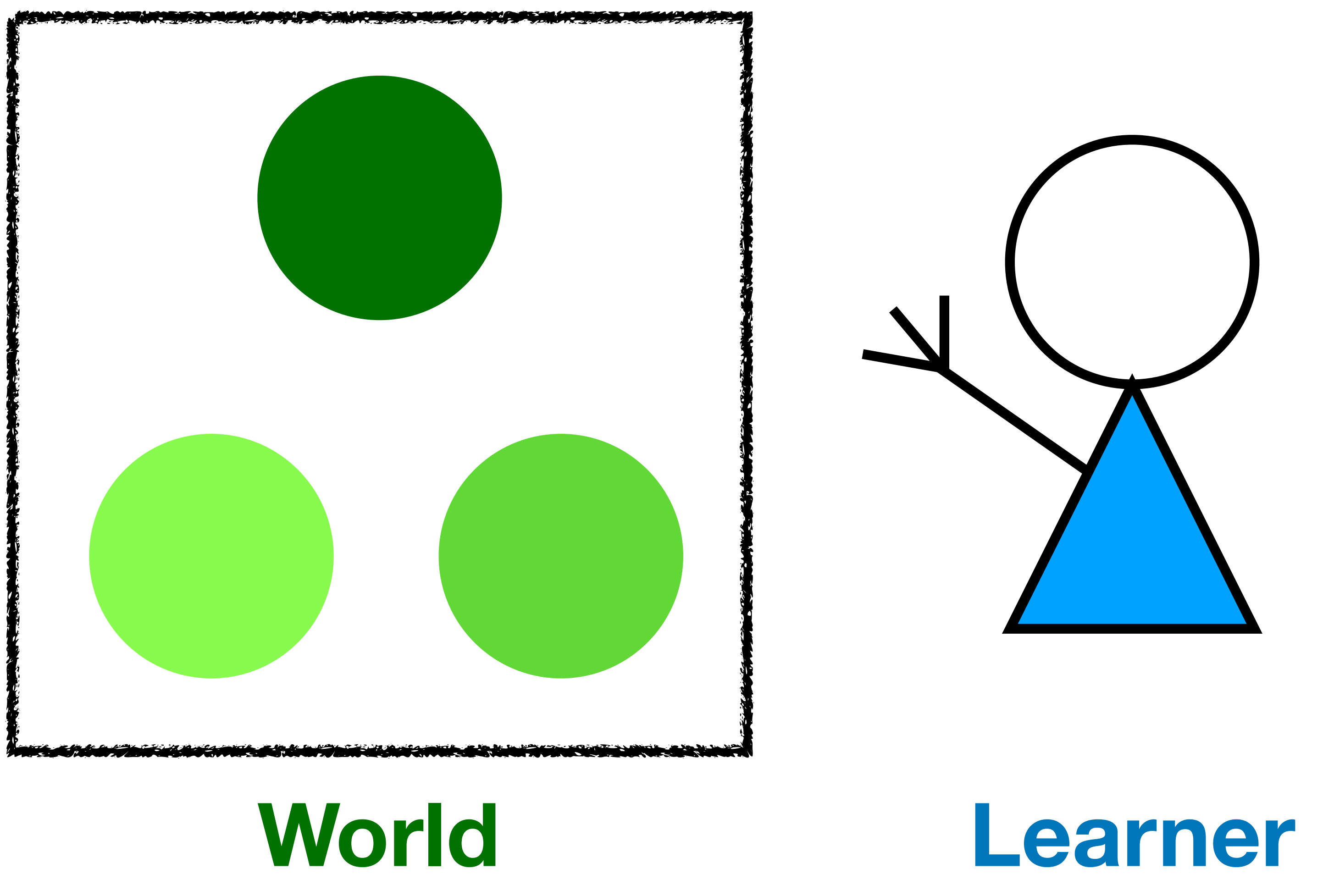
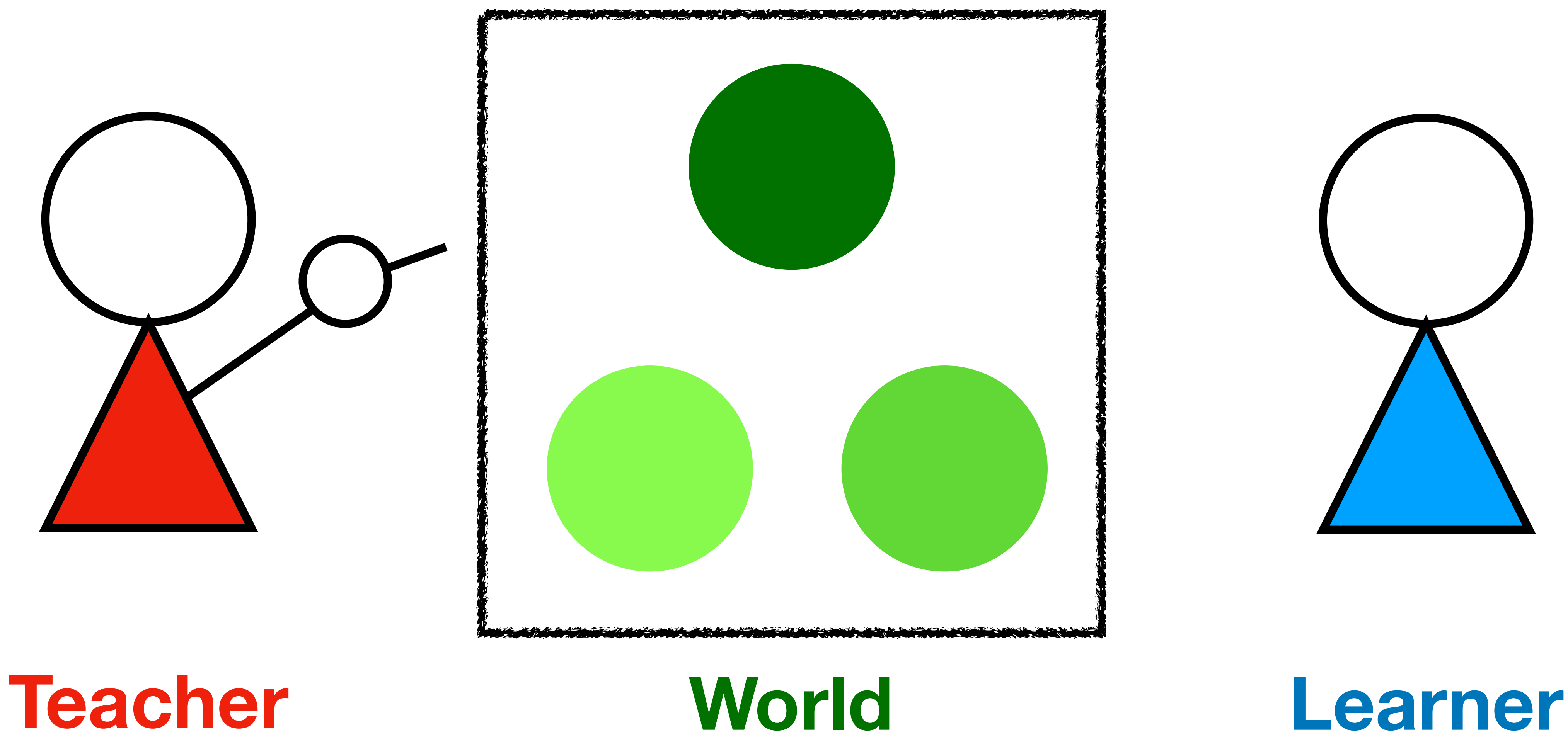
A unifying computational framework for teaching and active learning

