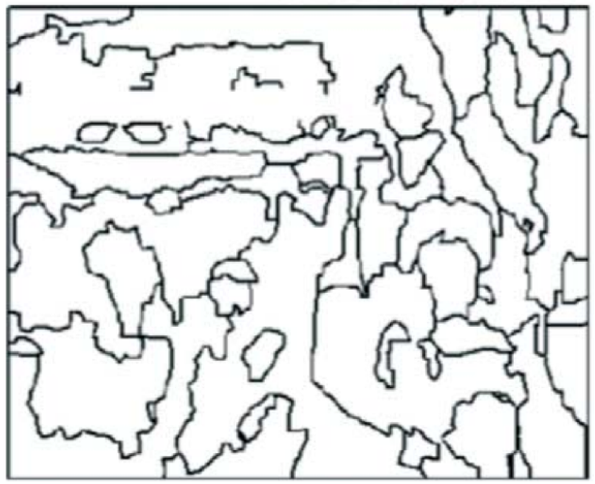
generative model for natural images

N number of nodes 32



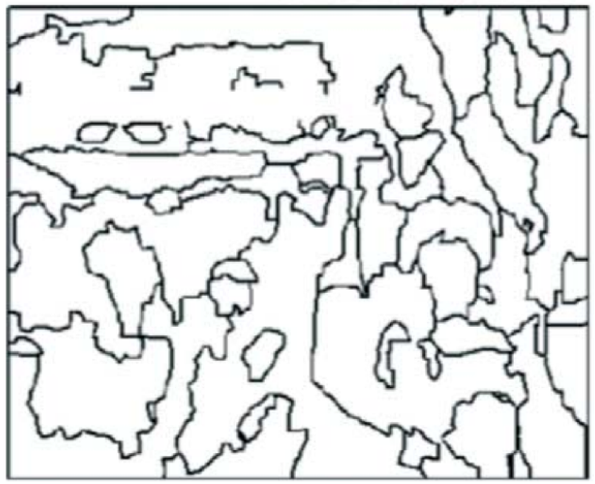
number of nodes 32

region



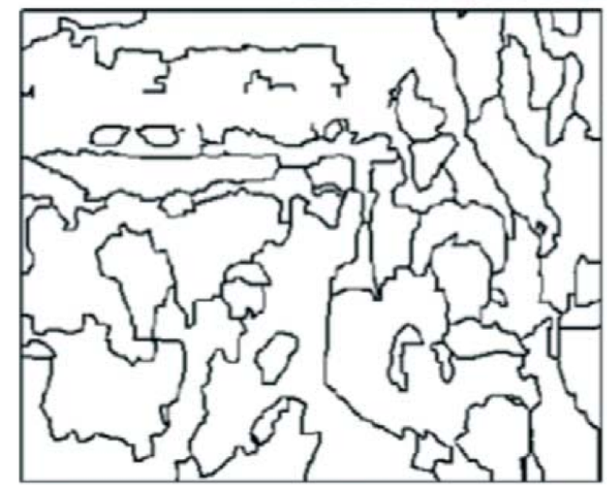
N number of nodes 32

$R(L)$ region



ζ type of model text/face/background

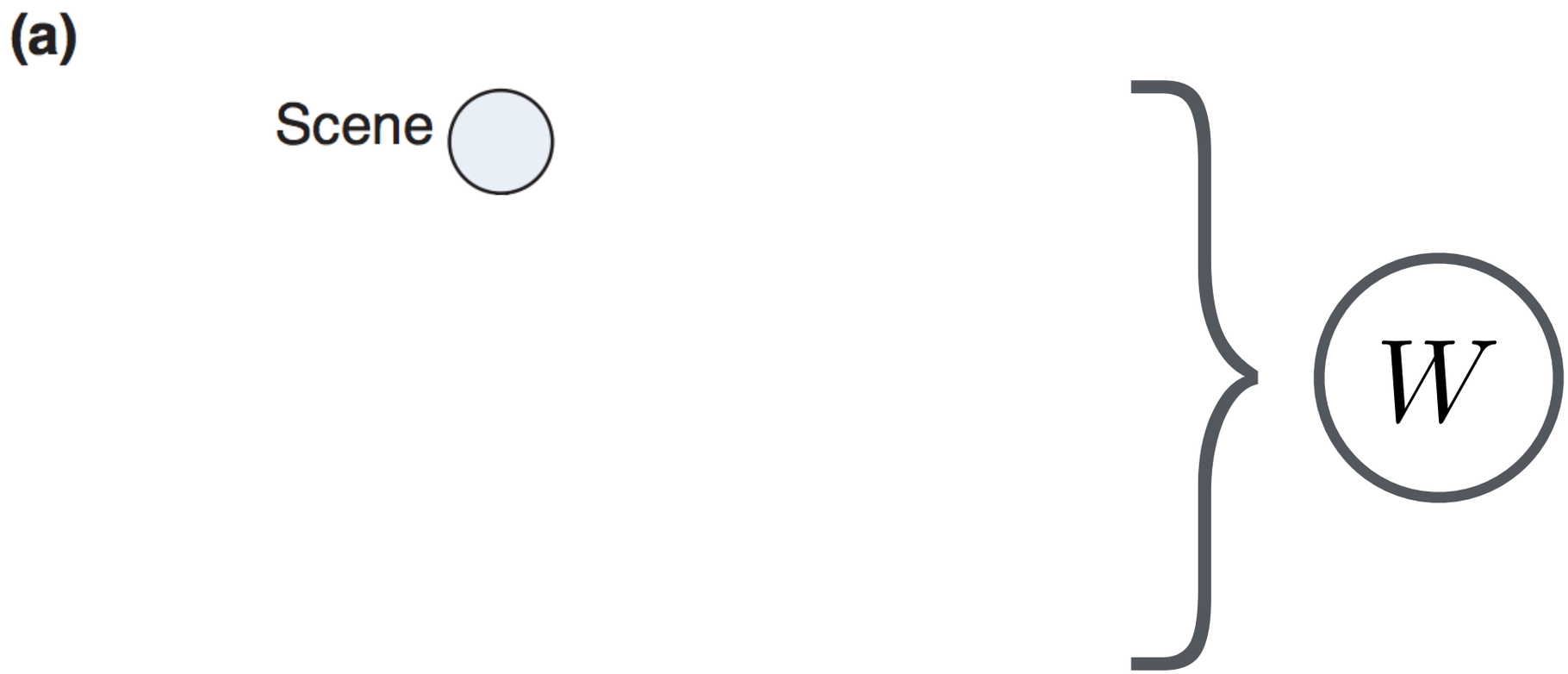
N number of nodes 32

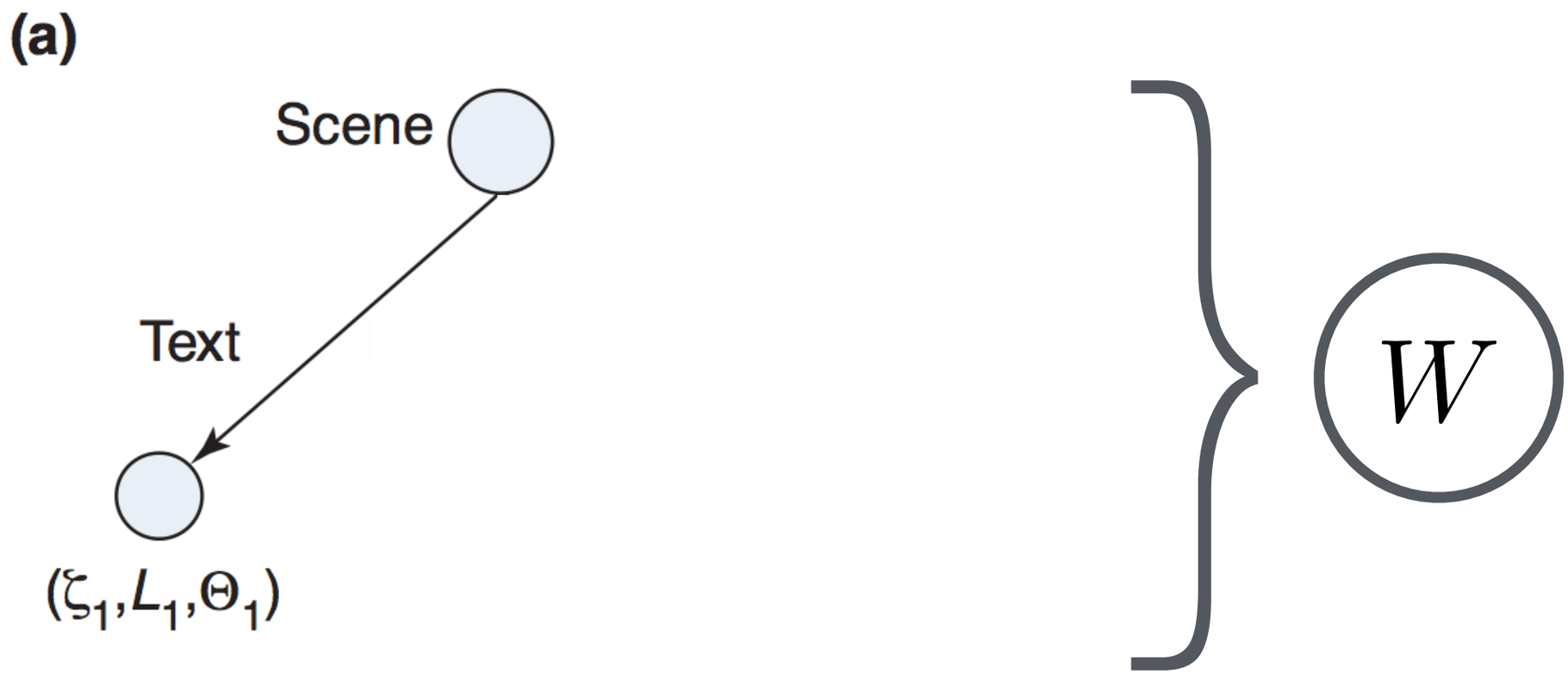


