

Transfer in Human Concept Learning

Long Ouyang
Noah Goodman

Stanford University

UAI Uncertainty in Natural Intelligence Workshop
August 18, 2012

What is transfer?

Using knowledge from the past
to deal with new stimuli, tasks

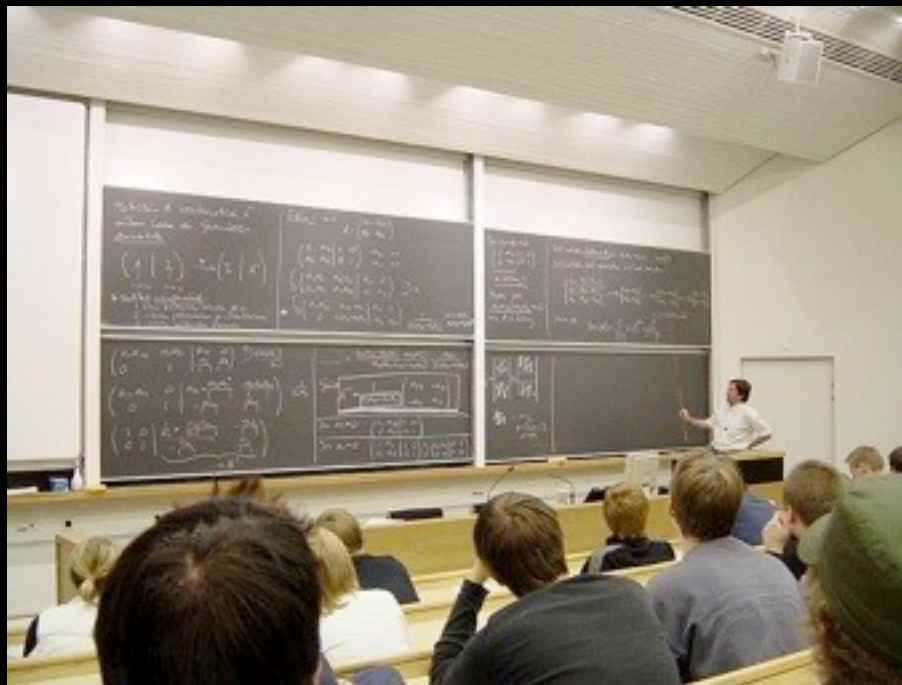
Transfer matters



Transfer matters



Transfer matters



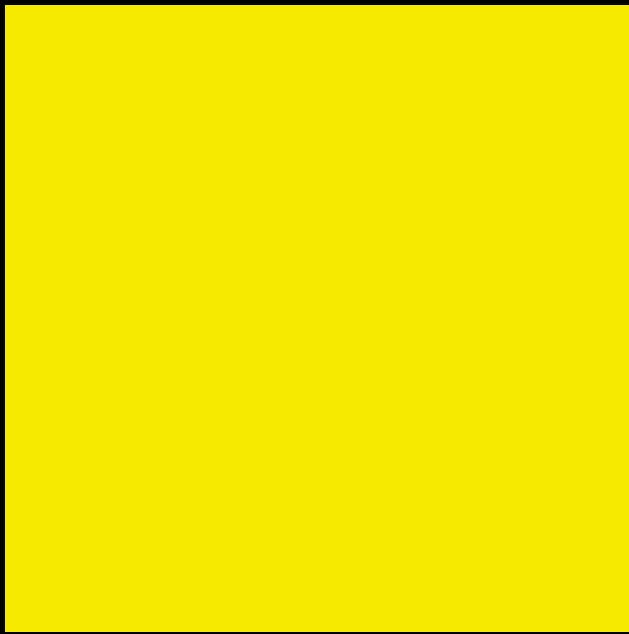
1. Let X be a gamma random variable with parameters λ and α , and let Y be a gamma random variable with parameters λ and β . Further, suppose that $X \perp\!\!\!\perp Y$. Find the joint and marginal distributions of $U = X + Y$ and $V = X/Y$.
2. Mike's bowl of spaghetti contains n strands. He selects two ends at random and joins them together. He does this until there are no ends left. What is the expected number of spaghetti hoops in the bowl?
3. A quality control plan for an assembly line involves sampling n finished items per day and counting X , the number of defective items. If p denotes the probability of observing a defective item, then X has a binomial distribution when the number of items produced by the line is large. However, p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $1/4$.
 - (a) Find the expected value of X for any given day.
 - (b) Find the standard deviation of X .
 - (c) Given $n = 10$ items are sampled, find the expected value and the standard deviation of the number of defectives for that day.
4. Let X be a continuous random variable with density function $f(x) = 2x$, for $0 \leq x \leq 1$. Find the moment generating function, $M_X(t)$, of X , and verify that $E(X) = M'_X(0)$ and that $E(X^2) = M''_X(0)$.

Problem: transfer occurs and
transfer doesn't occur

Transfer occurs

Shepard, 1987

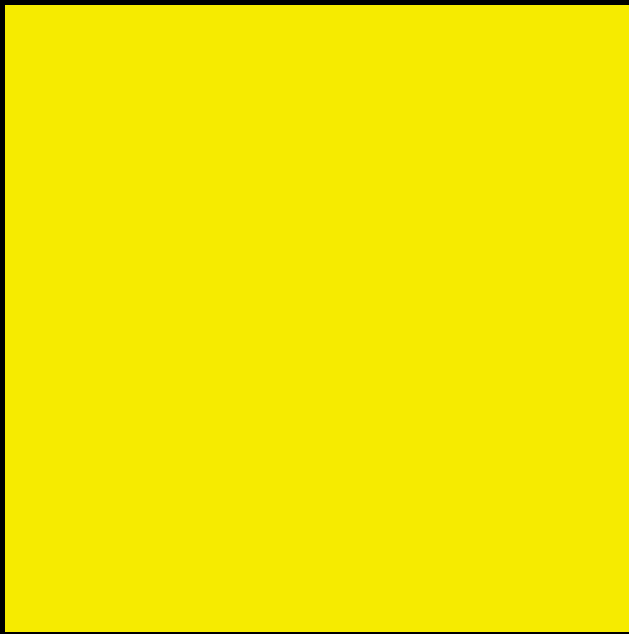
Transfer occurs



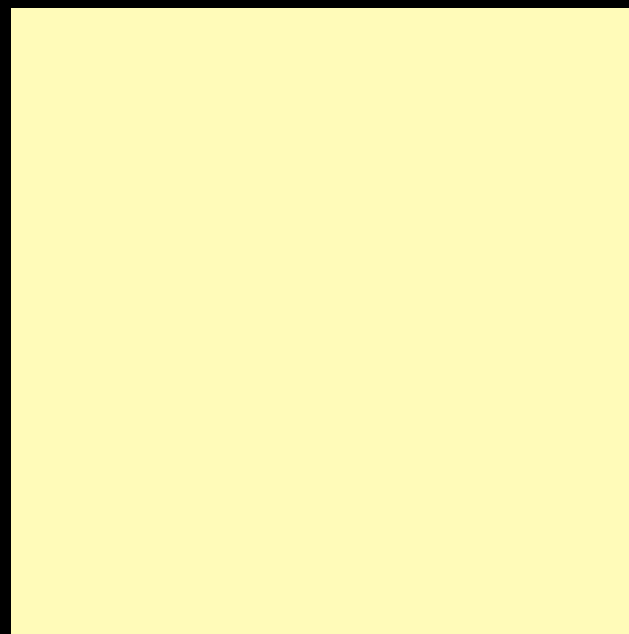
\$

Shepard, 1987

Transfer occurs

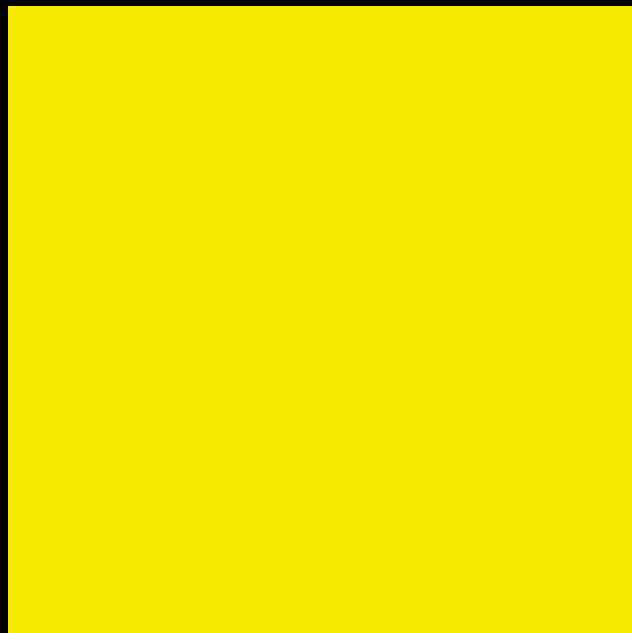


\$

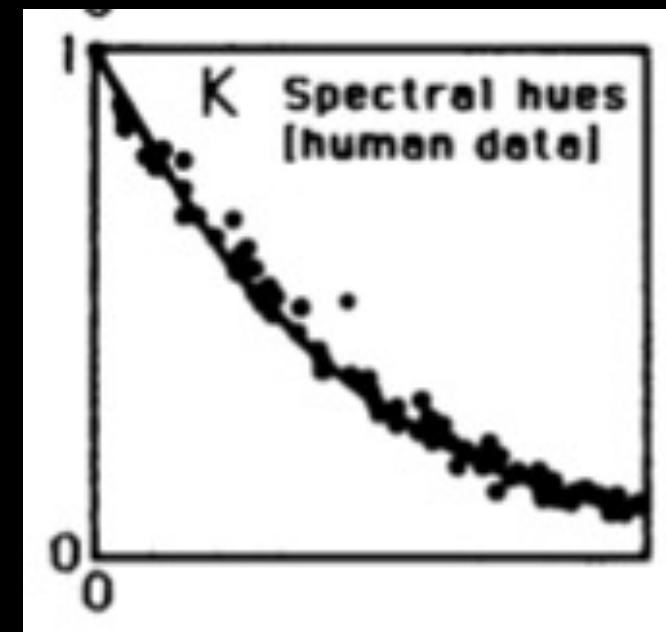
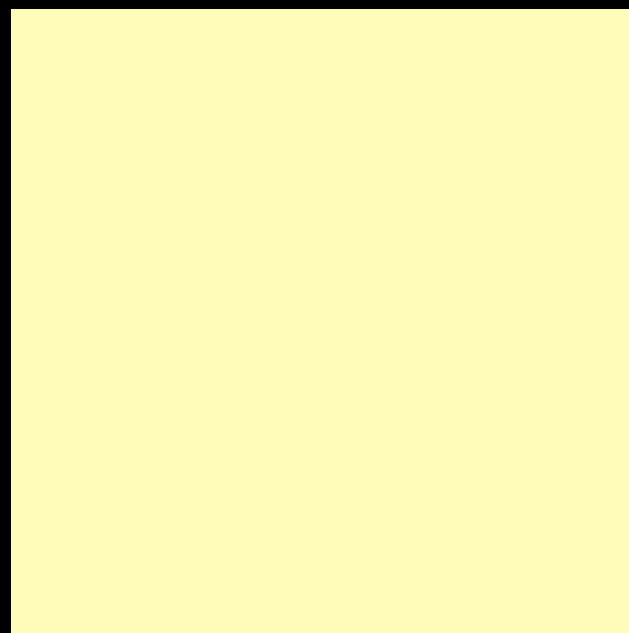


Shepard, 1987

Transfer occurs



\$



Shepard, 1987

Transfer occurs

Scheepers et al., 2011

Transfer occurs

$$80 - (9+1)*5$$

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that _____

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune

...(bells of the church) that ...

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune
...(bells of the church) that ...

80 - 9+1*5

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune
...(bells of the church) that ...

80 - 9+1*5

I saw the bells of the church that _____

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune
...(bells of the church) that ...

80 - 9+1*5

I saw the bells of the church that is on 5th street

Scheepers et al., 2011

Transfer occurs

80 - (9+1)*5

I saw the bells of the church that were out of tune
...(bells of the church) that ...