Scott Cheng-Hsin Yang, Wai Keen Vong,
Yue Yu & Patrick Shafto

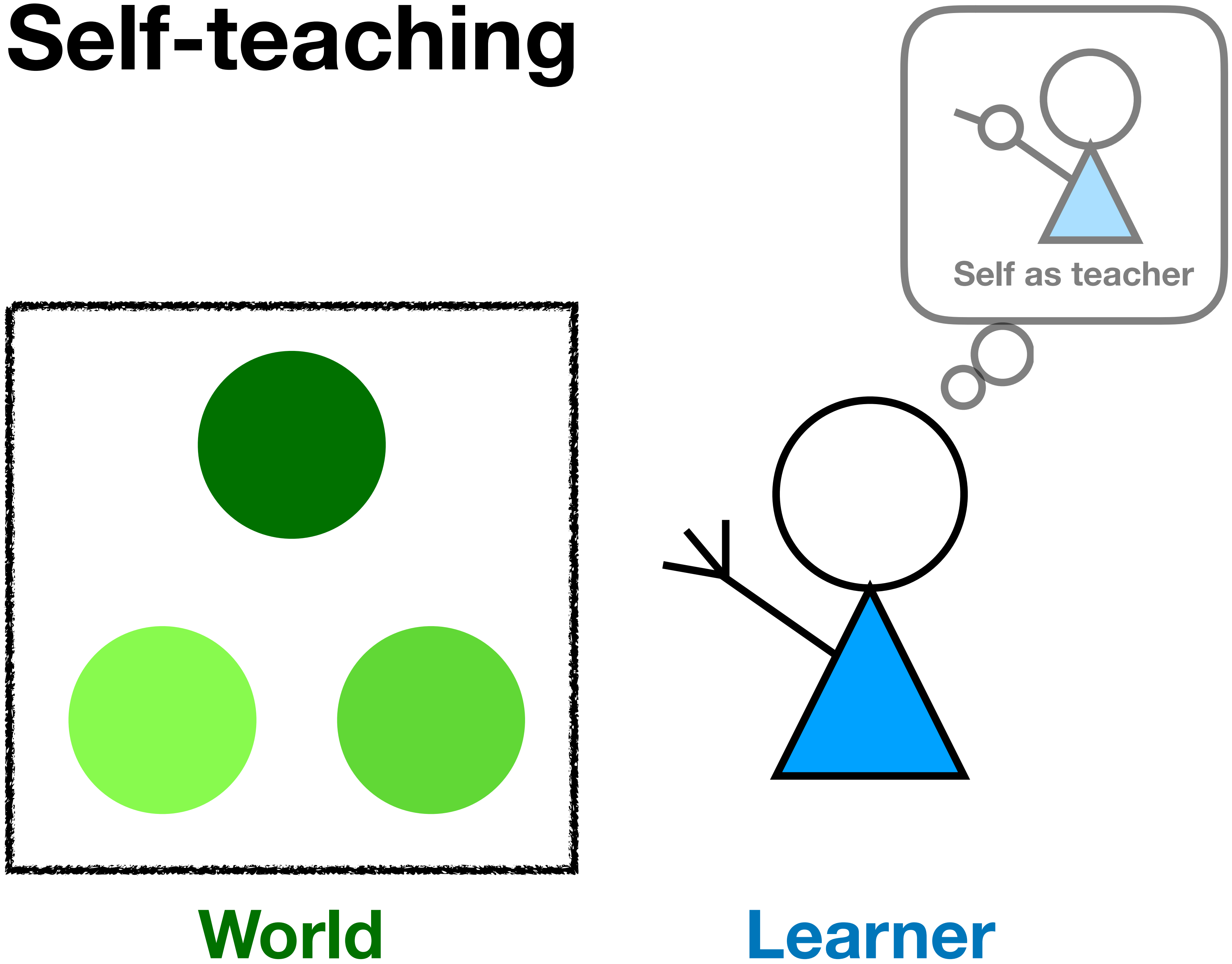
Active learning



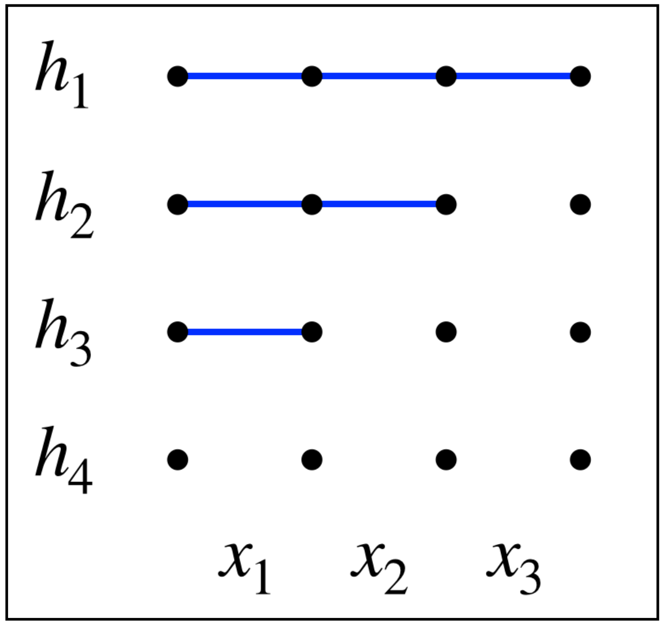
Teaching

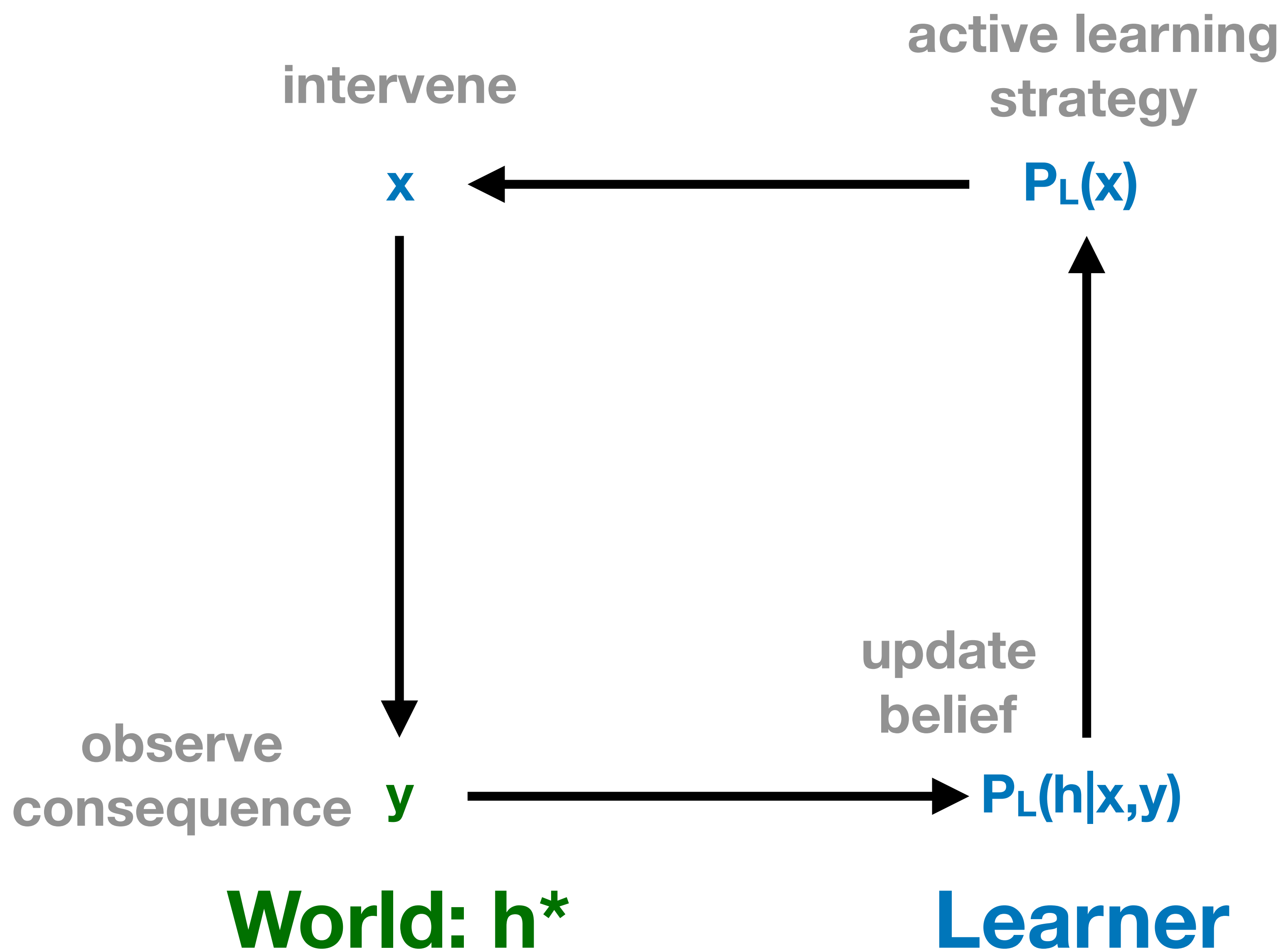