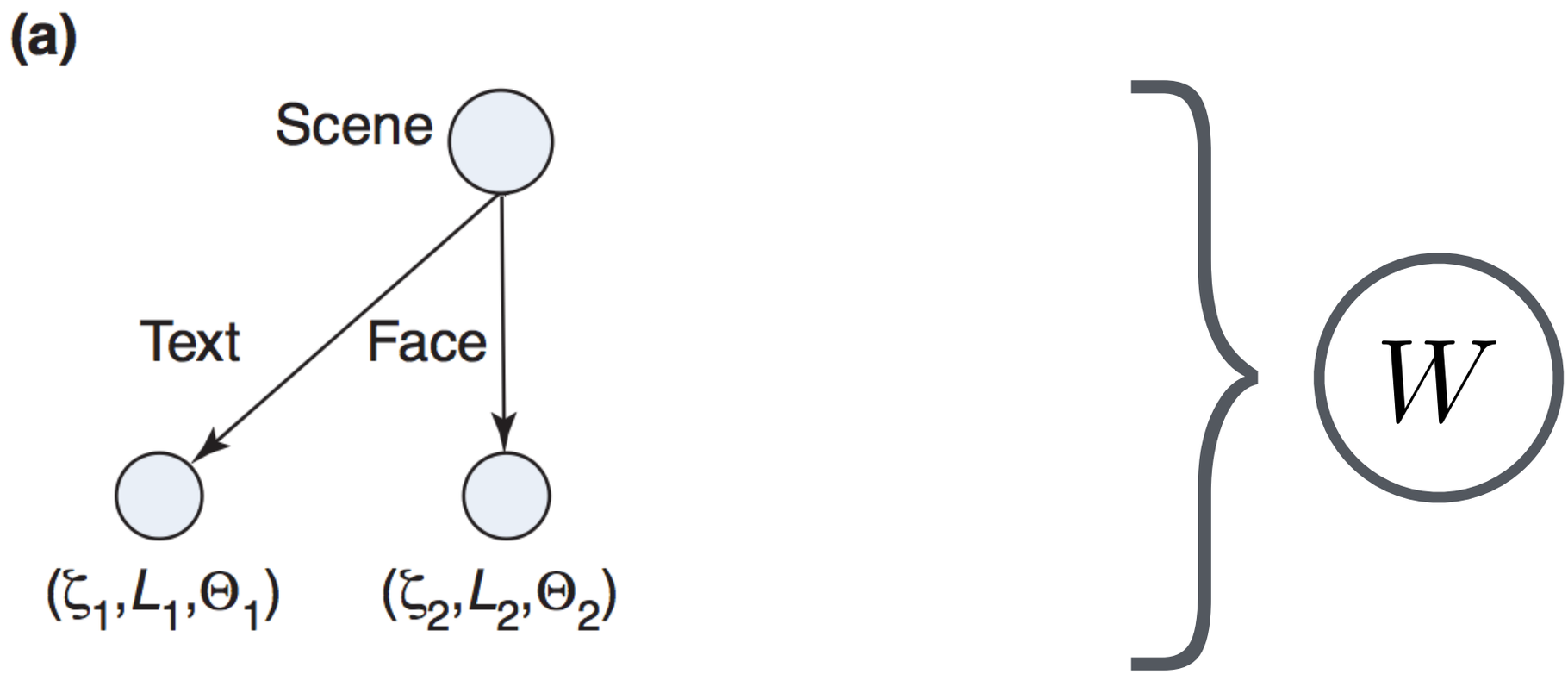
$R(L)$ region

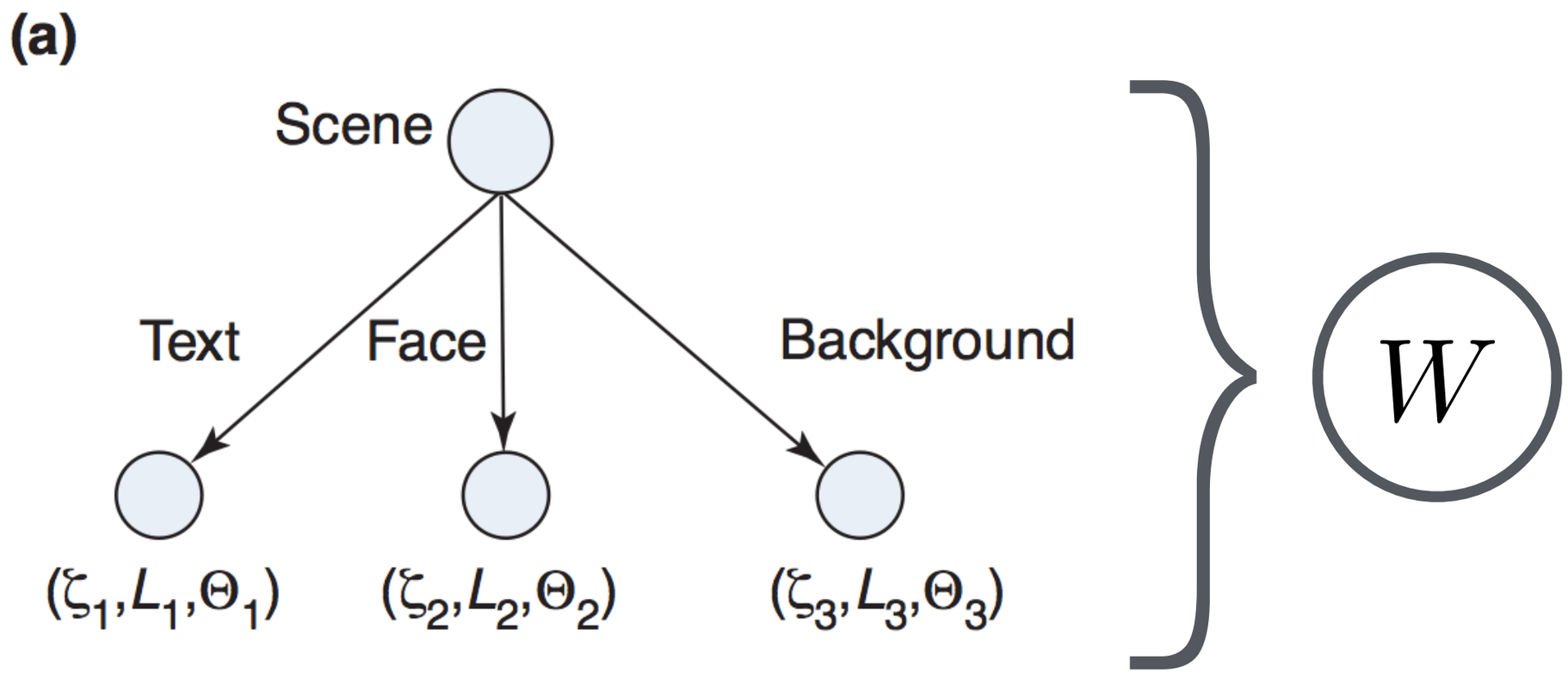
ζ type of model text/face/background

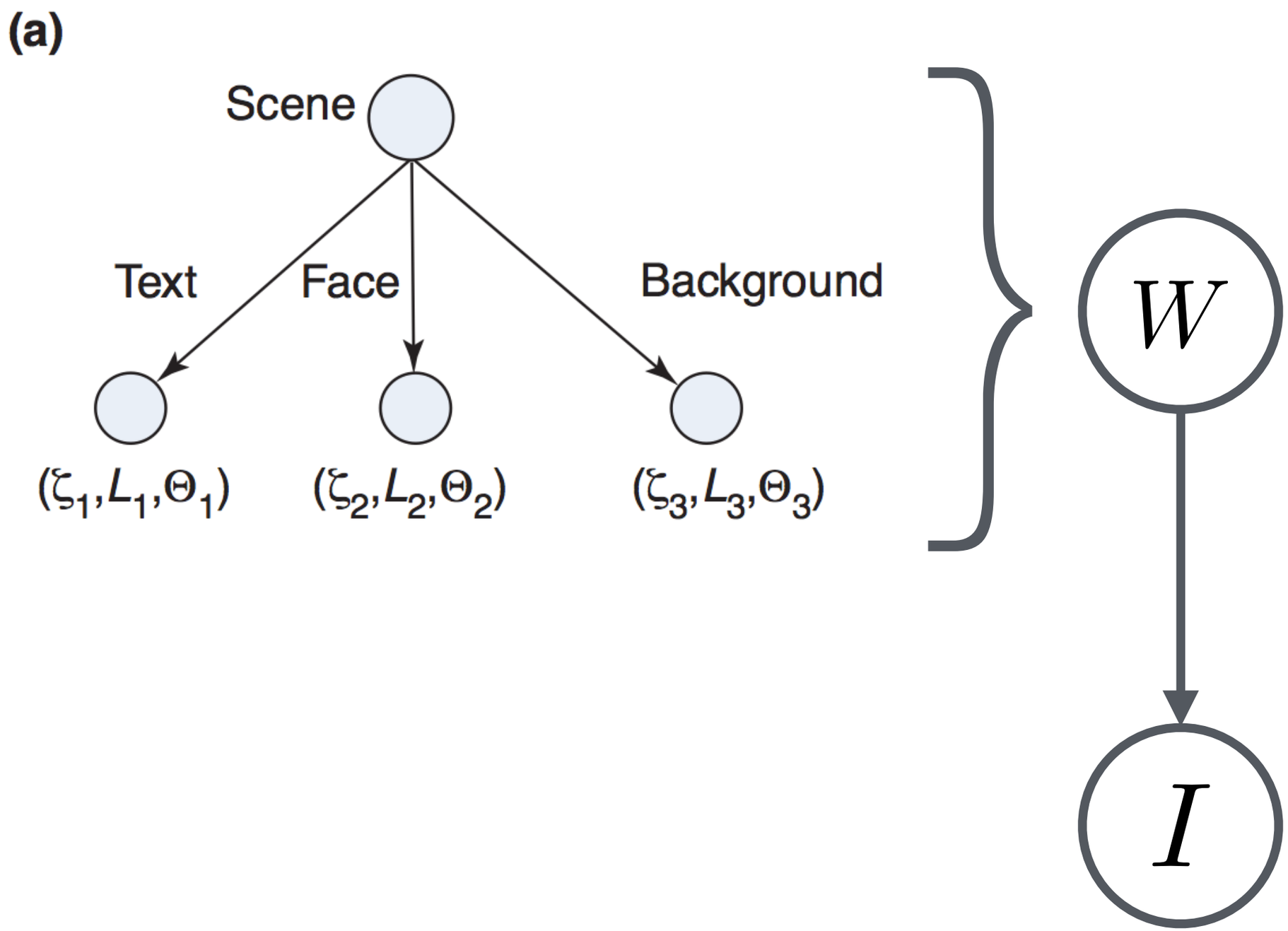
Θ parameters

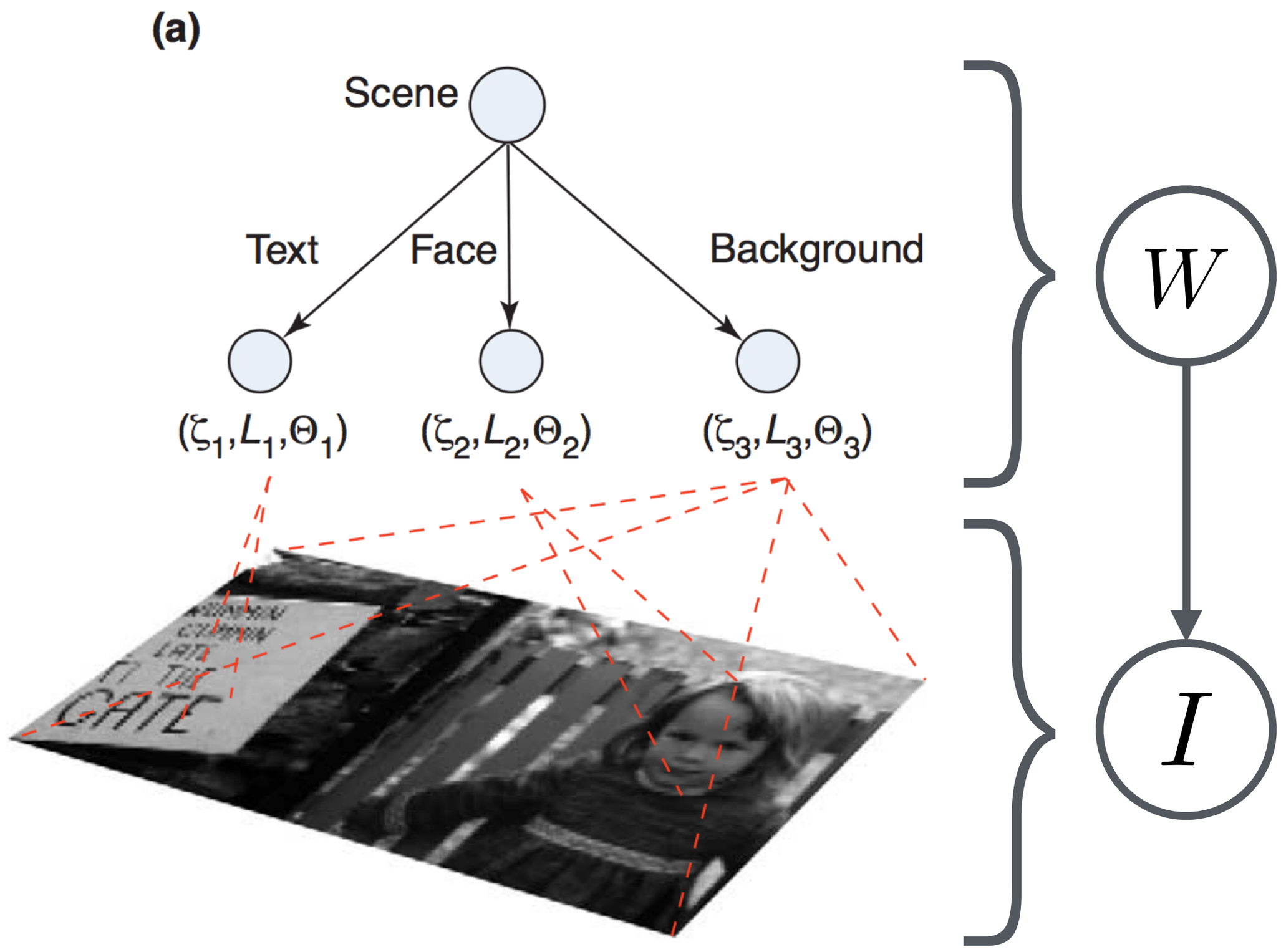

(a)
Scene
W

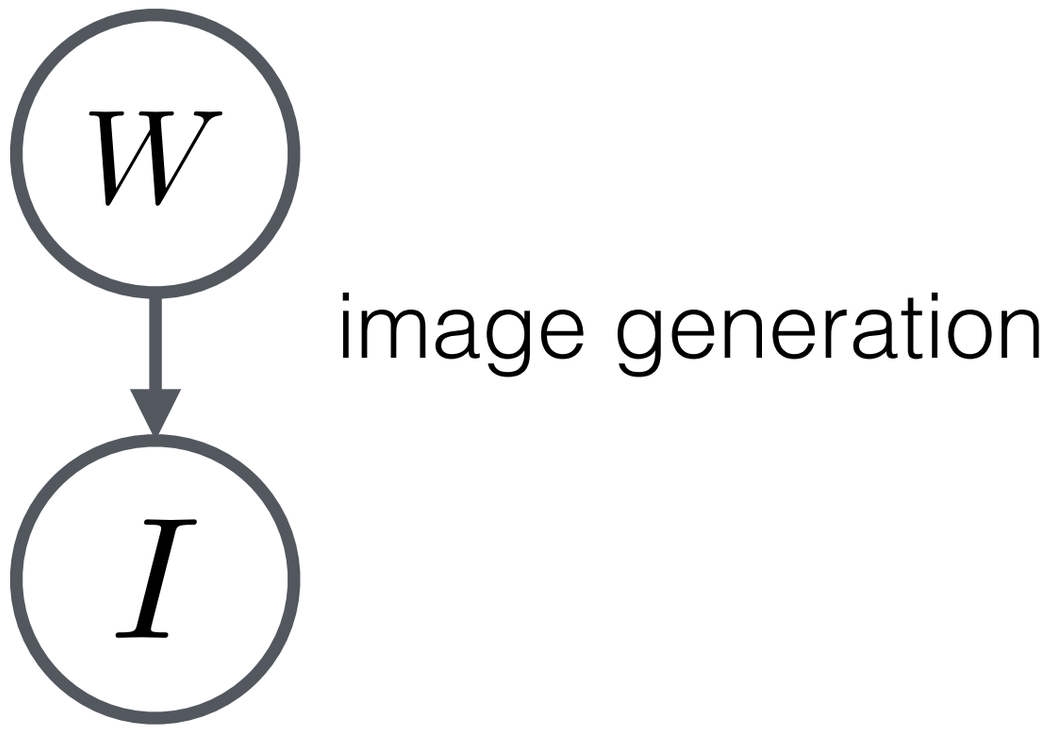

(a)
Scene
Text
(ζ_1, L_1, Θ_1)
W


(a)
Scene
Text
Face
(ζ_1, L_1, Θ_1)
(ζ_2, L_2, Θ_2)
W

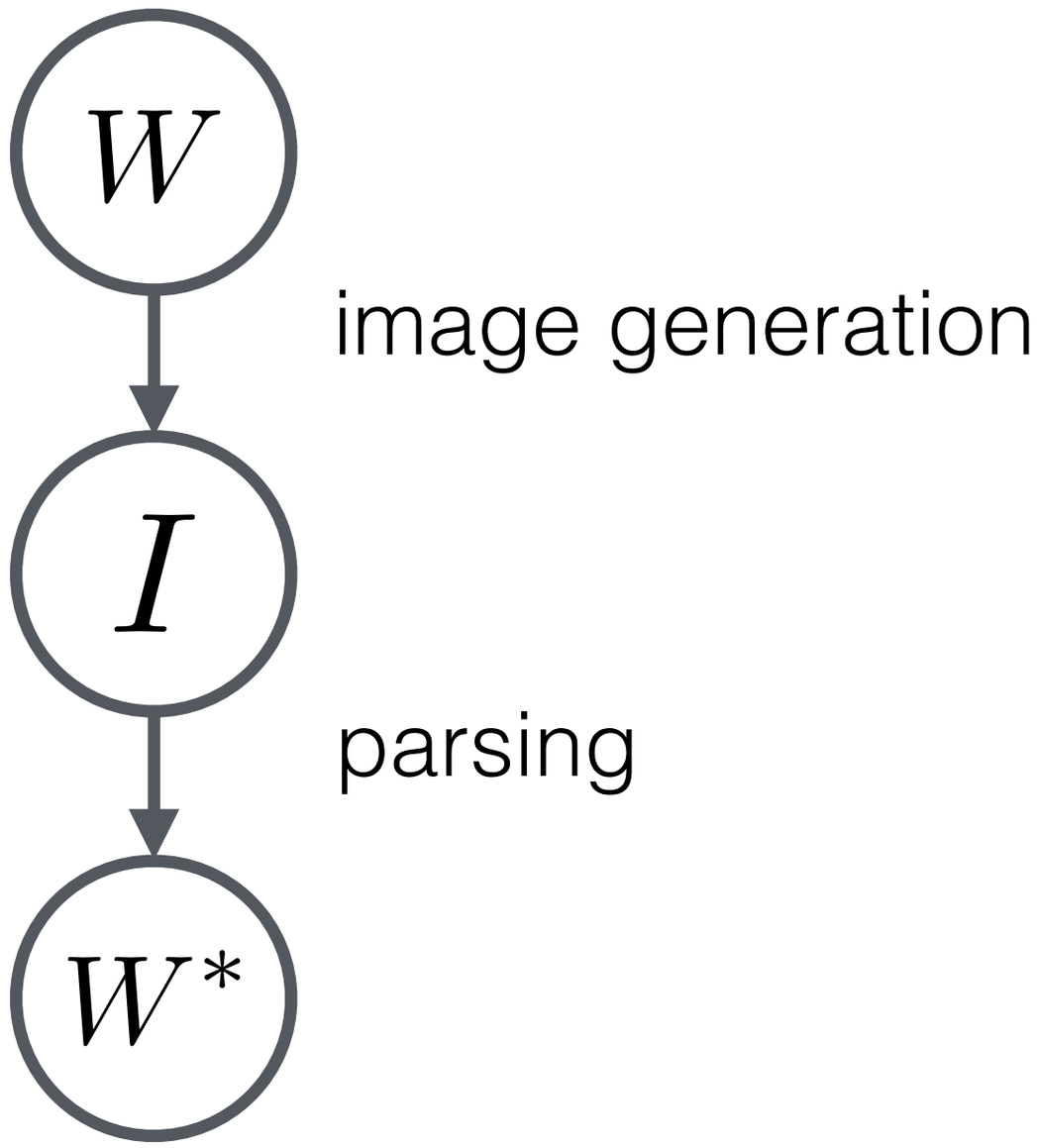

(a)
Scene
Text
Face