80 - 9+1*5

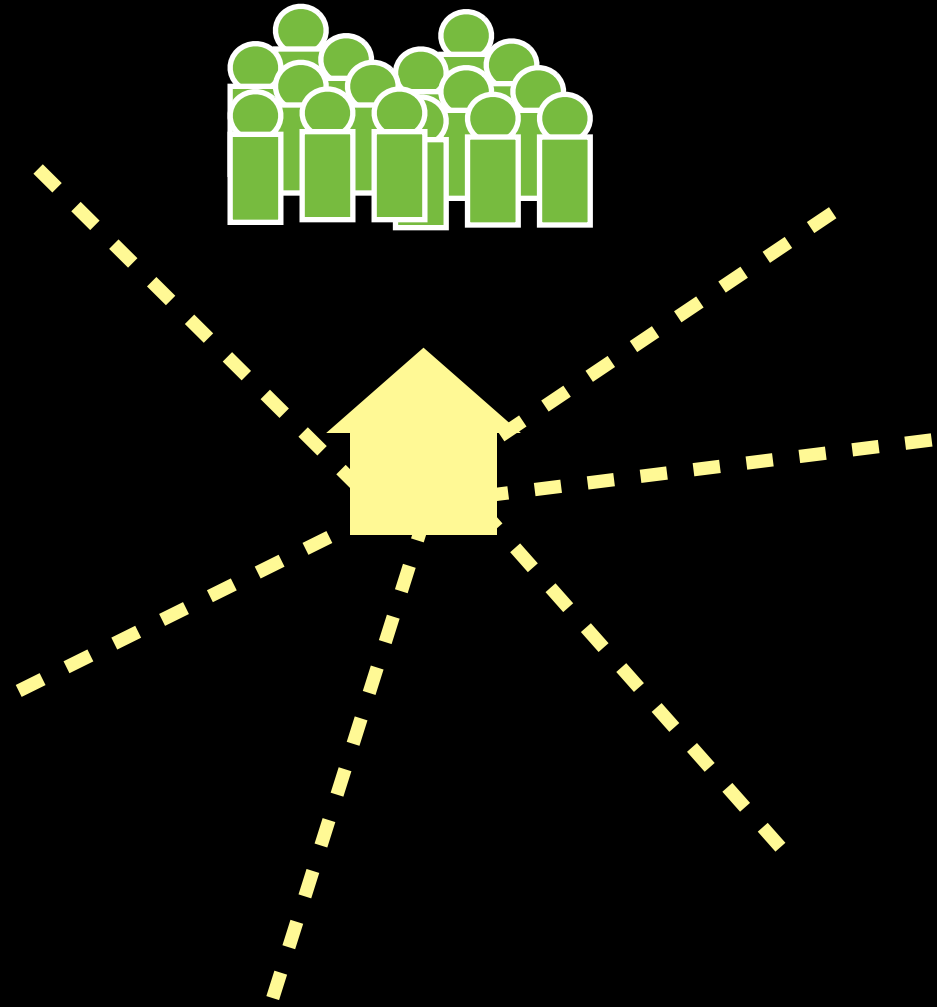
I saw the bells of the church that is on 5th street
...bells of the (church that ...)

Scheepers et al., 2011

Transfer doesn't occur

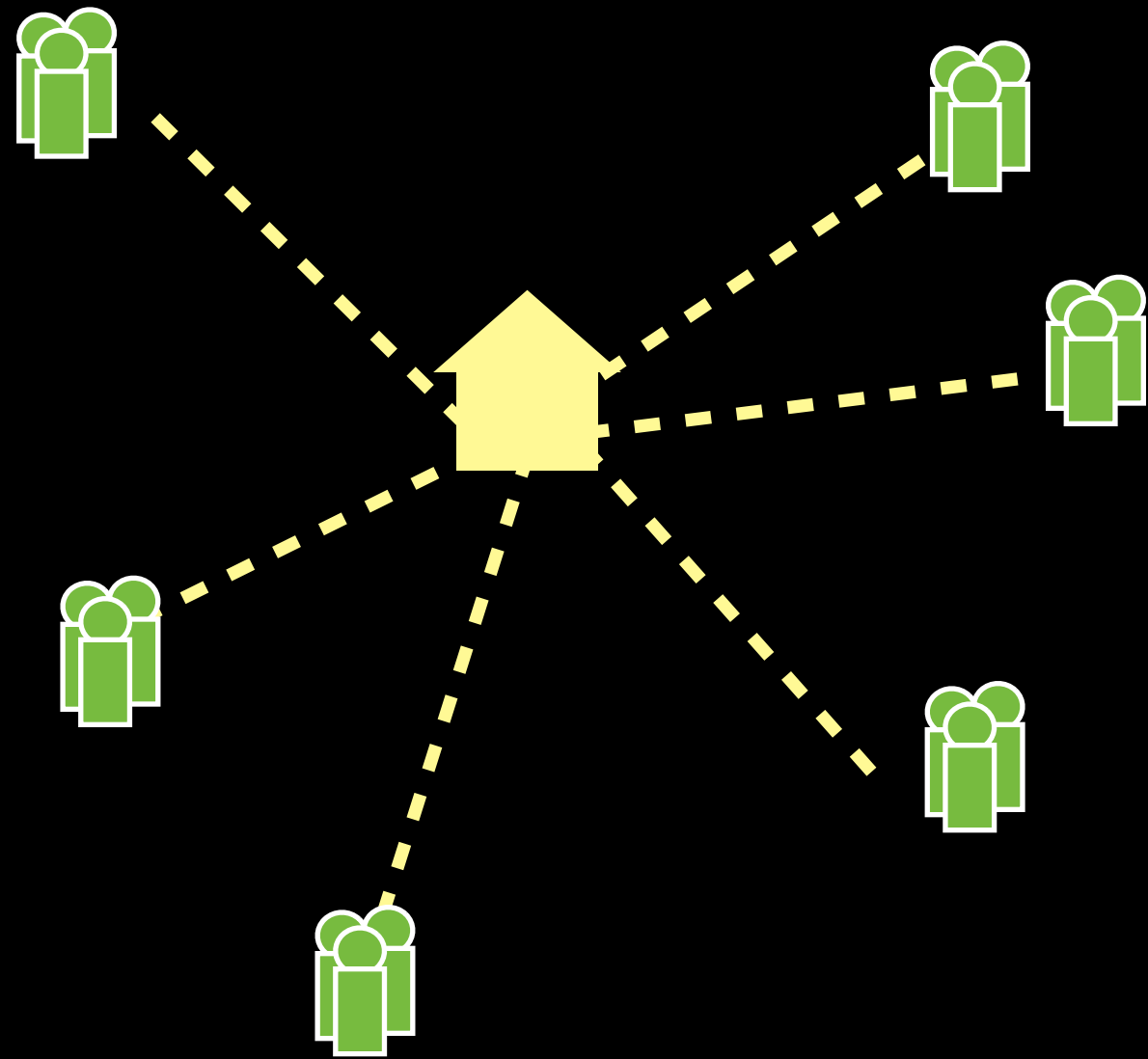
Gick & Holyoak (1980)

Transfer doesn't occur



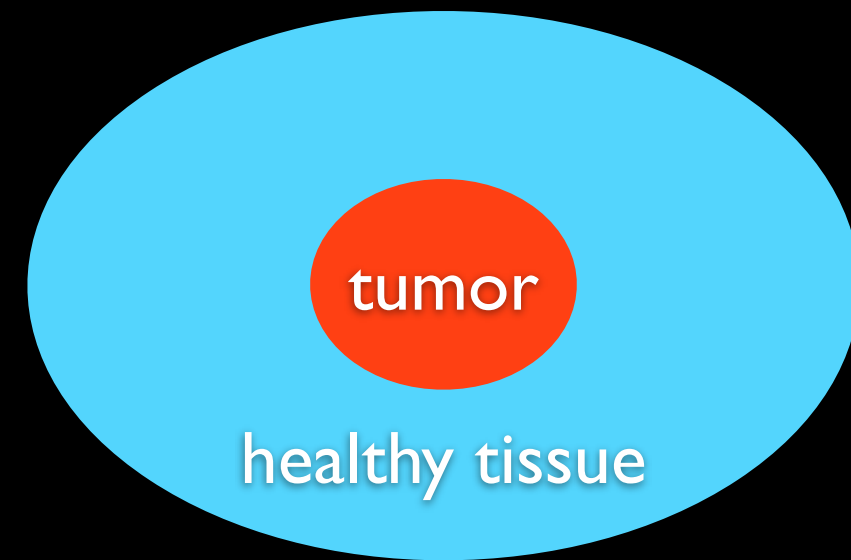
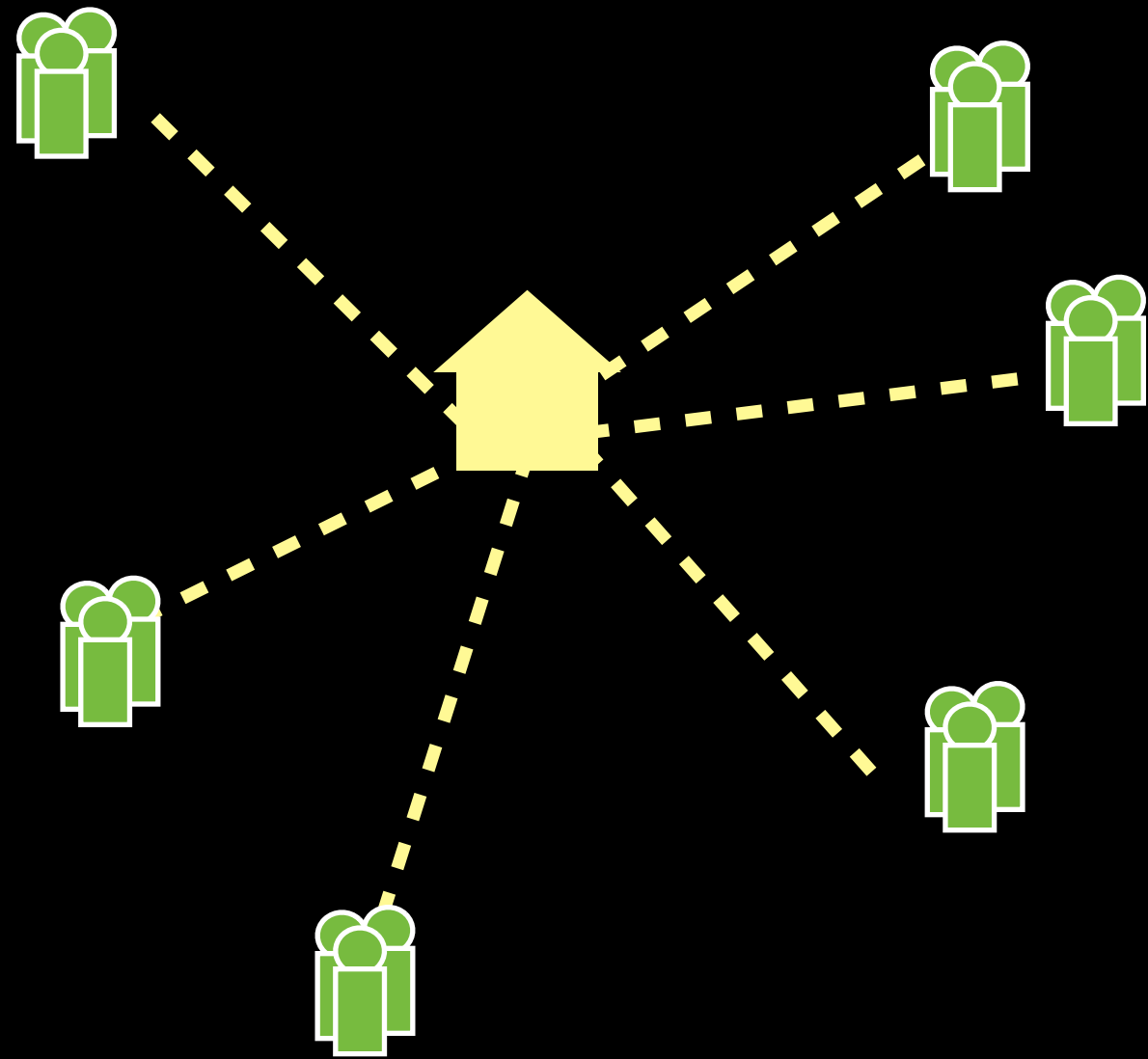
Gick & Holyoak (1980)