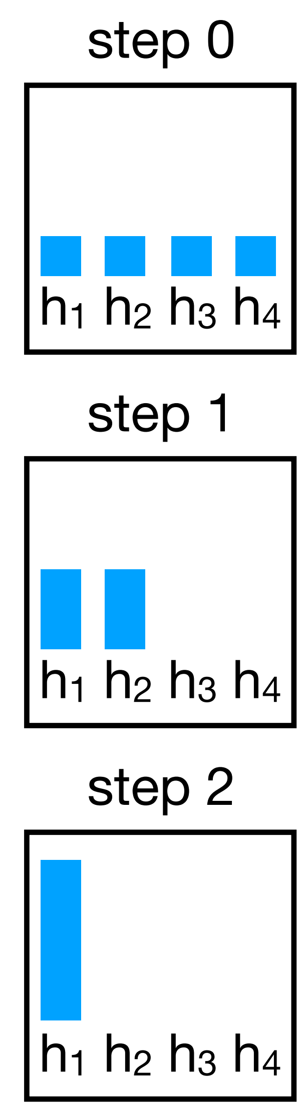
Self-teaching



Active learning







Teaching

teaching strategy

$P_T(x,y|h^*)$

show

x,y

active learning strategy

$P_L(x)$

Teacher knows y and h*; learner does not.