Background
(ζ_1, L_1, Θ_1)
(ζ_2, L_2, Θ_2)
(ζ_3, L_3, Θ_3)
W


(a)
Scene
Text
Face
Background
(ζ_1, L_1, Θ_1)
(ζ_2, L_2, Θ_2)
(ζ_3, L_3, Θ_3)
W
I


(a)
Scene
Text
Face
Background
(ζ_1, L_1, Θ_1)
(ζ_2, L_2, Θ_2)
(ζ_3, L_3, Θ_3)
W
I



$P(I|W)$



$P(I|W)$

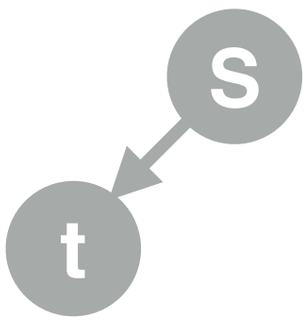
$argmax\ P(W|I)$

inference

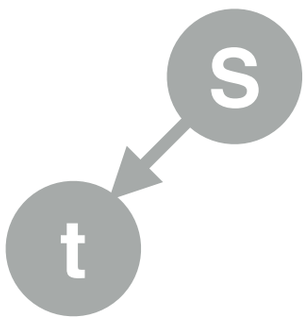
Data driven Markov Chain Monte Carlo

(DDMCMC)