Transfer doesn't occur



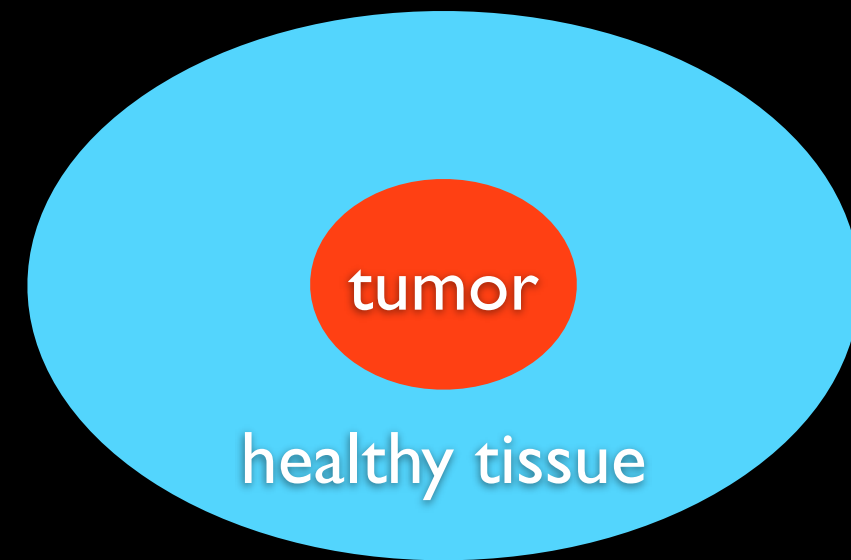
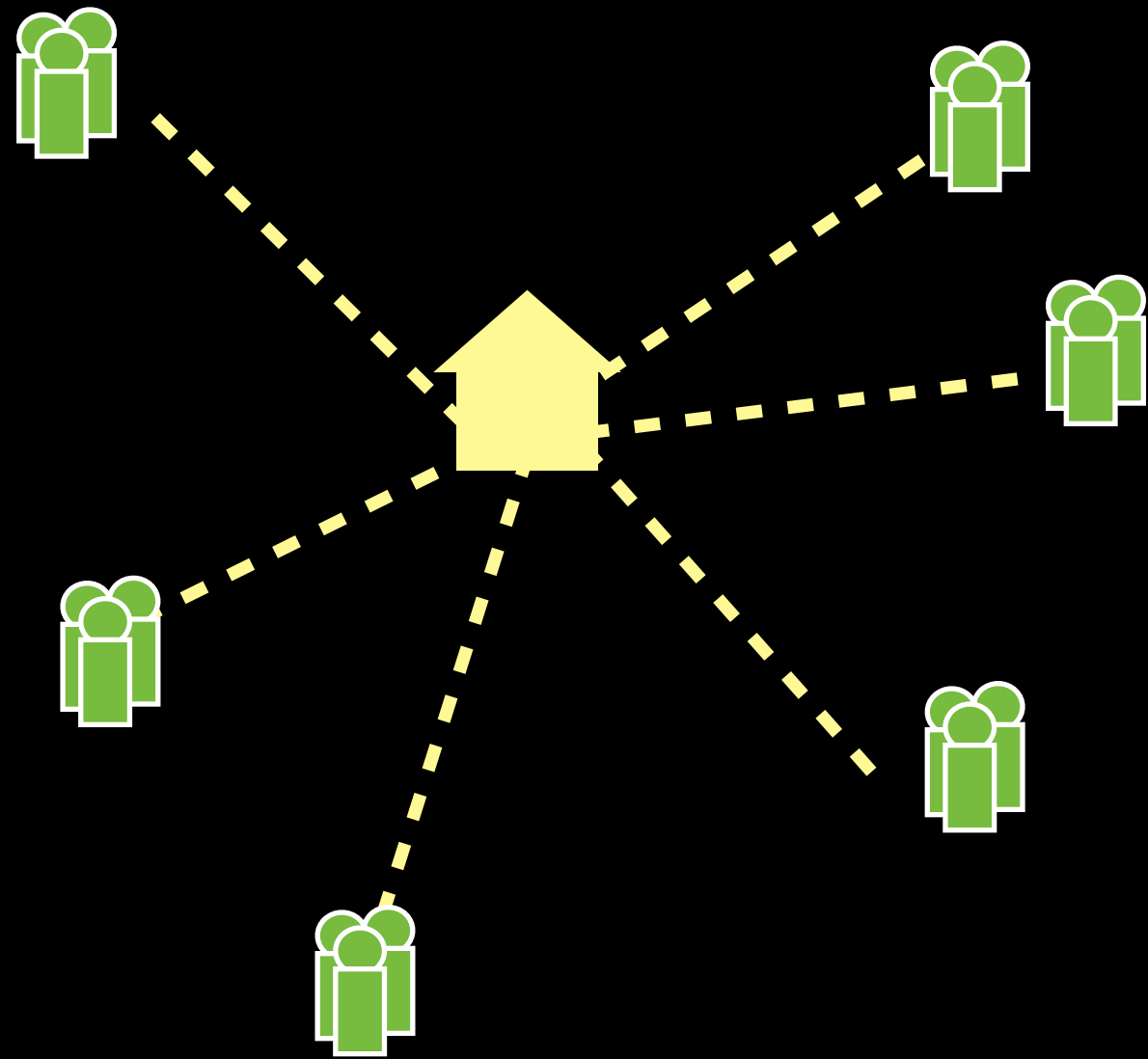
Gick & Holyoak (1980)

Transfer doesn't occur



Gick & Holyoak (1980)

Transfer doesn't occur



without hint,
only 30% use
dispersion
strategy

Gick & Holyoak (1980)

Transfer doesn't occur

Cheng et al. (1986)

Transfer doesn't occur

Formal logic training doesn't seem to help much

Cheng et al. (1986)

Transfer doesn't occur

Formal logic training doesn't seem to help much

“40 h of lectures ... in propositional logic, including *modus ponens*, *modus tollens*, affirming the consequent, and denying the antecedent, and the distinction between the conditional and the biconditional.”

Cheng et al. (1986)

Transfer doesn't occur

Formal logic training doesn't seem to help much

“40 h of lectures ... in propositional logic, including *modus ponens*, *modus tollens*, affirming the consequent, and denying the antecedent, and the distinction between the conditional and the biconditional.”

“...The mean improvement was a bare $3 \pm 7\%$ ”

Cheng et al. (1986)

Question

Broad

When does transfer occur?

Specific

How does learning one (Boolean) concept change learning for future concepts?

Transfer in the sense of *preparation for future learning* (Bransford & Schwartz, 2001)

Intuition: bootstrapping

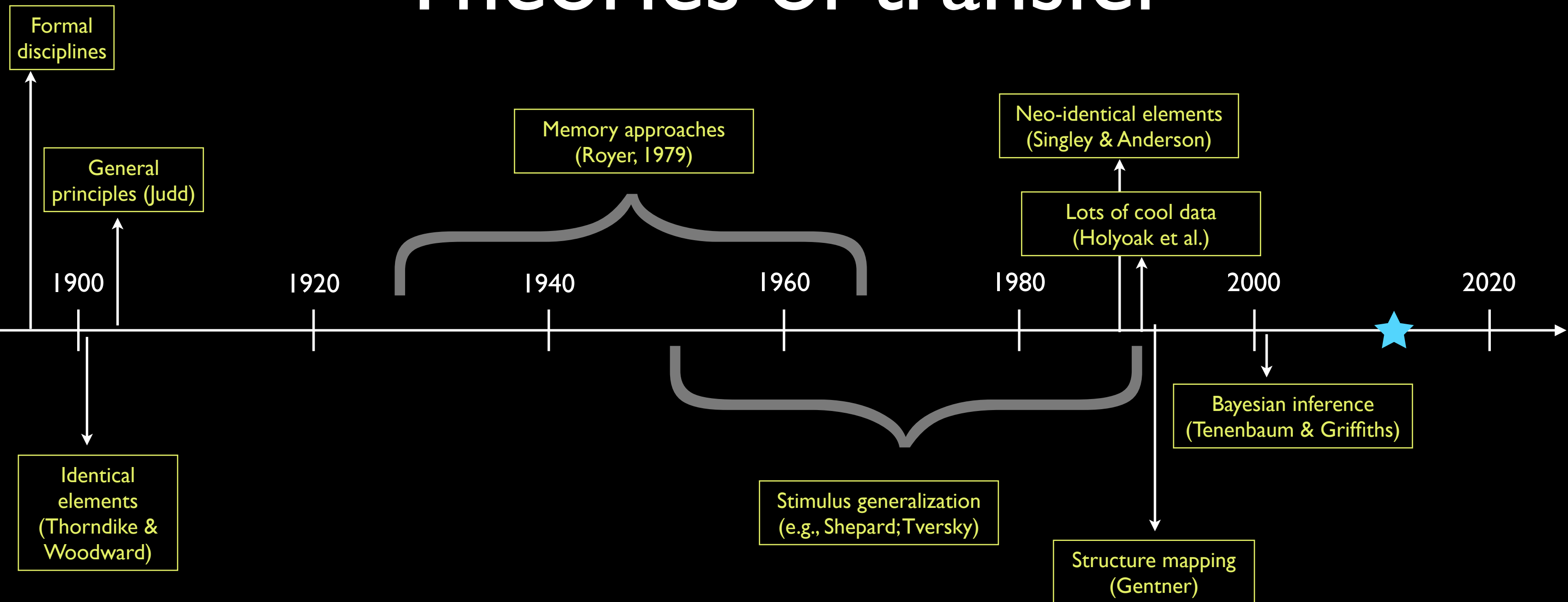
e.g., Easier to learn about transistors if you know about electricity

Intuition: bootstrapping

e.g., Easier to learn about transistors if you know about electricity



Theories of transfer



Bayesian state of the art

Jointly learn concepts (cf. “molecules”) and features (cf. “atoms”)

- Kemp, Goodman, & Tenenbaum, 2010
- Canini & Griffiths, 2010
- Lake, Salakhutdinov, Gross, & Tenenbaum, 2011
- Austerweil & Griffiths, 2011

Bayesian state of the art

Jointly learn concepts (cf. “molecules”) and features (cf. “atoms”)

- Kemp, Goodman, & Tenenbaum, 2010
- Canini & Griffiths, 2010
- Lake, Salakhutdinov, Gross, & Tenenbaum, 2011
- Austerweil & Griffiths, 2011

Challenge: extend this to cases of richer compositionality

→ Blur the distinction between concepts and features



Experiment

Experiment



Stanford Computation and Cognition Lab

This experiment has 2 parts. In each part, there are 8 objects that we have divided into two groups, P and Q. Your task is to learn which objects belong to each group. We'll present the objects to you one at a time and you'll respond by pressing P or Q on your keyboard. The entire experiment takes about 10 minutes – please do the HIT without any interruptions. If you need to go to the bathroom or get a snack, do so right now, before the experiment starts.

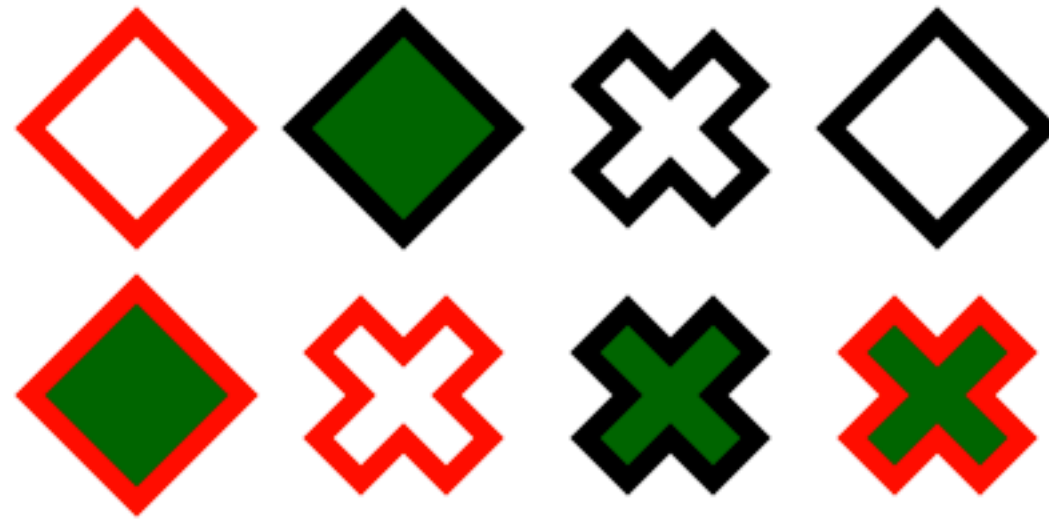
Begin

Legal information: By answering the following questions, you are participating in a study being performed by cognitive scientists in the Stanford Department of Psychology. If you have questions about this research, please contact Long Ouyang at louyang@stanford.edu or Noah Goodman, at ngoodman@stanford.edu. You must be at least 18 years old to participate. Your participation in this research is voluntary. You may decline to answer any or all of the following questions. You may decline further participation, at any time, without adverse consequences. Your anonymity is assured; the researchers who have requested your participation will not receive any personal information about you.