update belief

$P_L(h|x,y)$

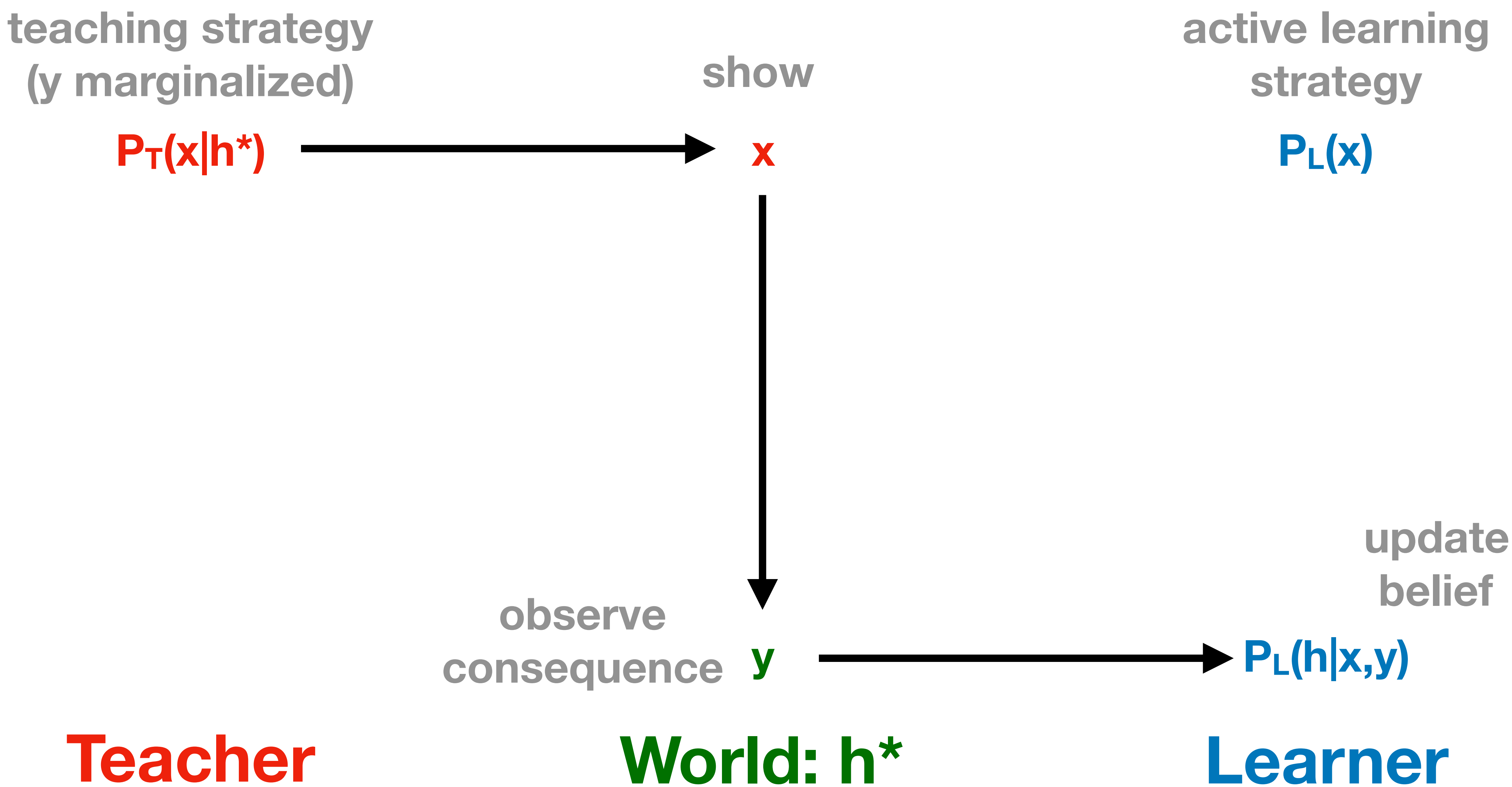
Teacher **World: h*** **Learner**

learner's inference $P_L(h|x,y) \propto P_T(x,y|h) P_L(h)$

teacher's selection $P_T(x,y|h) \propto P_L(h|x,y) P_T(x,y)$

Shafto et al. 2008, 2014

Teaching (marginalize out y)



learner's inference $$\colorbox{lightblue}{$P_L(h|x,y)$} \propto P(y|x,h)\,P_T(x|h)\,P_L(h)$$

teacher's selection $$P_T(x|h) = \sum_{y \in \mathcal{Y}} P_T(x,y|h)$$

Yang & Shafto 2017

Knowledgeability (marginalize out "h")

teaching strategy
(y marginalized)

$P_T(x|h^*)$

teaching strategy
(y & h* marginalized) = active learning strategy

$P_T(x) = P_L(x)$

$\delta(g|h)$: truth

teacher's belief

	h_1	h_2	h_3	h_4
g_1	1	0	0	0
g_2	0	1	0	0
g_3	0	0	1	0
g_4	0	0	0	1

Teacher

World: h*

$\delta_{ST}(g|h) = P_L(h)$: truth

learner's belief

	h_1	h_2	h_3	h_4
g_1	1/4	1/4	1/4	1/4
g_2	1/4	1/4	1/4	1/4
g_3	1/4	1/4	1/4	1/4
g_4	1/4	1/4	1/4	1/4