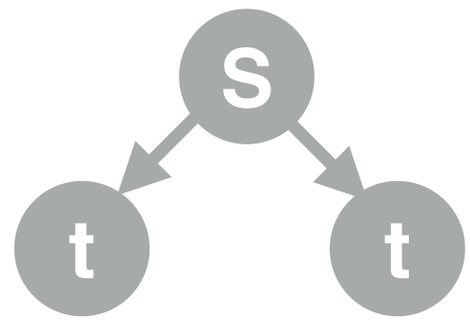
W_1



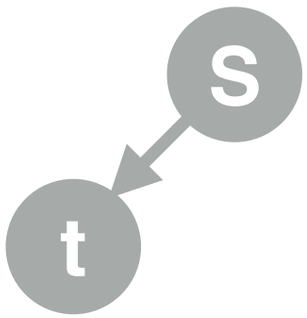
W_1



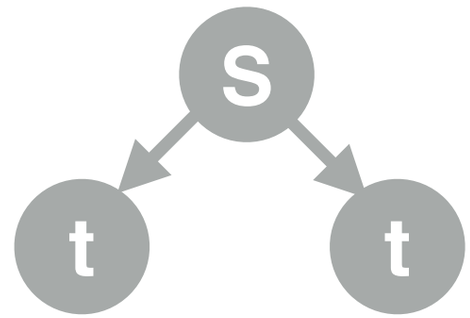
W_2



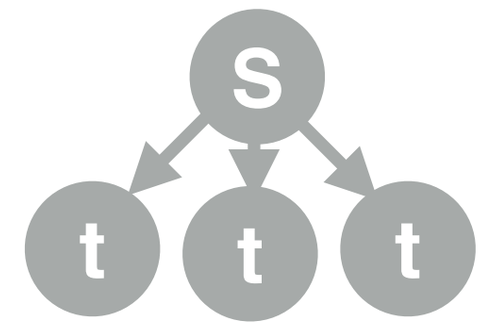
W_1



W_2

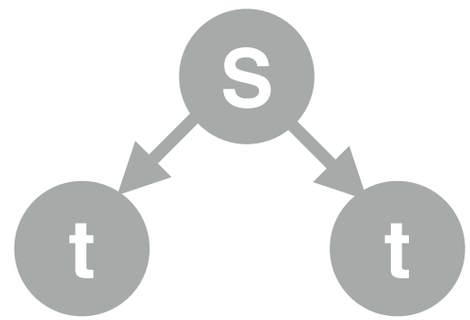


W_3

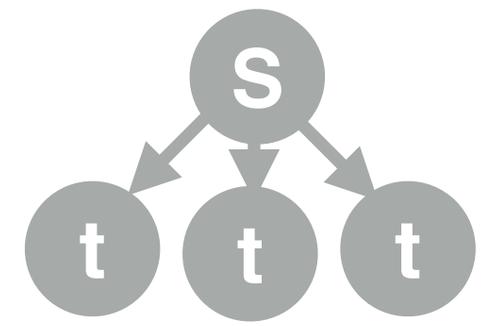


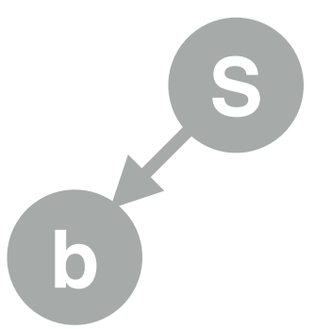
W_1

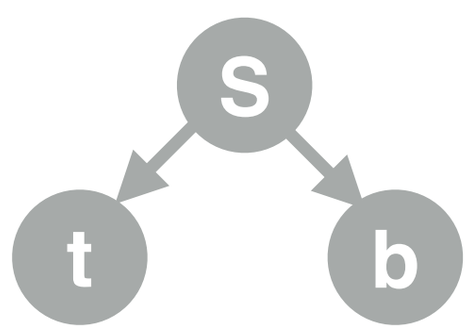
W_2

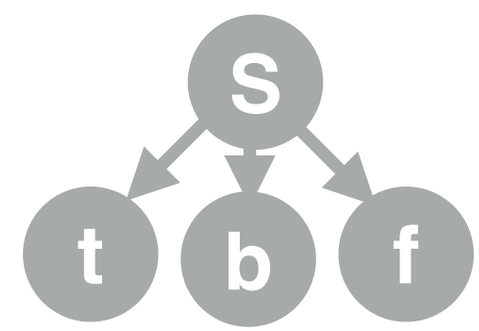