Experiment

Part 1 / 2

These are the 8 objects in this part:



There are 64 trials in this part.

Begin

Experiment

Part 1 / 2

Is this object a Q or a P?
Press Q/P to respond.

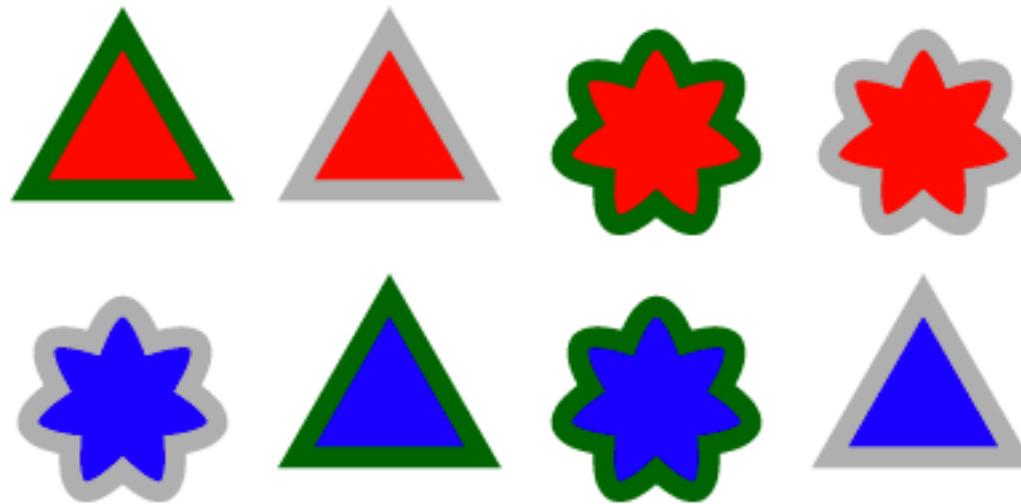


Right!

Experiment

Part 2 / 2

These are the 8 objects in this part:



There are 64 trials in this part.

Begin

Experiment

Part 2 / 2

Is this object a Q or a P?
Press Q/P to respond.



Wrong. The answer was P.

Boolean concept learning

Boolean concept learning

Three binary dimensions, a , b , c (8 objects total)

Boolean concept learning

Three binary dimensions, a , b , c (8 objects total)

→ Randomly instantiated to {shape, inner color, outer color}

Boolean concept learning

Three binary dimensions, a , b , c (8 objects total)

→ Randomly instantiated to {shape, inner color, outer color}

P/Q classification based on Boolean function, h , of (a, b, c)

→ $||$ different combinations of $(h^{(1)}, h^{(2)})$

Boolean concept learning

Three binary dimensions, a , b , c (8 objects total)

→ Randomly instantiated to {shape, inner color, outer color}

P/Q classification based on Boolean function, h , of (a, b, c)

→ 11 different combinations of $(h^{(1)}, h^{(2)})$

Fixed, blocked pseudorandom trial order

Boolean concept learning

Three binary dimensions, a , b , c (8 objects total)

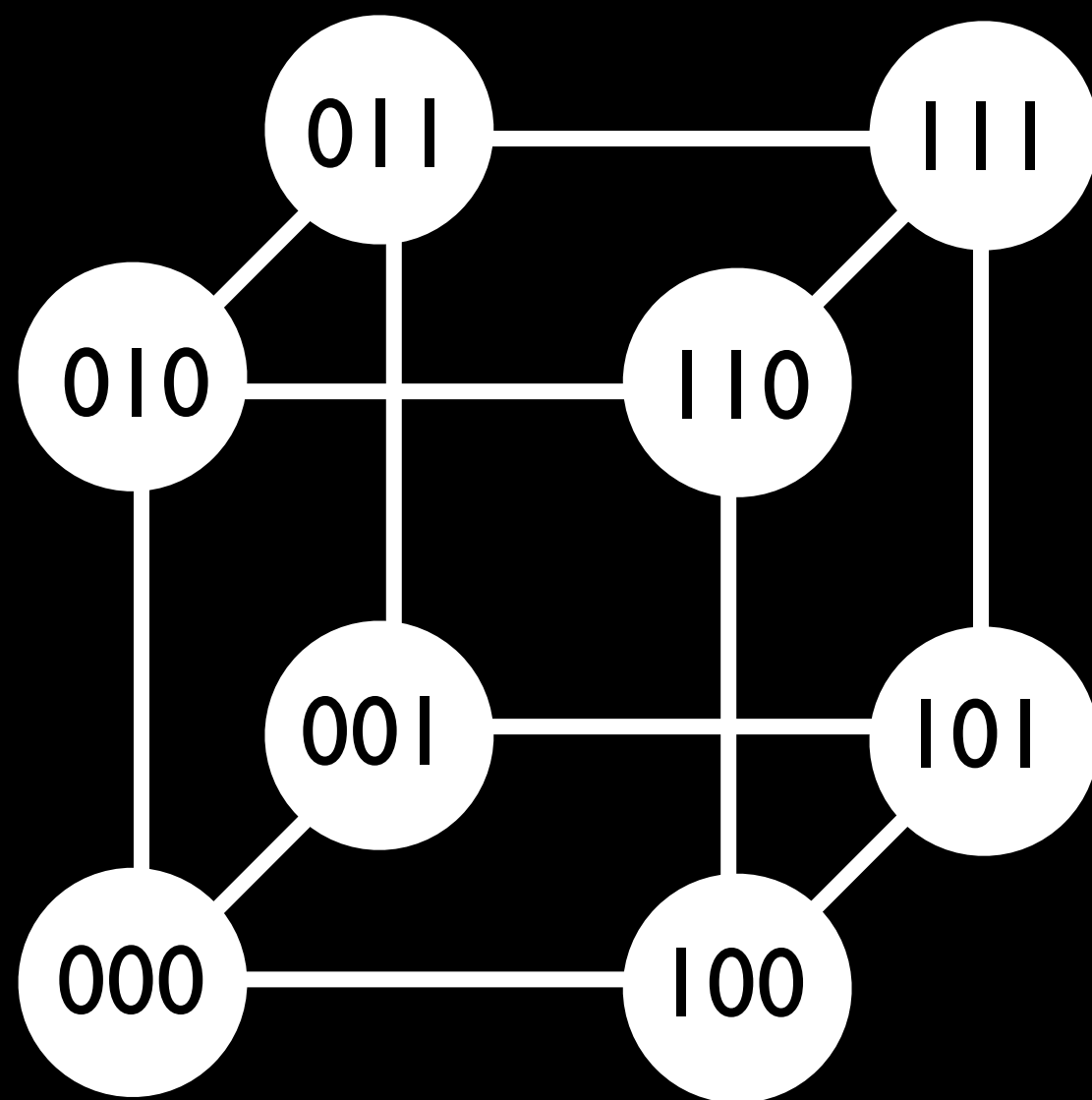
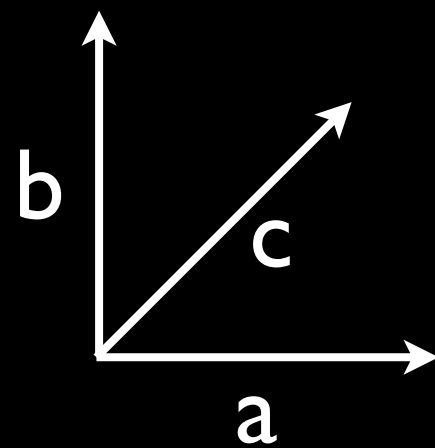
→ Randomly instantiated to {shape, inner color, outer color}

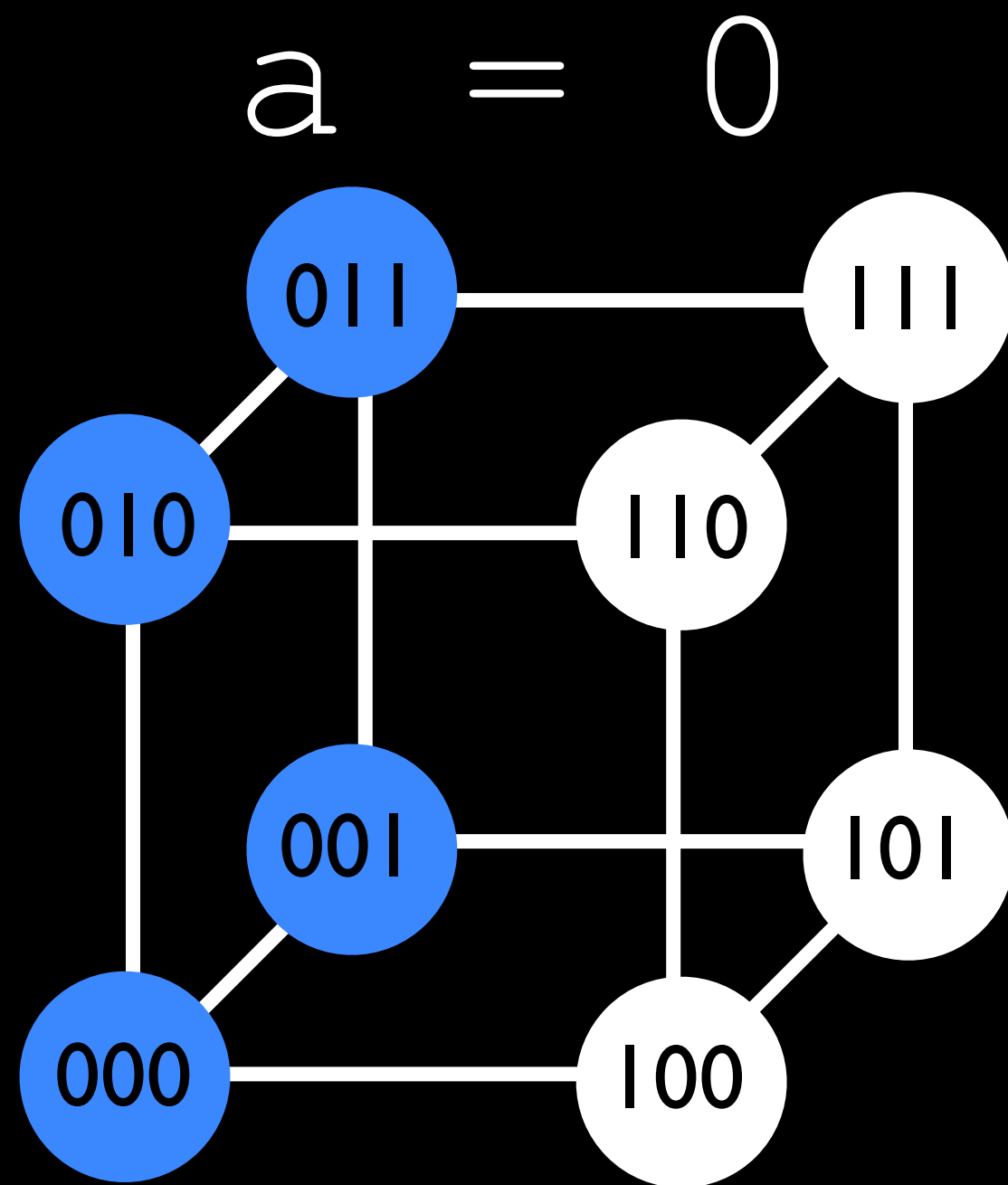
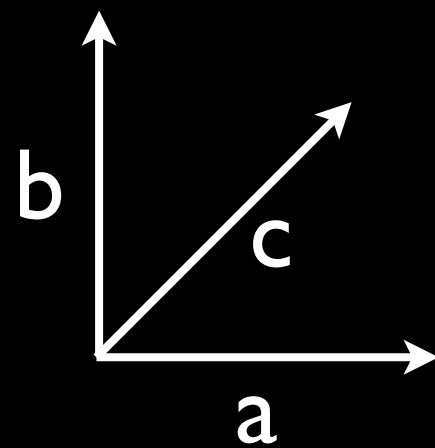
P/Q classification based on Boolean function, h , of (a, b, c)