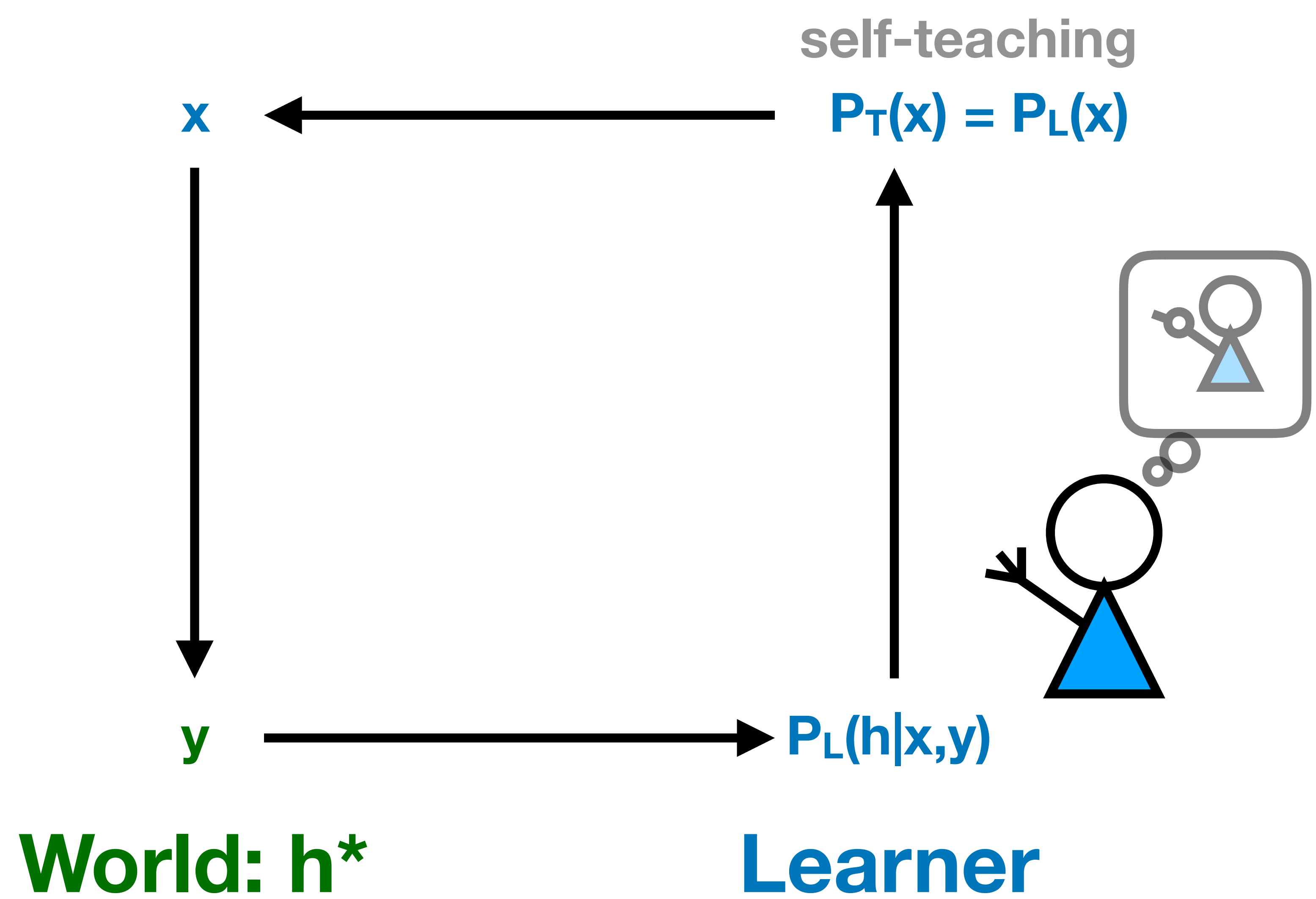
Learner

$$P_T(x|h) = \sum_{g \in \mathcal{H}} P_T(x|g)\delta(g|h)$$

$$P_T(x) = \sum_{g \in \mathcal{H}} P_T(x|g)P_L(g)$$

Shafto, Eaves, et al. 2012

Self-teaching



learner's inference

$$P_L(h|x,y) = \frac{P(y|x,h)\cancel{P_T(x)}P_L(h)}{\sum_{h'\in\mathcal{H}} P\left(y|x,h'\right)\cancel{P_T(x)}P_L\left(h'\right)}$$

self-teacher's selection

$$P_T(x) = \sum_{g\in\mathcal{H}} P_T(x|g)P_L(g)$$

How is the Self-Teaching model different from the most common model of active learning objective —optimizing for expected information gain?

Does the Self-Teaching model capture human's active learning behavior?

Self-Teaching

$$P_T(x) = \sum_{g \in H} \sum_{y \in Y} \boxed{\frac{P_L(g|x,y)P_T(x,y)}{Z(g)}} P_L(g)$$

- Uses only the rules of probability
- Meta-reasons about oneself as the teacher
- Hypothesis testing for distinctive hypothesis

Expected information gain

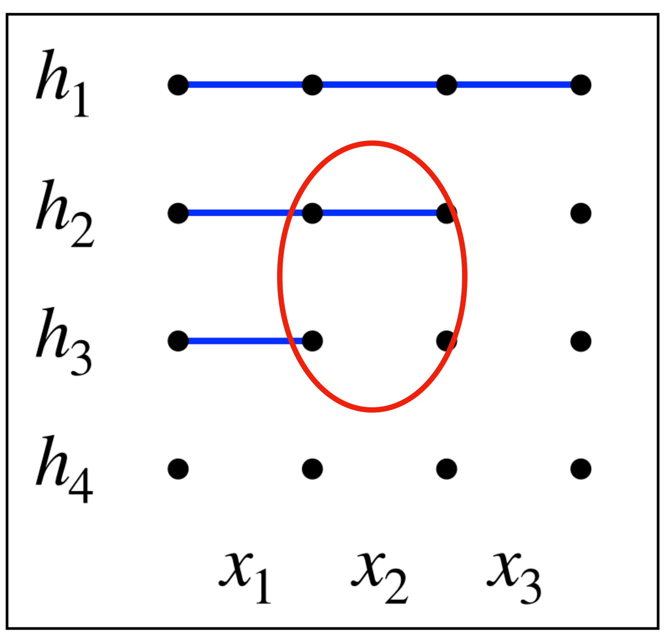
$$EIG(x) = H(h) - \sum_{y \in Y} \boxed{P_L(y|x)} H(h|x,y)$$