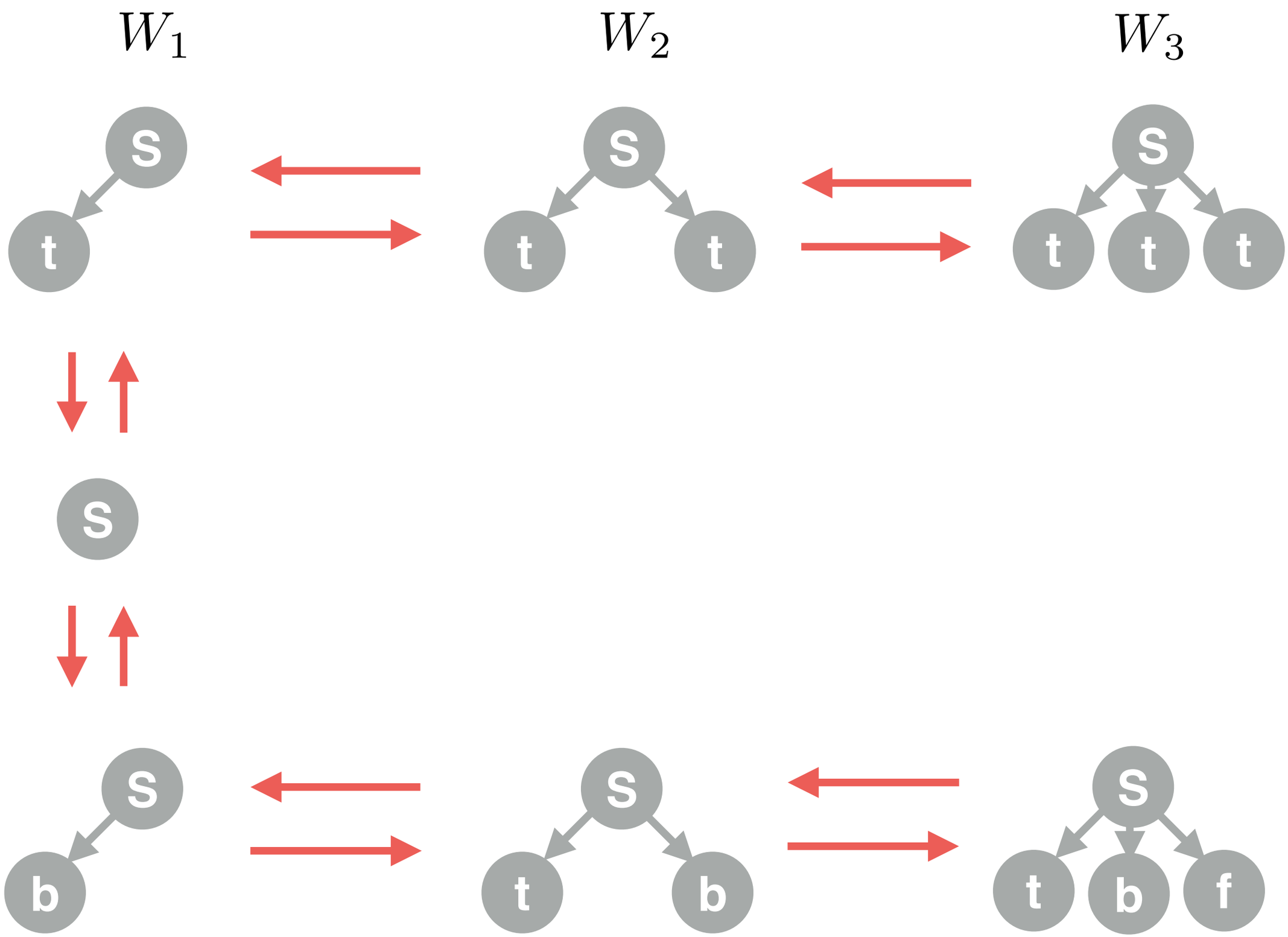
W_3

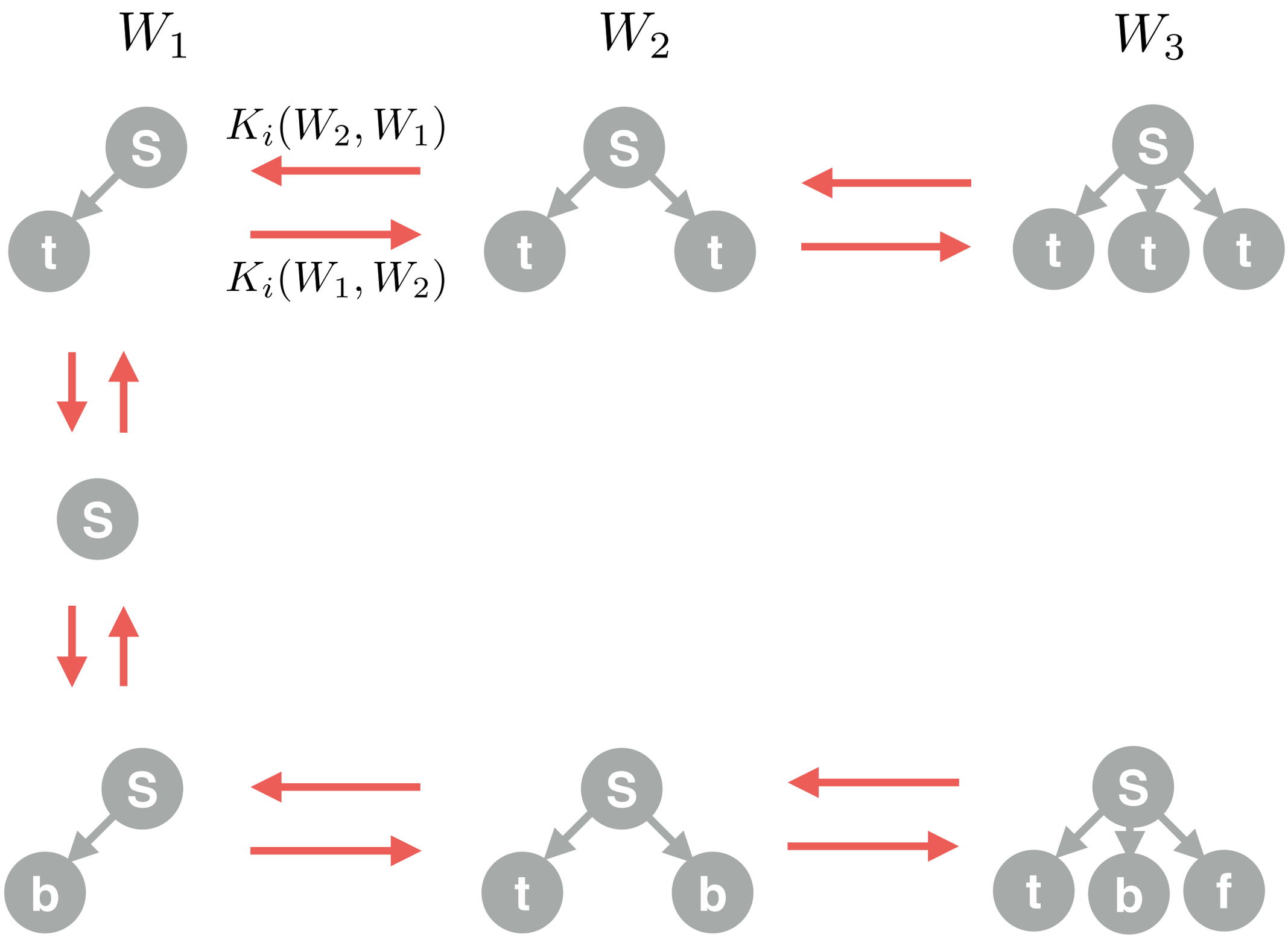


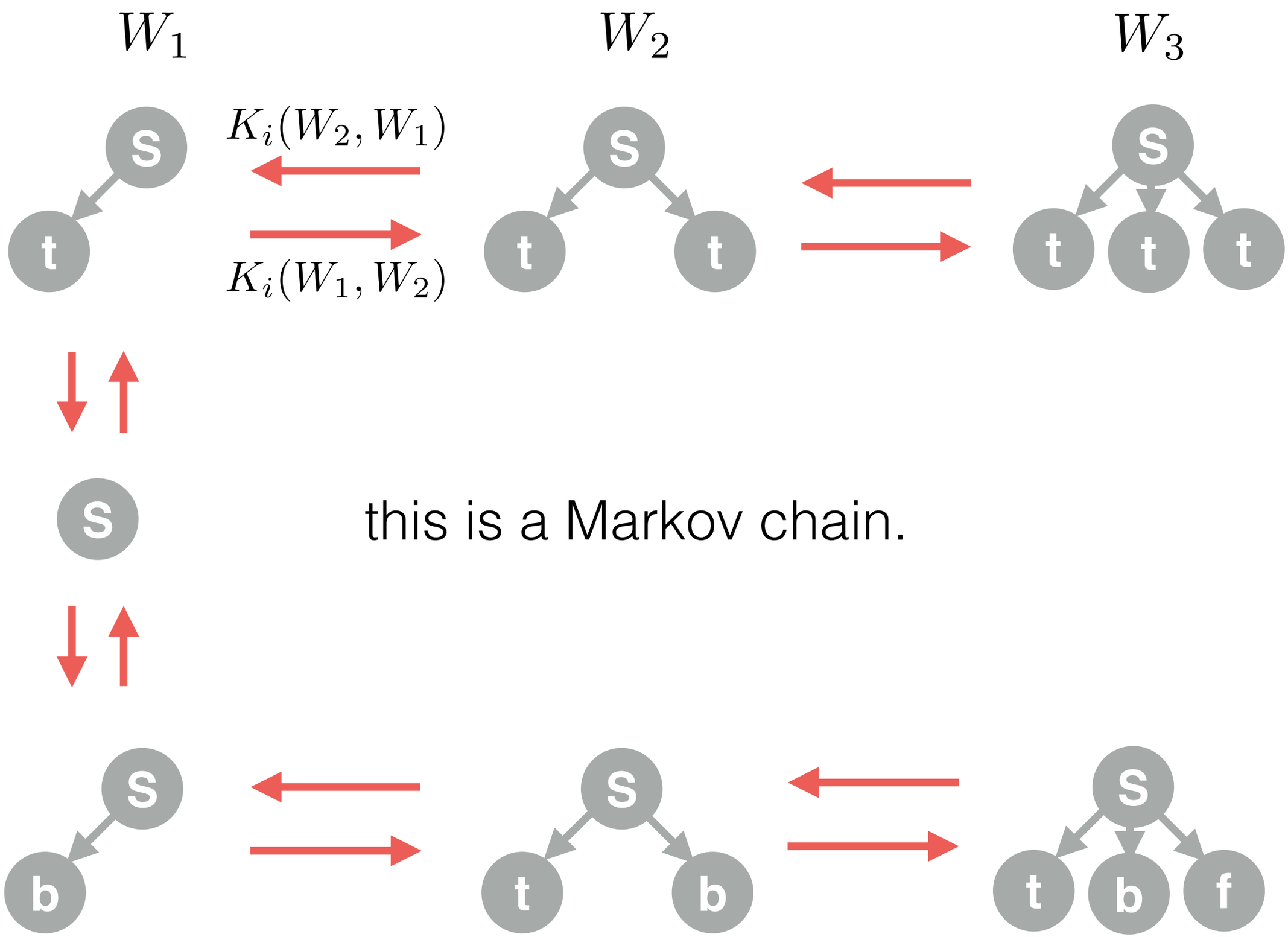




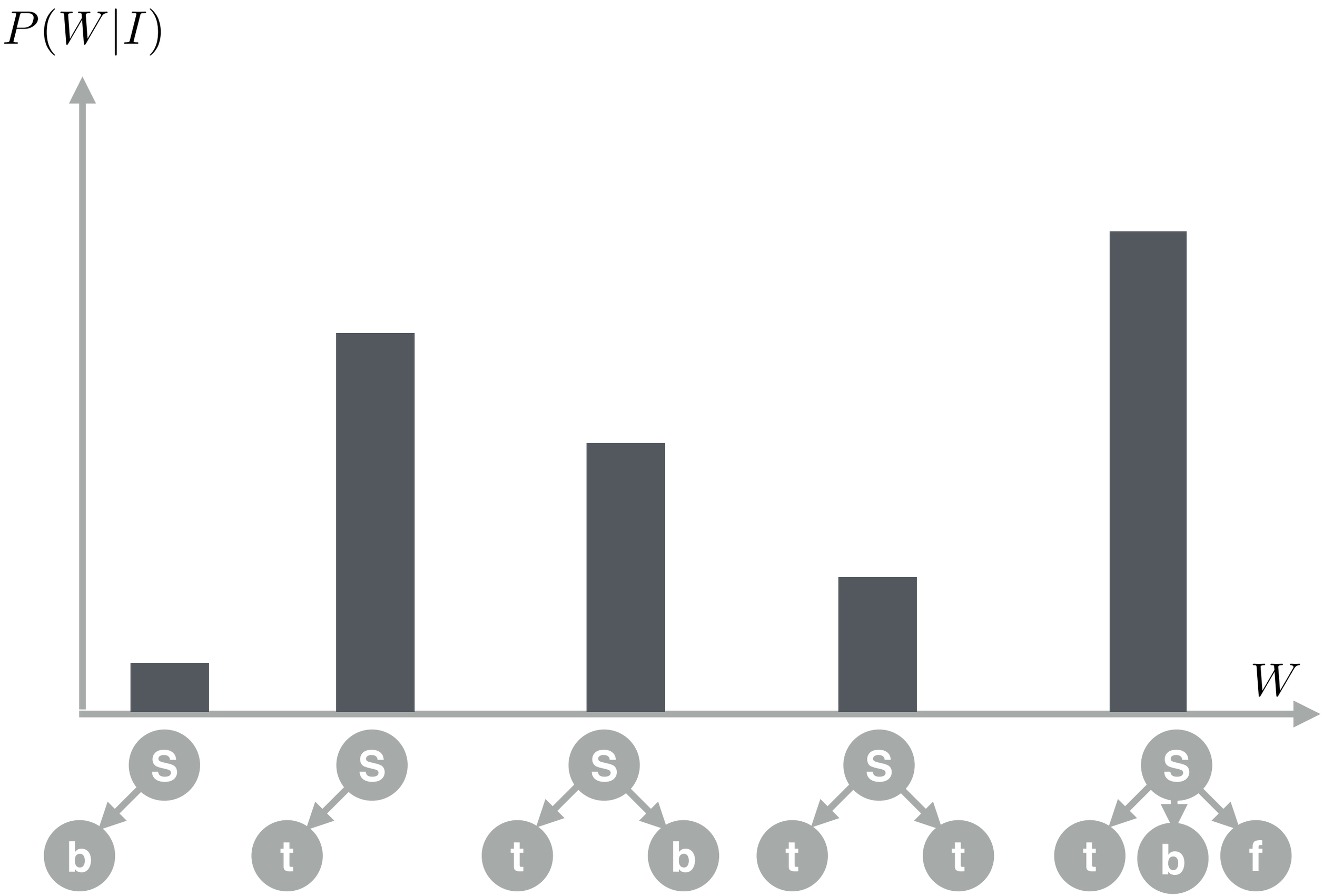


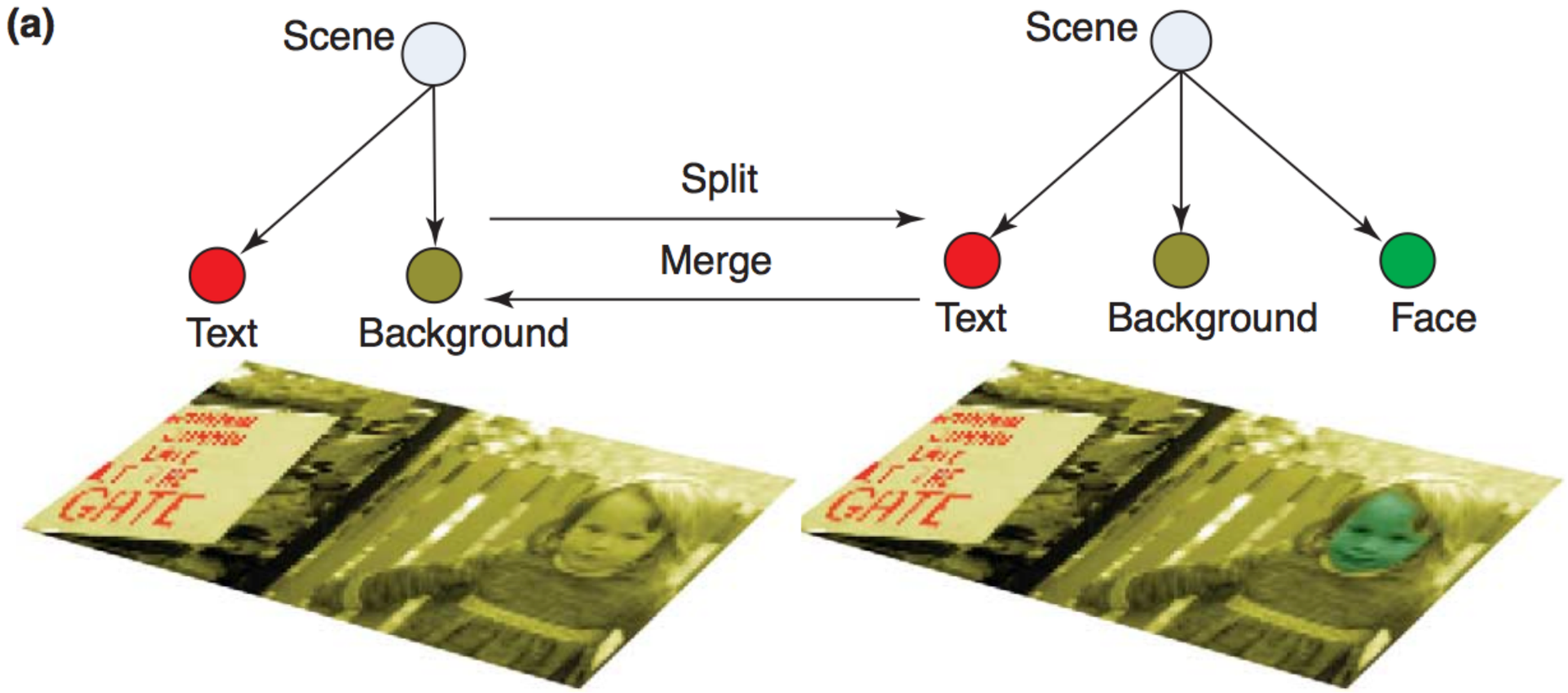

W_1
W_2
W_3
s
t
s
t
t
s
t
t
t
s
s
b
s
t
b
s
t
b
f

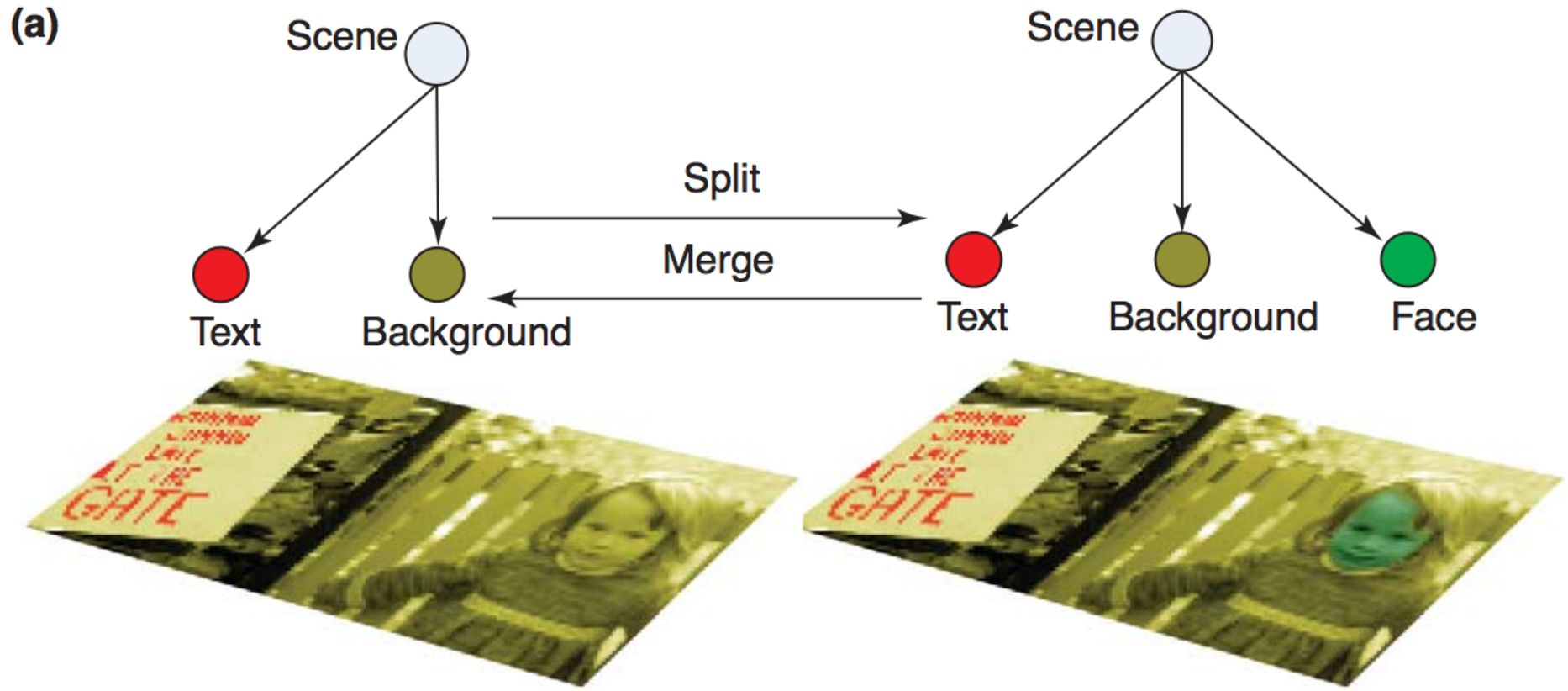

W_1
W_2
W_3
$K_i(W_2, W_1)$
$K_i(W_1, W_2)$
s
t
s
t
t
s
t
t
t
s
s
b
s
t
b
s
t
b
f