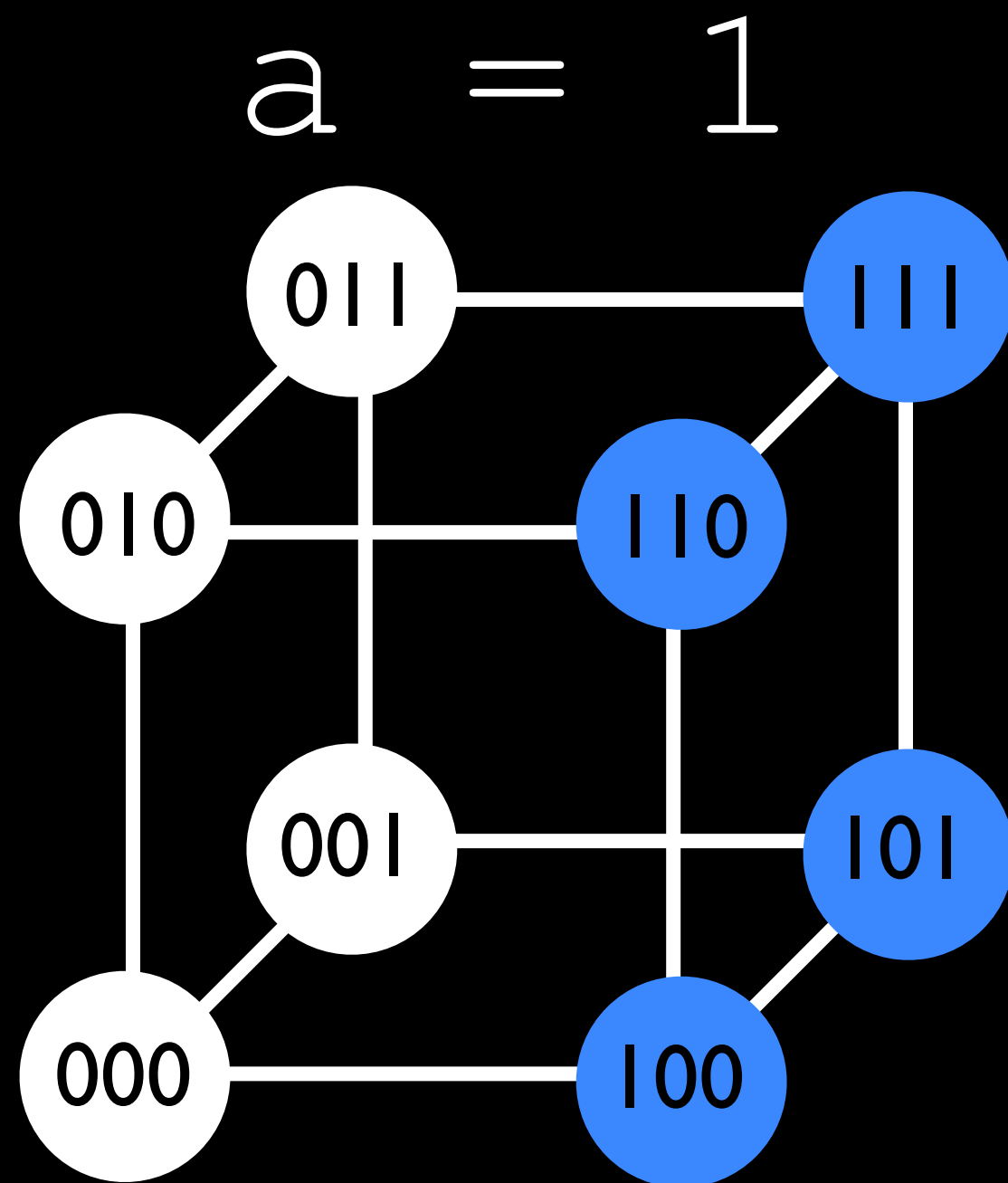
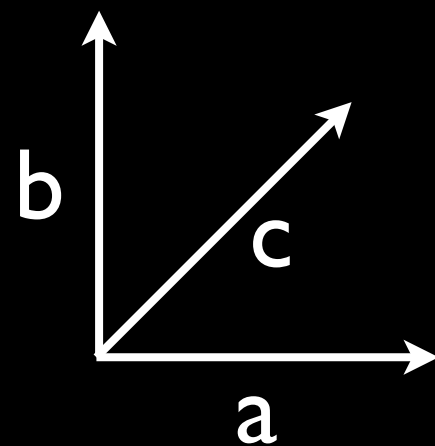
→ 11 different combinations of $(h^{(1)}, h^{(2)})$

Fixed, blocked pseudorandom trial order

Set of objects is different across domains







Notation

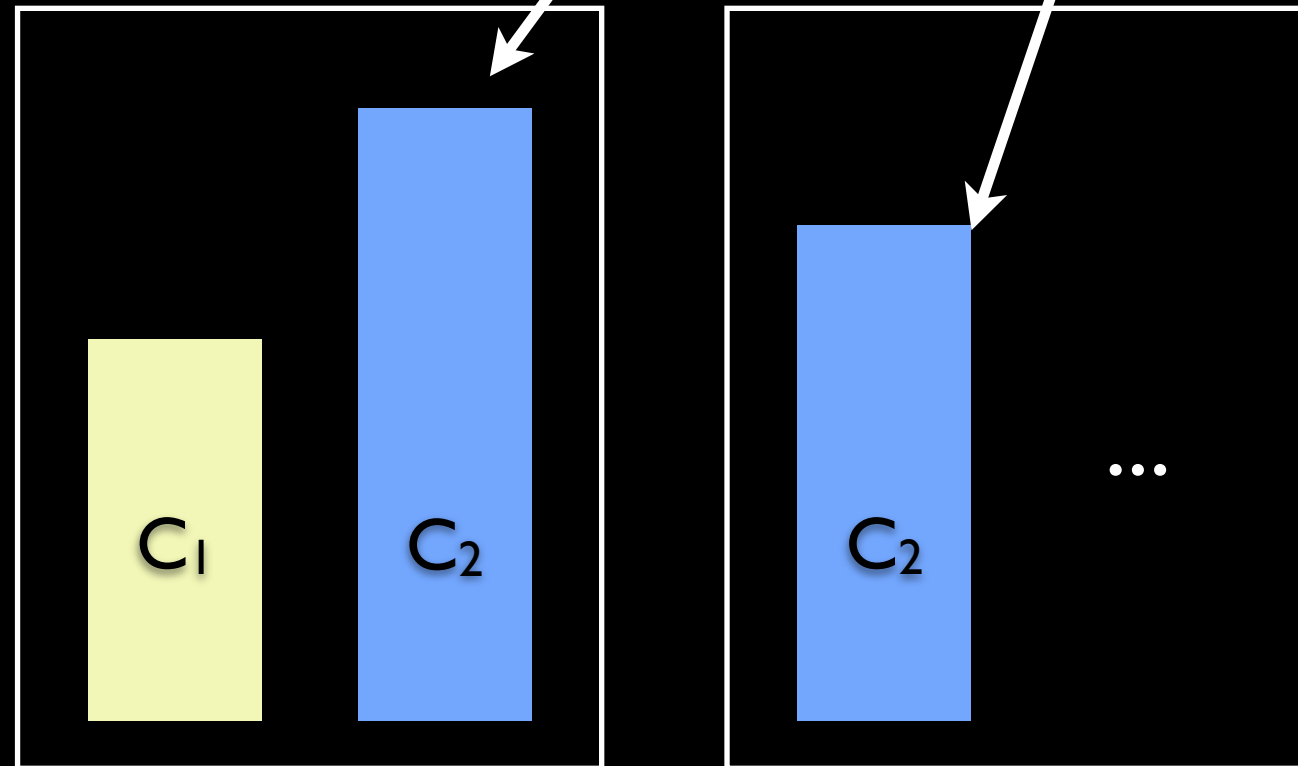
Logical	Algebraic
a	a
$\neg a$	a'
$a \wedge b$	ab
$a \vee b$	$a+b$
$a \oplus b$	$a \wedge b$

Transfer operationalized

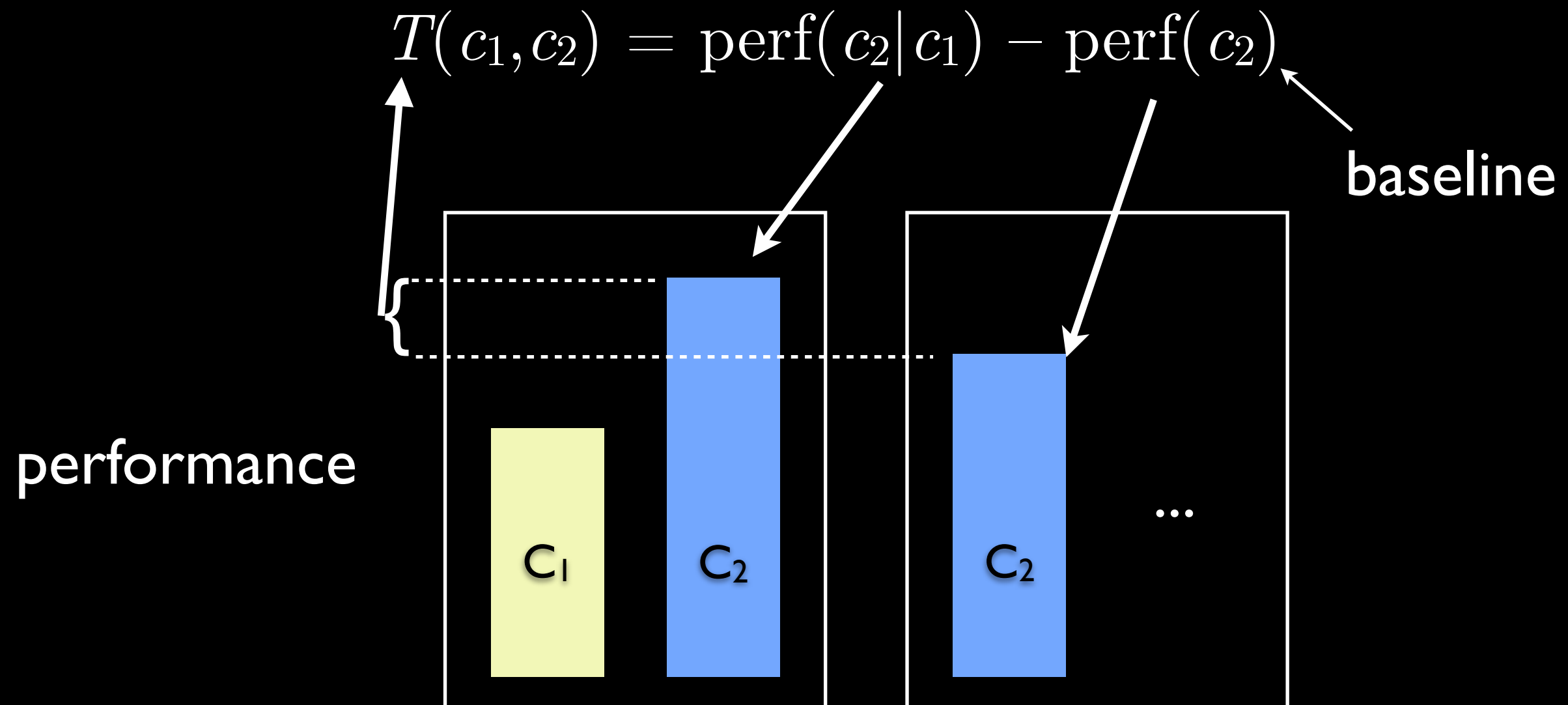
$$T(c_1, c_2) = \text{perf}(c_2|c_1) - \text{perf}(c_2)$$

baseline

performance

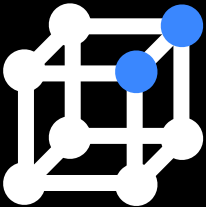
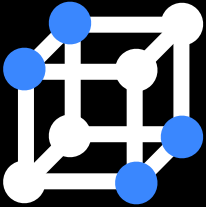
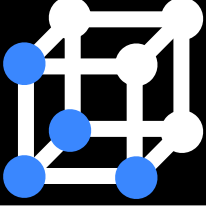
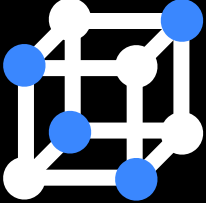