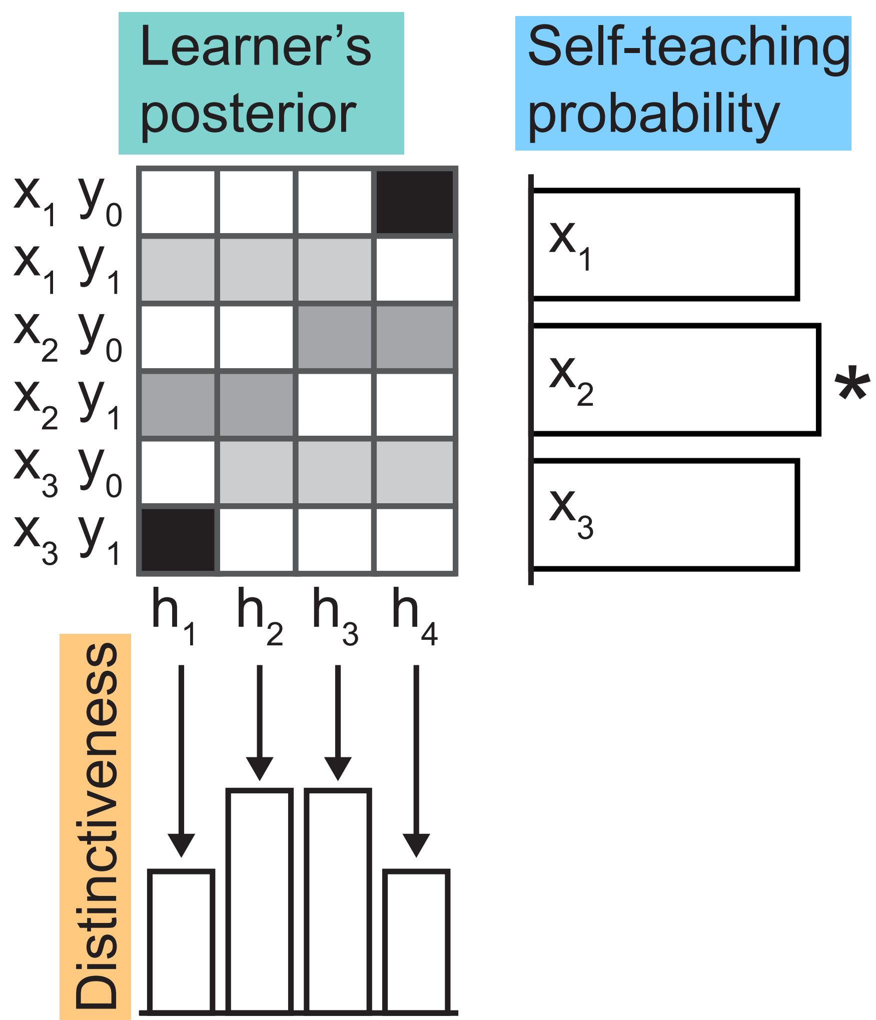
- Also uses entropy and subtraction
- Reasons about the world
- Overall uncertainty reduction

Self-teaching: confirming distinctive h

$$P_T(x) = \sum_{g \in \mathcal{H}} P_T(x|g) P_L(g) = \sum_{g \in \mathcal{H}} \sum_{y \in \mathcal{Y}} P_L(g|x,y) P_T(x,y) P_L(g) Z(g)^{-1}$$