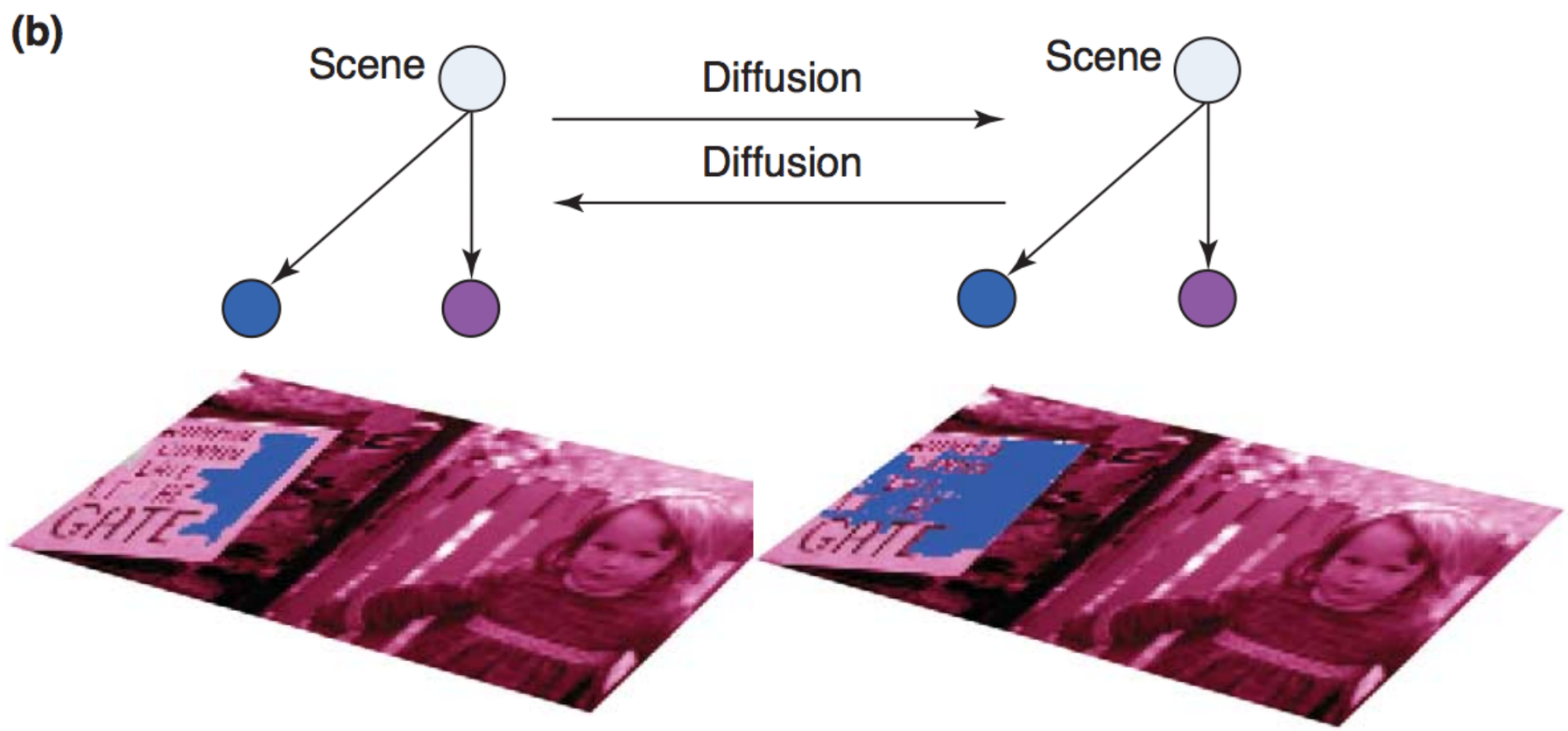


this is a Markov chain.


$P(W|I)$
W
s
b
s
t
s
t
b
s
t
t
s
t
b
f


(a)
Scene
Text
Background
Split
Merge
Scene
Text
Background
Face


(a)
Scene
Text
Background
Split
Merge
Scene
Text
Background
Face


(b)
Scene
Diffusion
Diffusion
Scene

Metropolis-Hastings kernel

$$K_i(W_1, W_2) = q_i(W_1, W_2)\, a_i(W_1, W_2)$$