Transfer operationalized

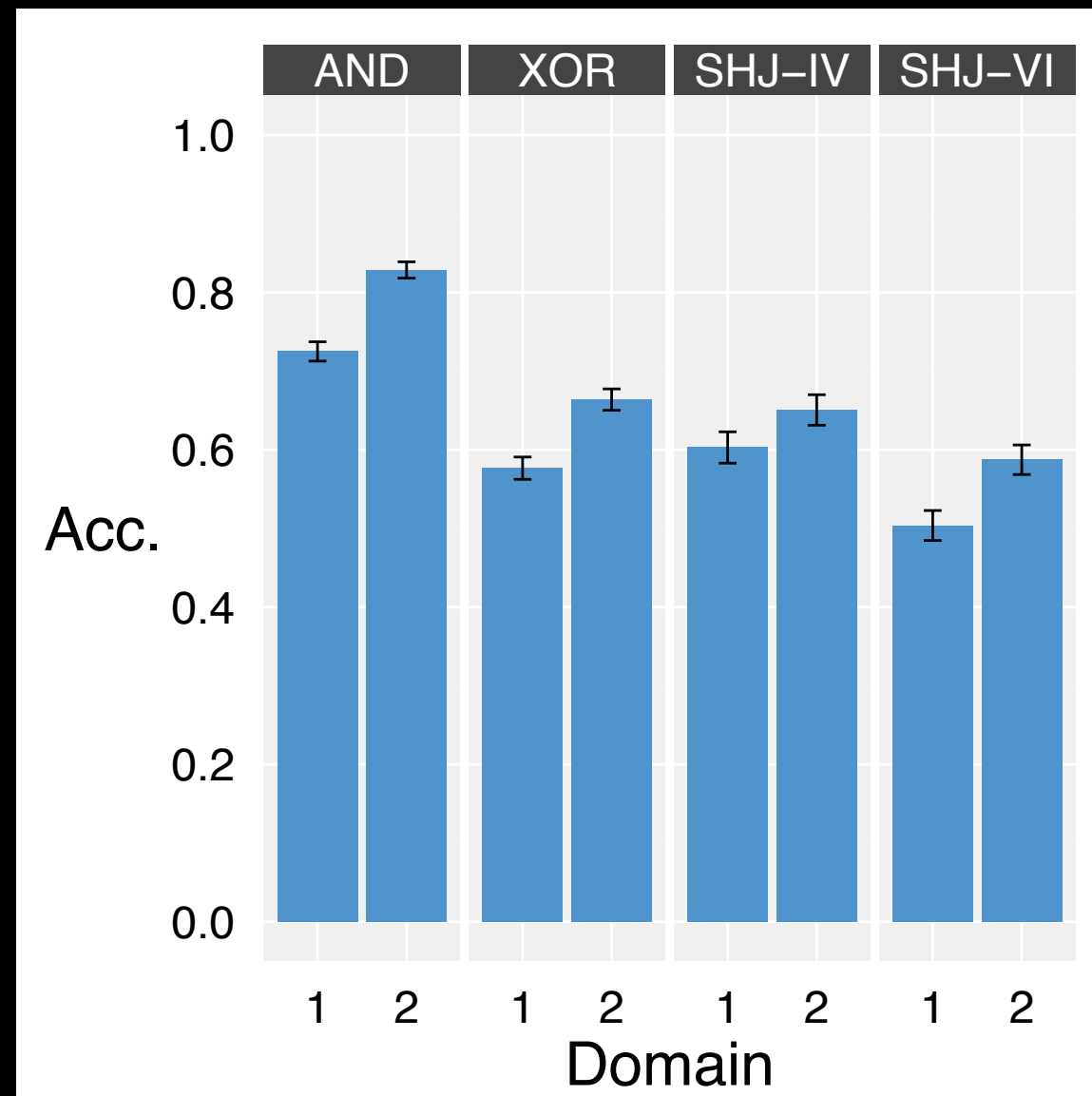


Results I: + self transfer

Positive self transfer for these concepts:

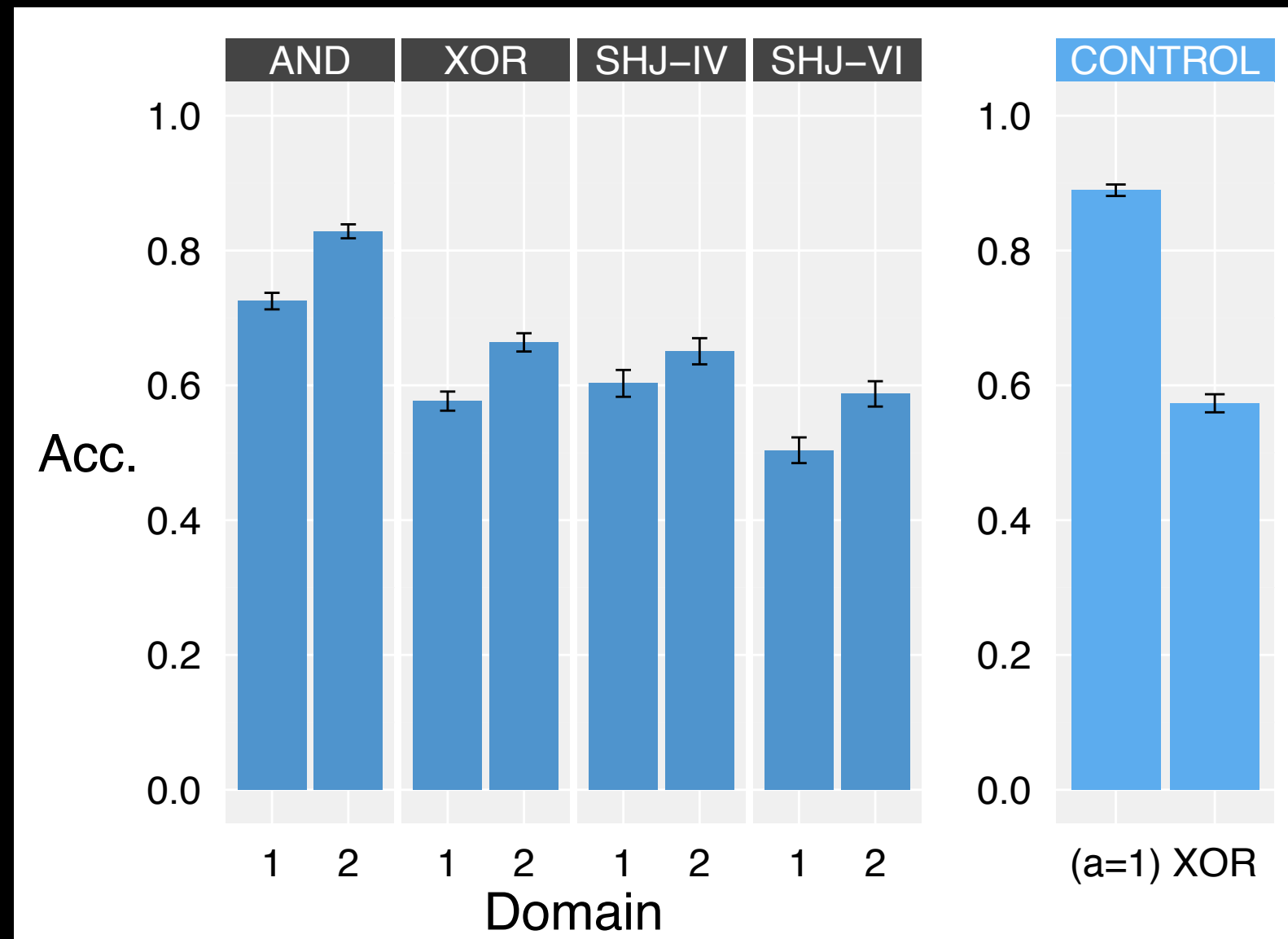
Shorthand	Formula	Cube
AND	ab	
XOR	$a \oplus b$	
SHJ-IV	$a' (bc)' + ab' c'$	
SHJ-VI	$a \oplus (b \oplus c)$	

Results I: + self transfer



cf. Bourne, 1970

Results I: + self transfer



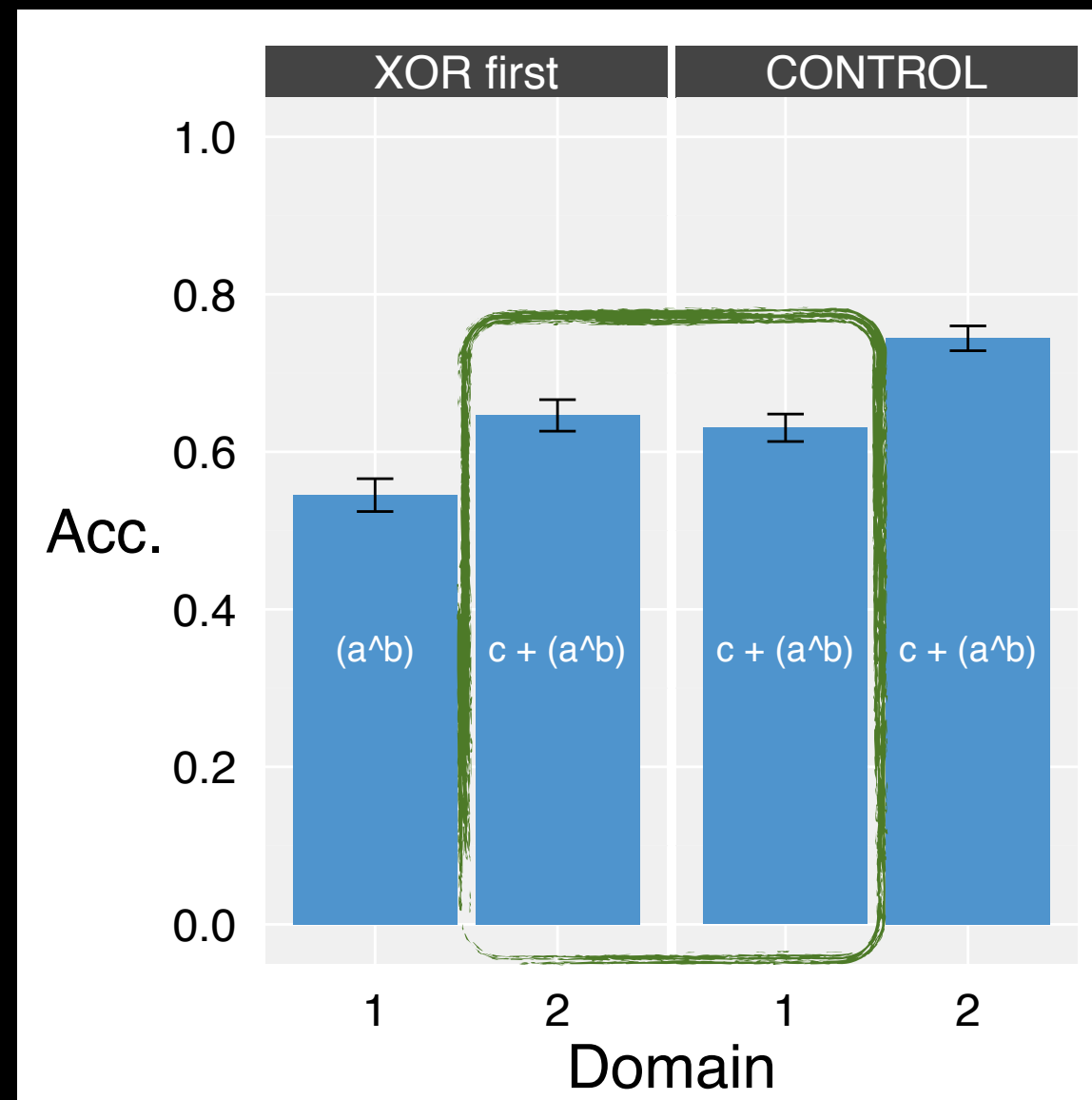
cf. Bourne, 1970

Results I: garden pathing

Does learning a^b transfer to $c + (a^b)$?