$$Z(g) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_L(g|x,y) P_T(x,y)$$



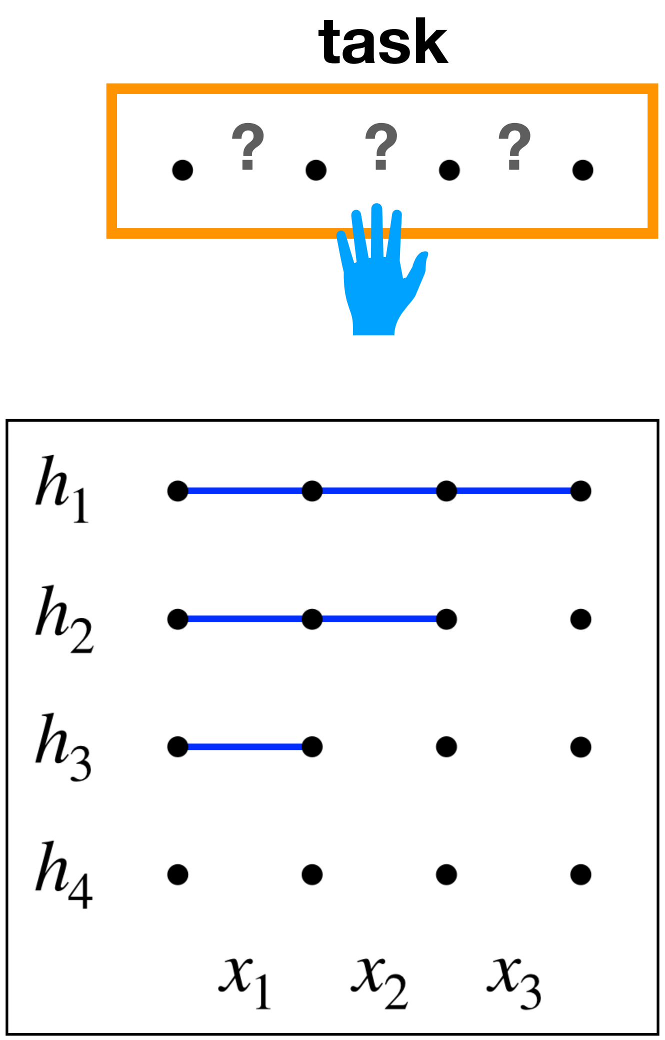


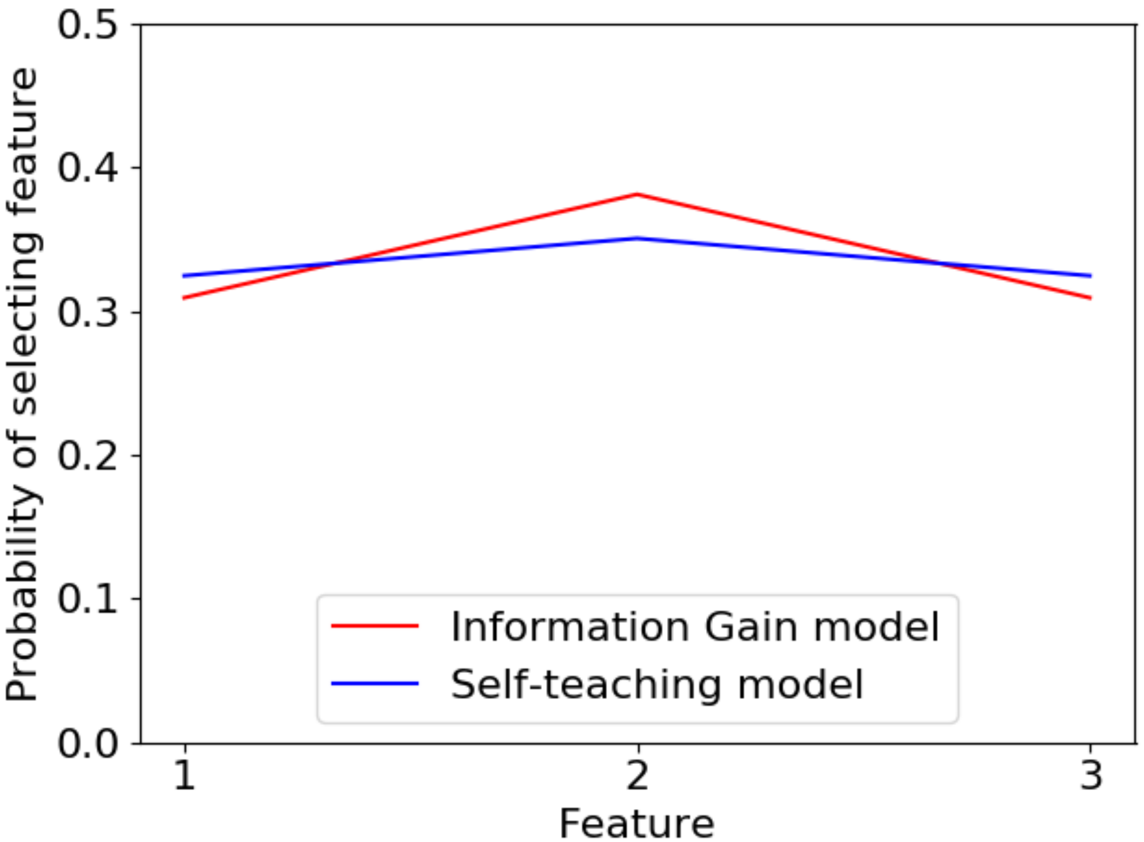
A distinctive hypothesis is one that is on average less likely to be inferred if all interventions and observations are equally likely to occur.

How is the Self-Teaching model different from the most common model of active learning objective —optimizing for expected information gain?

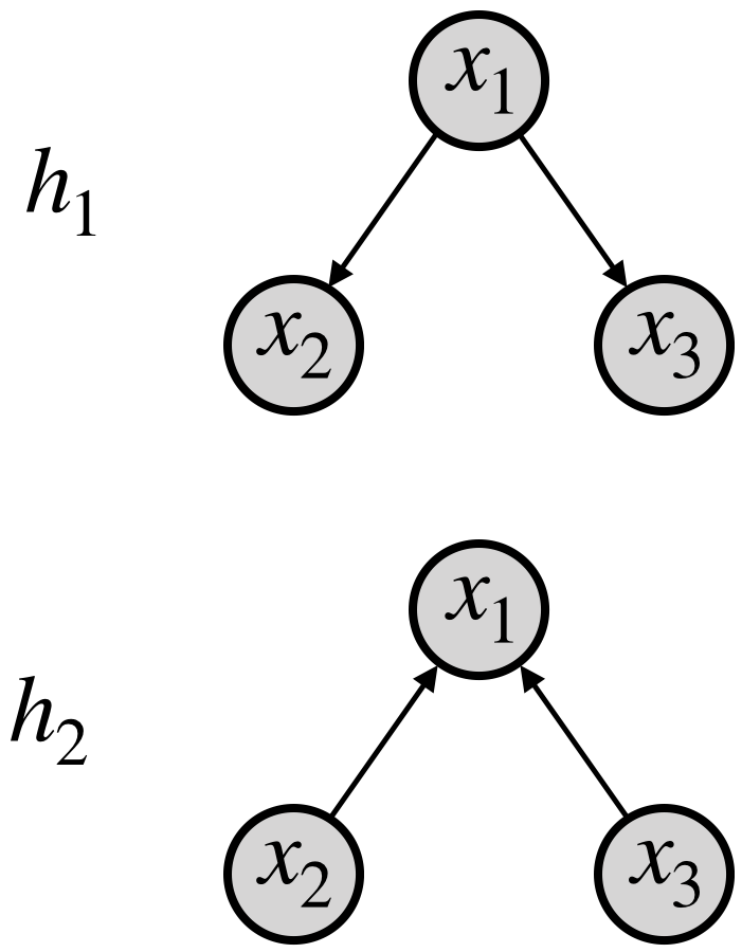
Does the Self-Teaching model capture human's active learning behavior?