$$K_i(W_1, W_2) = \overbrace{q_i(W_1, W_2)}^{\text{“bottom-up proposals” proposal distribution}} a_i(W_1, W_2)$$

$$K_i(W_1, W_2) = \overbrace{q_i(W_1, W_2)}^{\text{“bottom-up proposals” proposal distribution}} \underbrace{a_i(W_1, W_2)}_{\text{acceptance probability “high-level model”}}$$

original image

original image

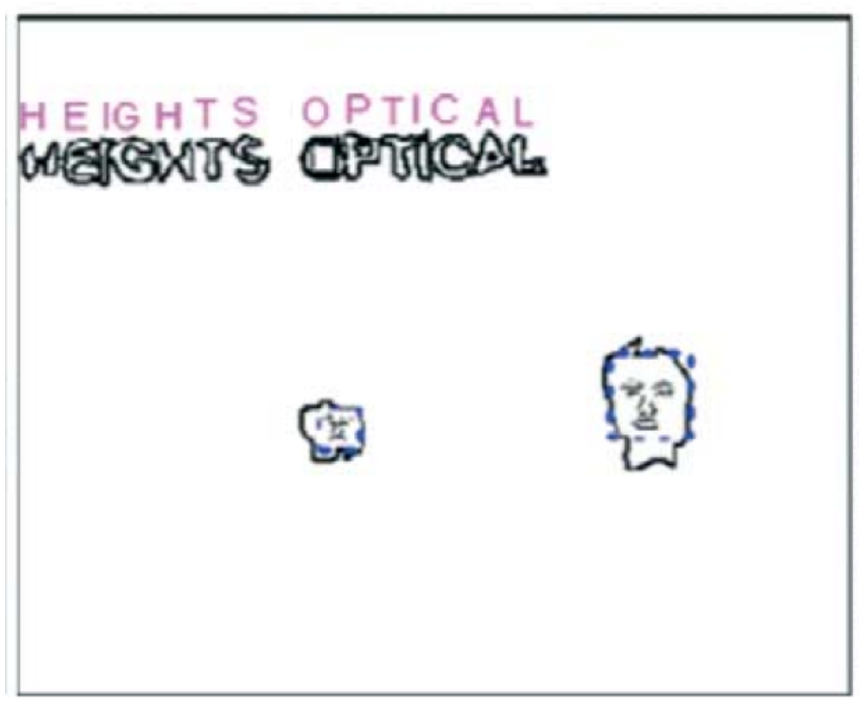
bottom-up proposals

note the tree-face!

original image

bottom-up proposals

note the tree-face!