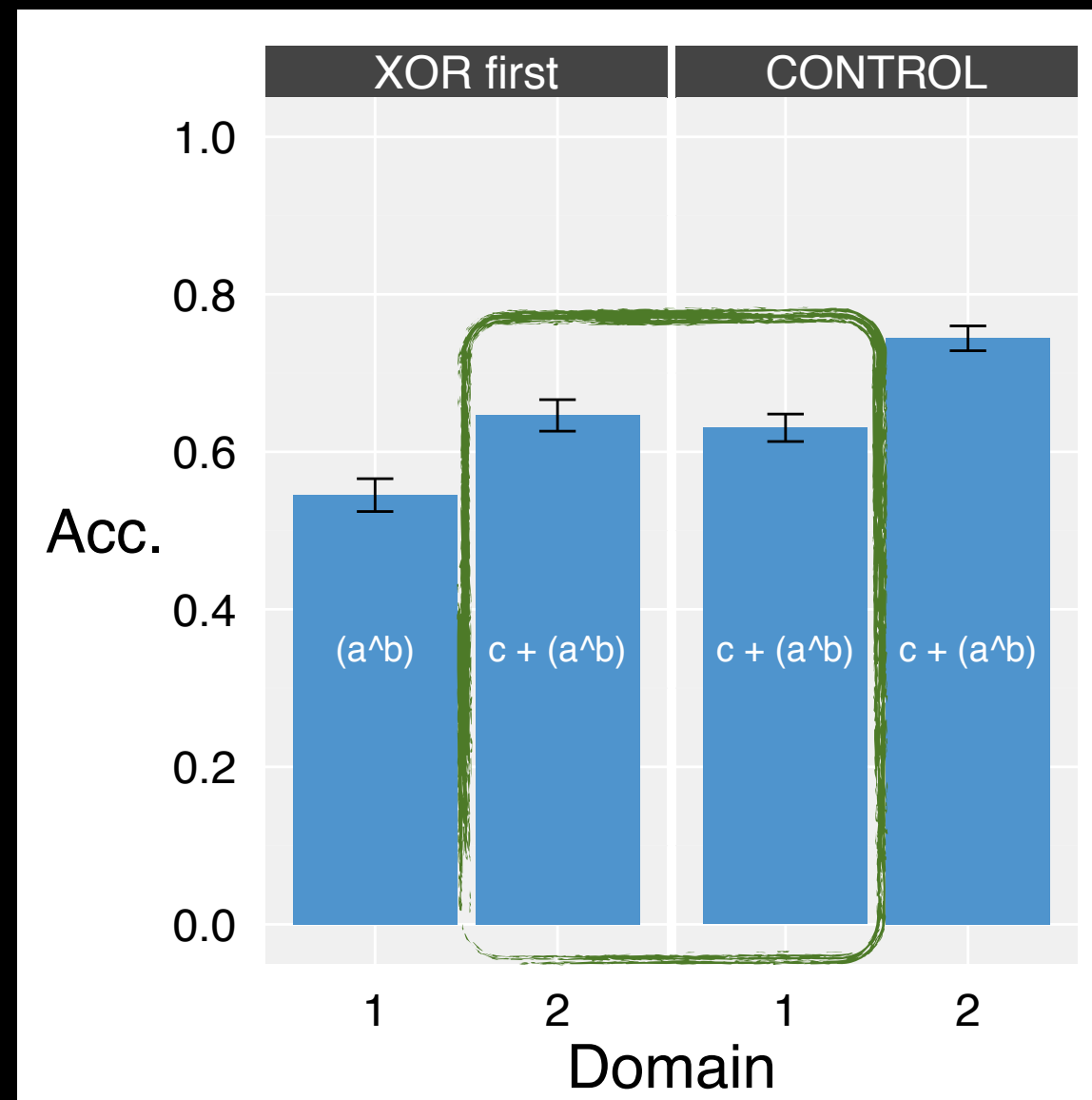
Results I: garden pathing

Does learning $a \wedge b$ transfer to $c + (a \wedge b)$?



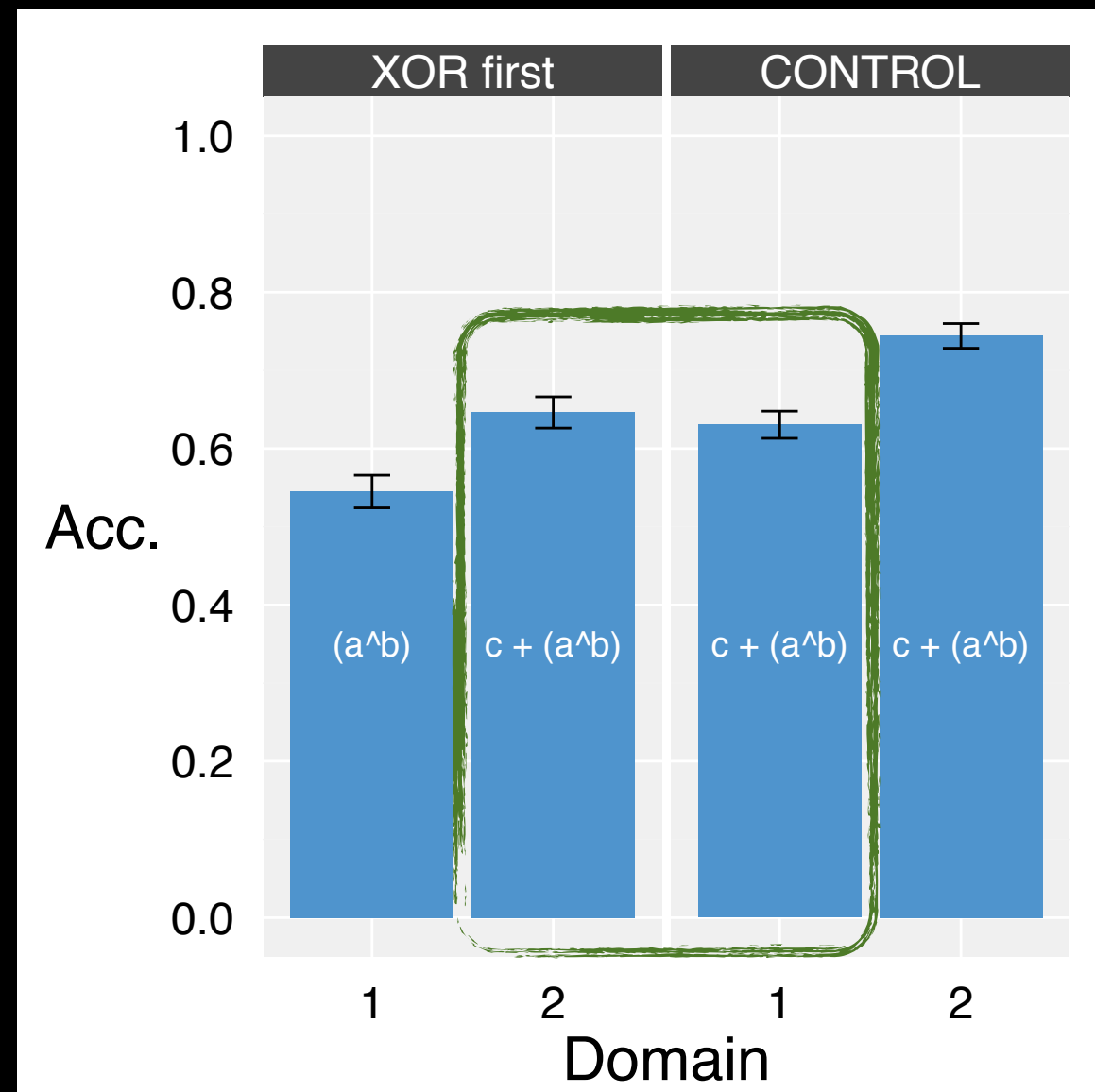
Results I: garden pathing

Does learning $a \wedge b$ transfer to $c + (a \wedge b)$?



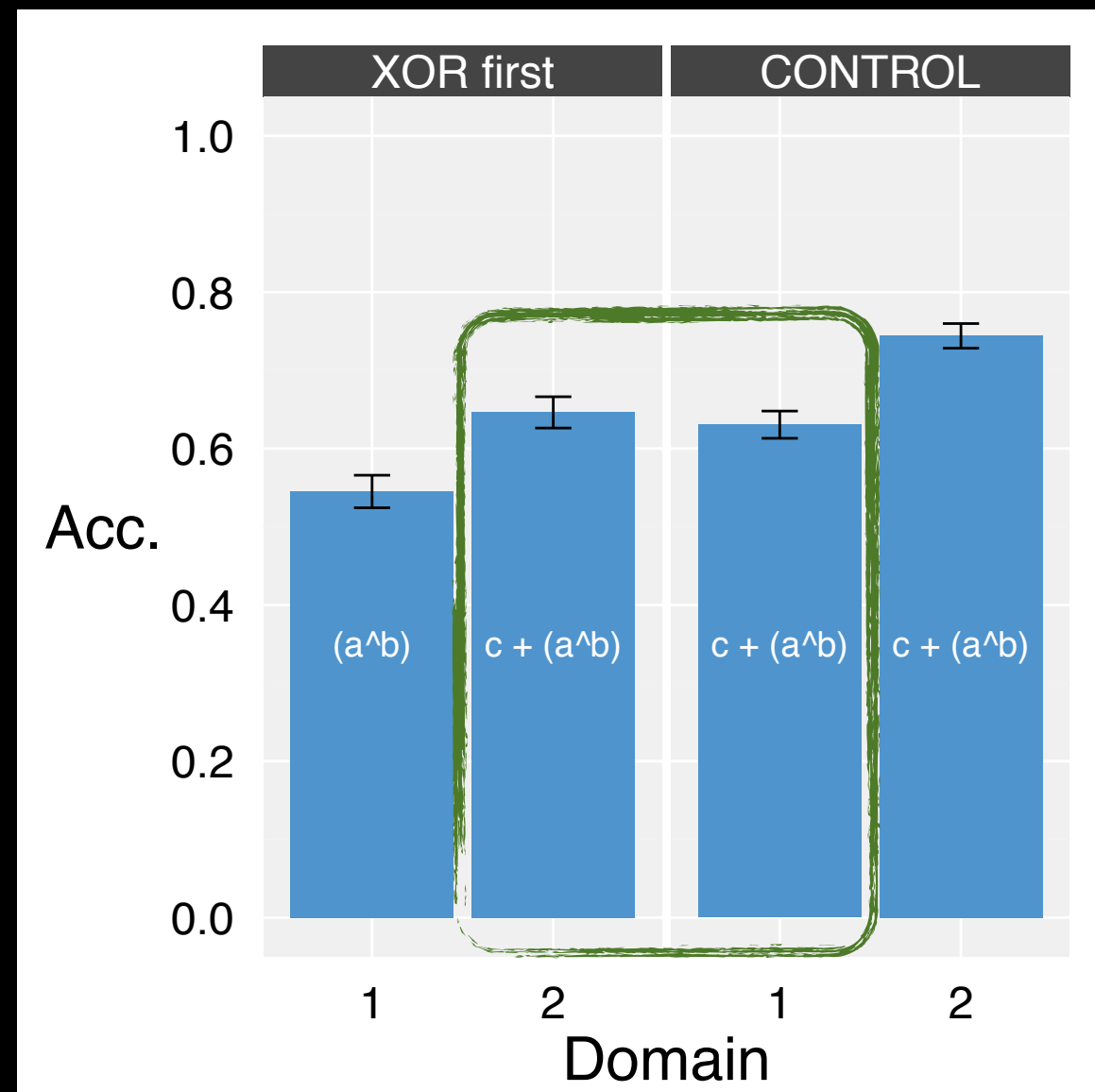
Results I: garden pathing

Does learning $a \oplus b$ transfer to $c + (a \oplus b)$?