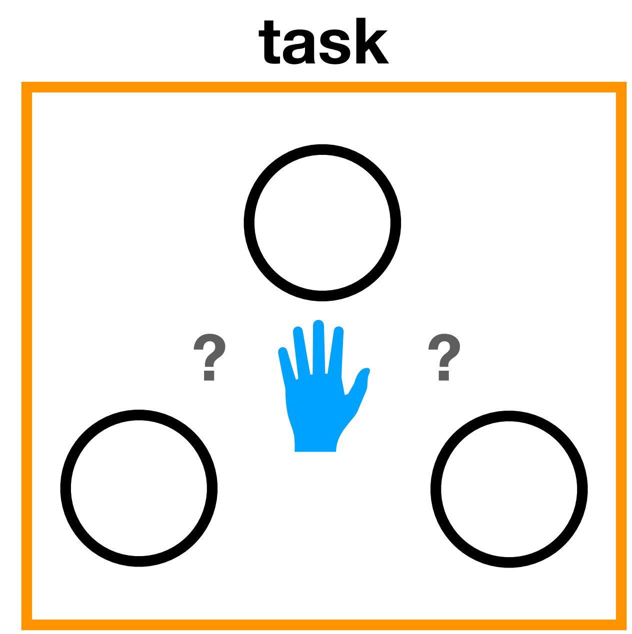
Boundary game



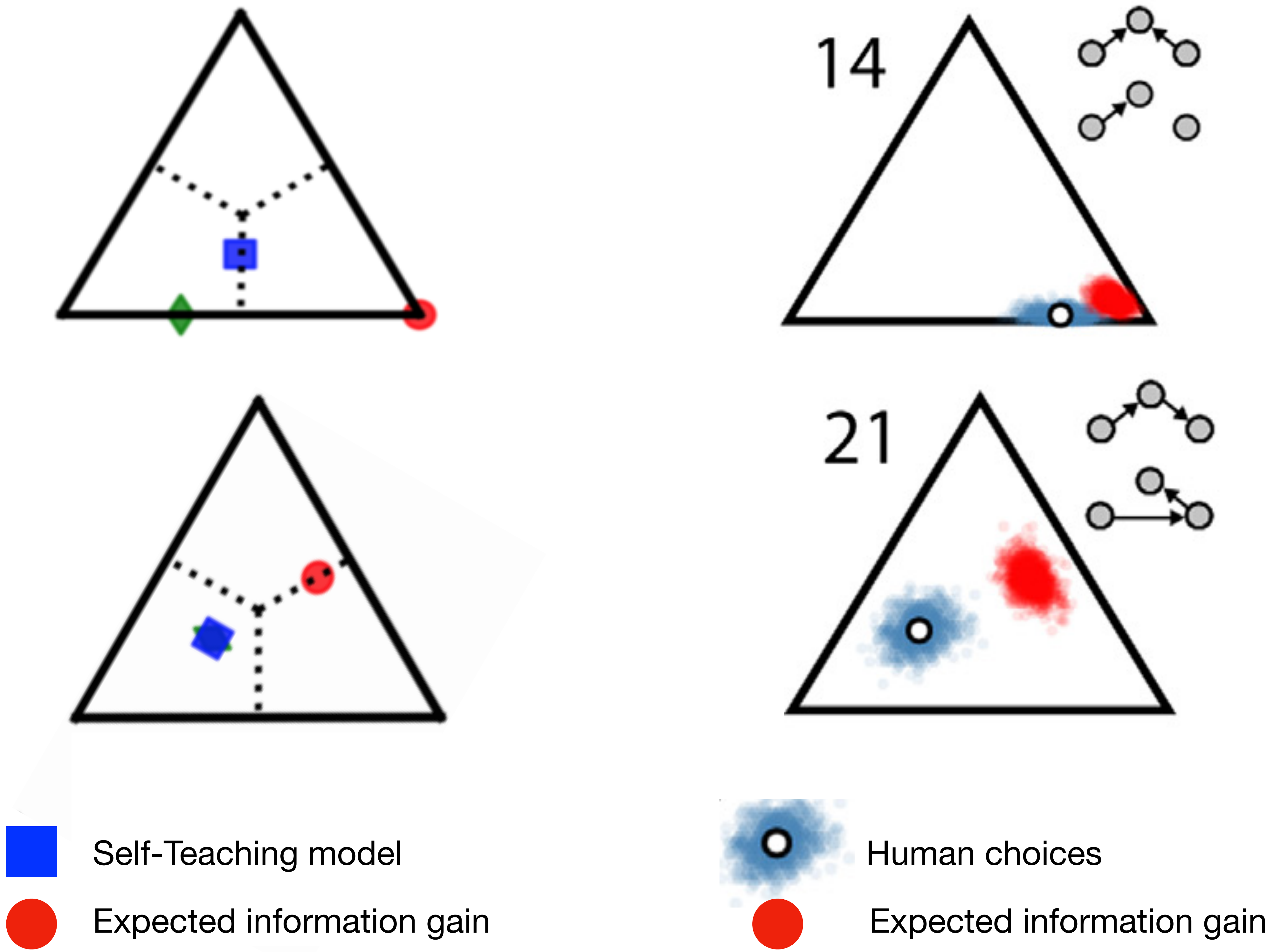


Causal graph learning





Coenen et al. 2015



Coenen, Rehder, & Gureckis. (2015). Strategies to intervene on causal systems are adaptively selected. Cognitive psychology, 79, 102-133.

Conclusions

- We derived a **Self-Teaching model**, a novel form of active learning.
- It depends on only the rules of probability (may have implications for active machine learning).
- It unifies teaching and active learning under a single learning mechanism.
- It matches human's active learning behavior in many cases.

Collaborators

Wai Keen Vong

Yue Yu

Patrick Shafto

Yang, Vong, Yu & Shafto. (2019). A unifying computational framework for teaching and active learning. Topics in Cognitive Science 11(2): 316-337.