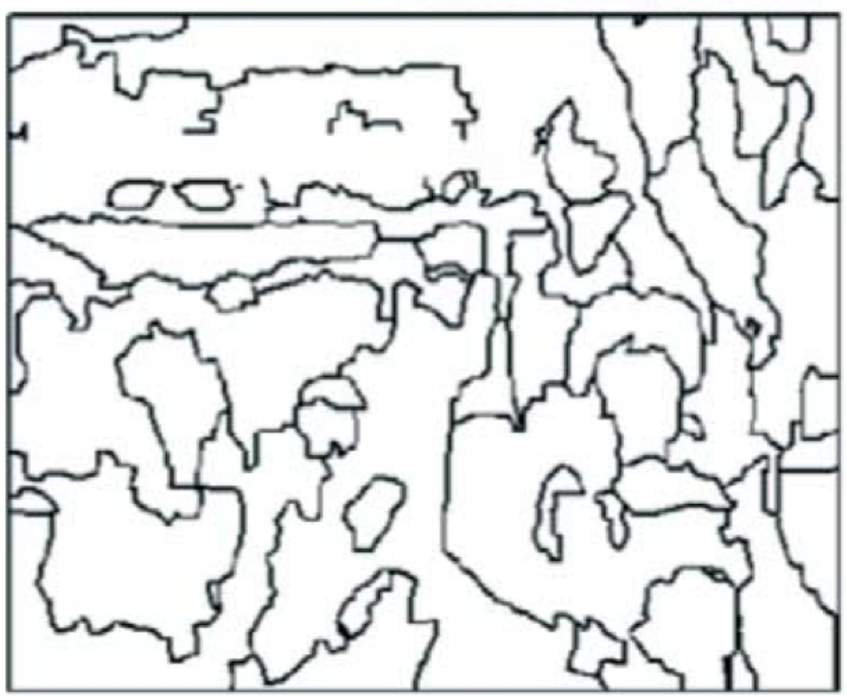


detection

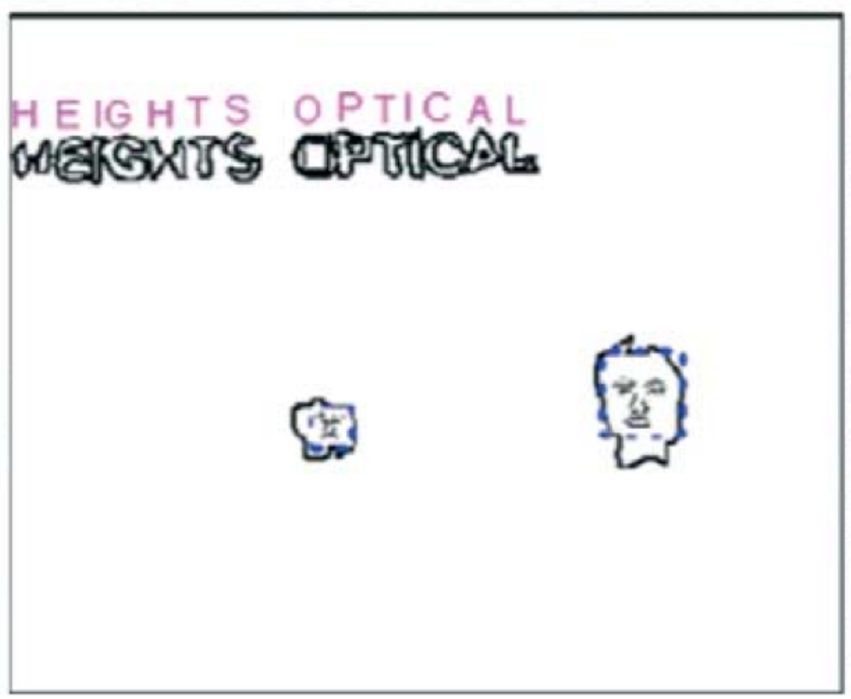
original image

bottom-up proposals

note the tree-face!



segmentation

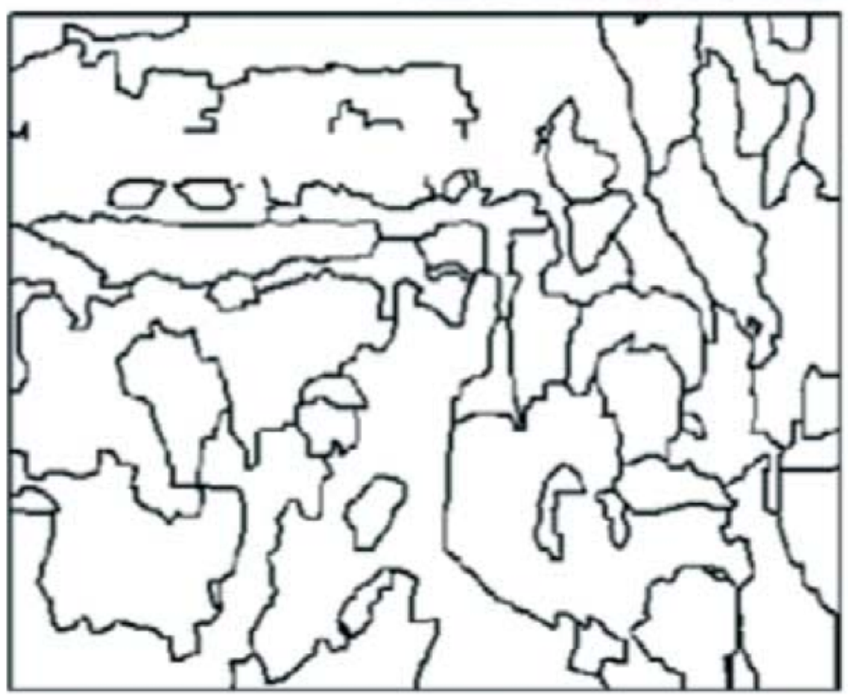


detection

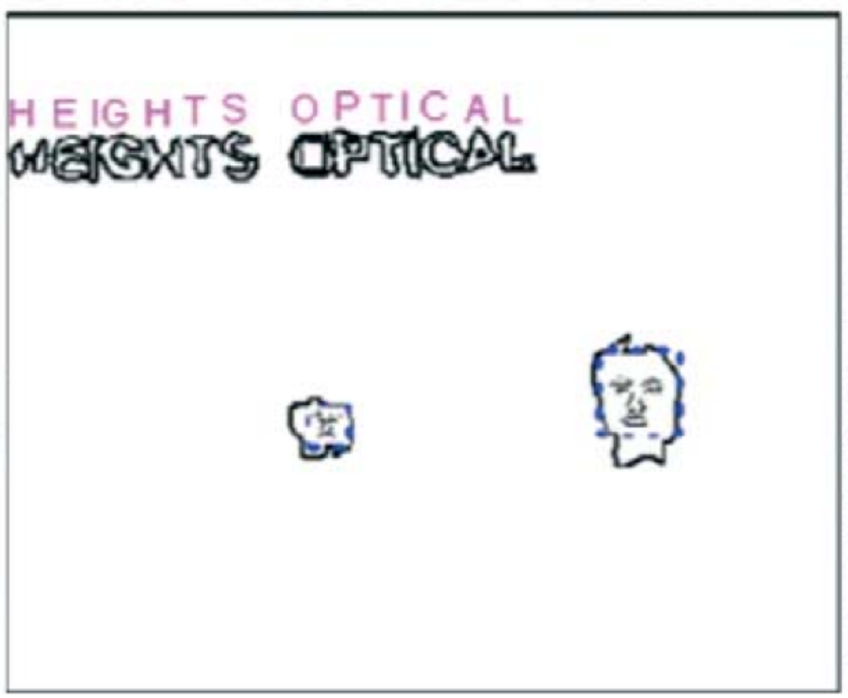
original image

bottom-up proposals

note the tree-face!



segmentation



detection

synthesised image

conclusions

- major goal in vision should be to determine how visual system works under natural conditions and natural tasks

conclusions

- major goal in vision should be to determine how visual system works under natural conditions and natural tasks

- probability distributions over structured representations offer the promise to model natural images

conclusions

- major goal in vision should be to determine how visual system works under natural conditions and natural tasks
- probability distributions over structured representations offer the promise to model natural images
- DDMCMC inference algorithm follows “analysis by synthesis” strategy, which may correspond to forward and backward pathways in the cortex