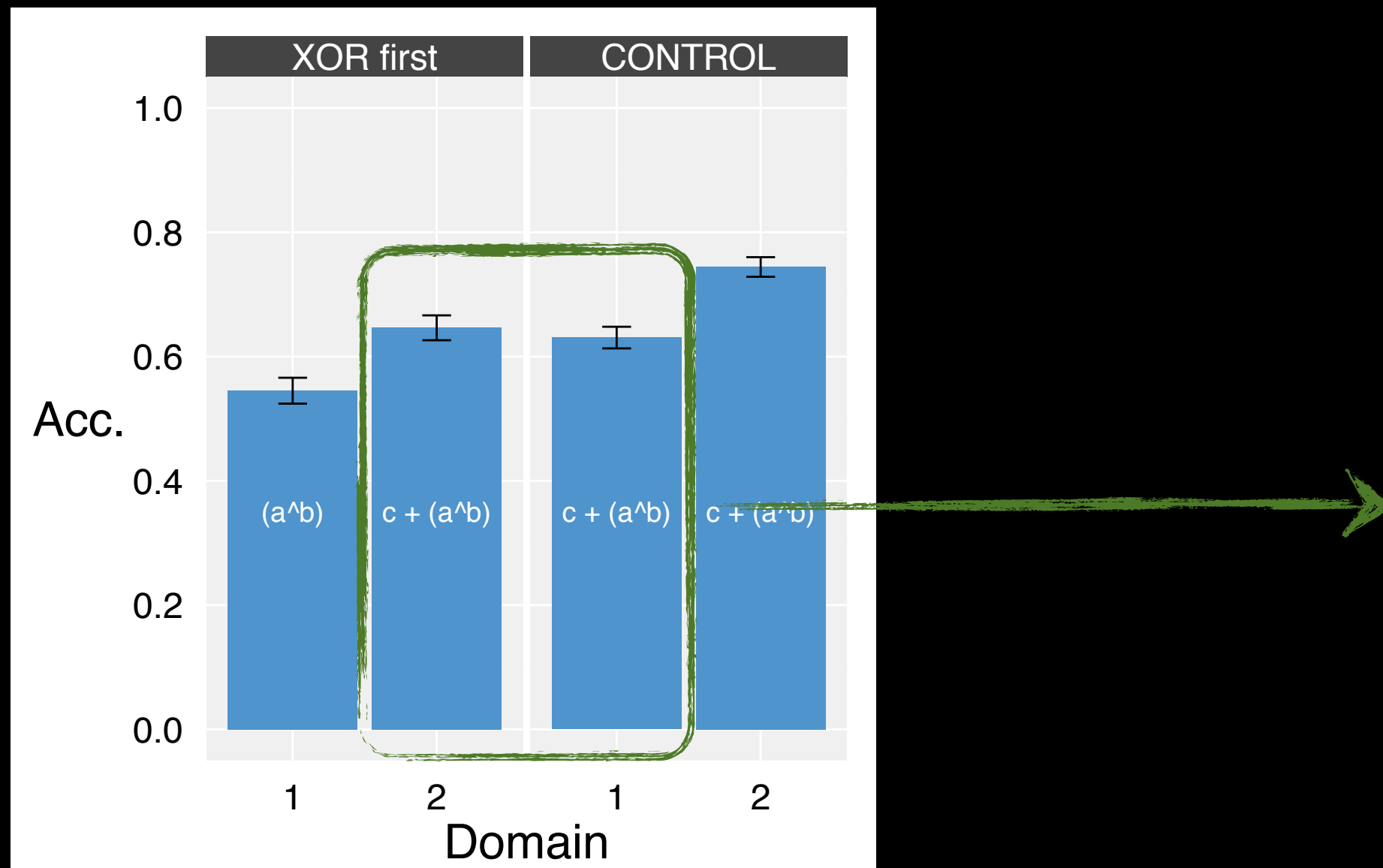
Results I: garden pathing

Does learning $a \oplus b$ transfer to $c + (a \oplus b)$?



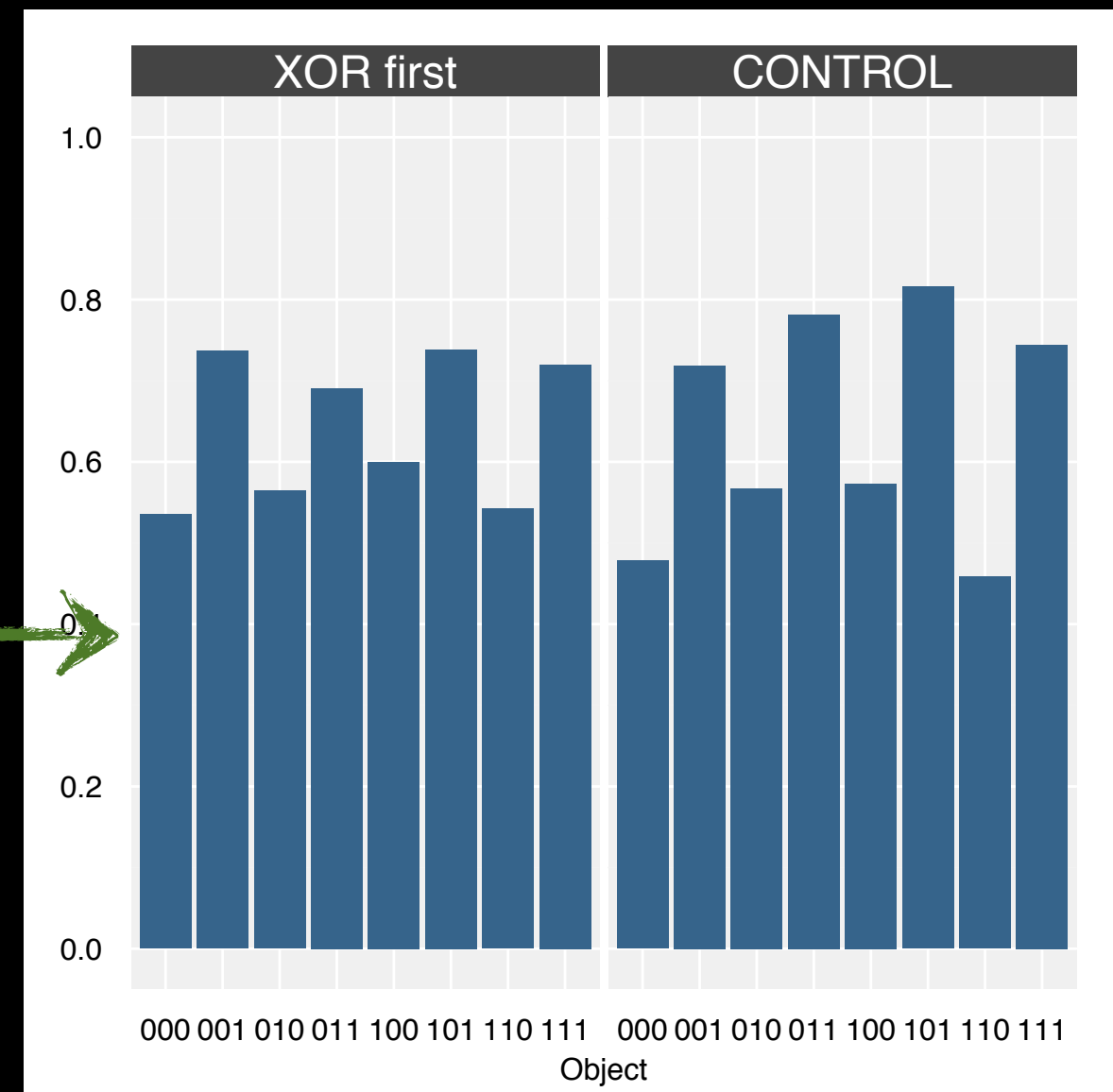
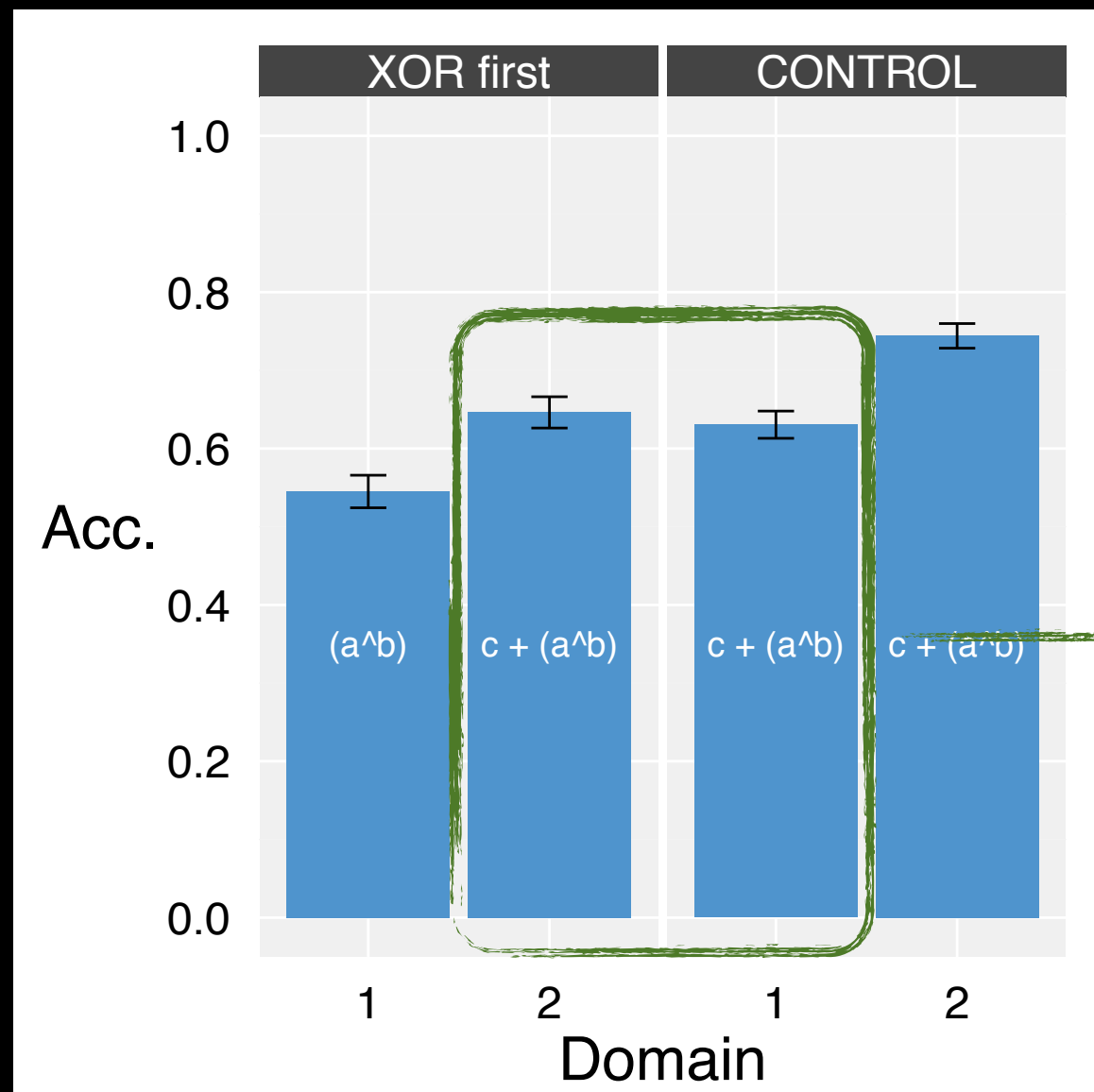
Results I: garden pathing

Does learning $a \oplus b$ transfer to $c + (a \oplus b)$?



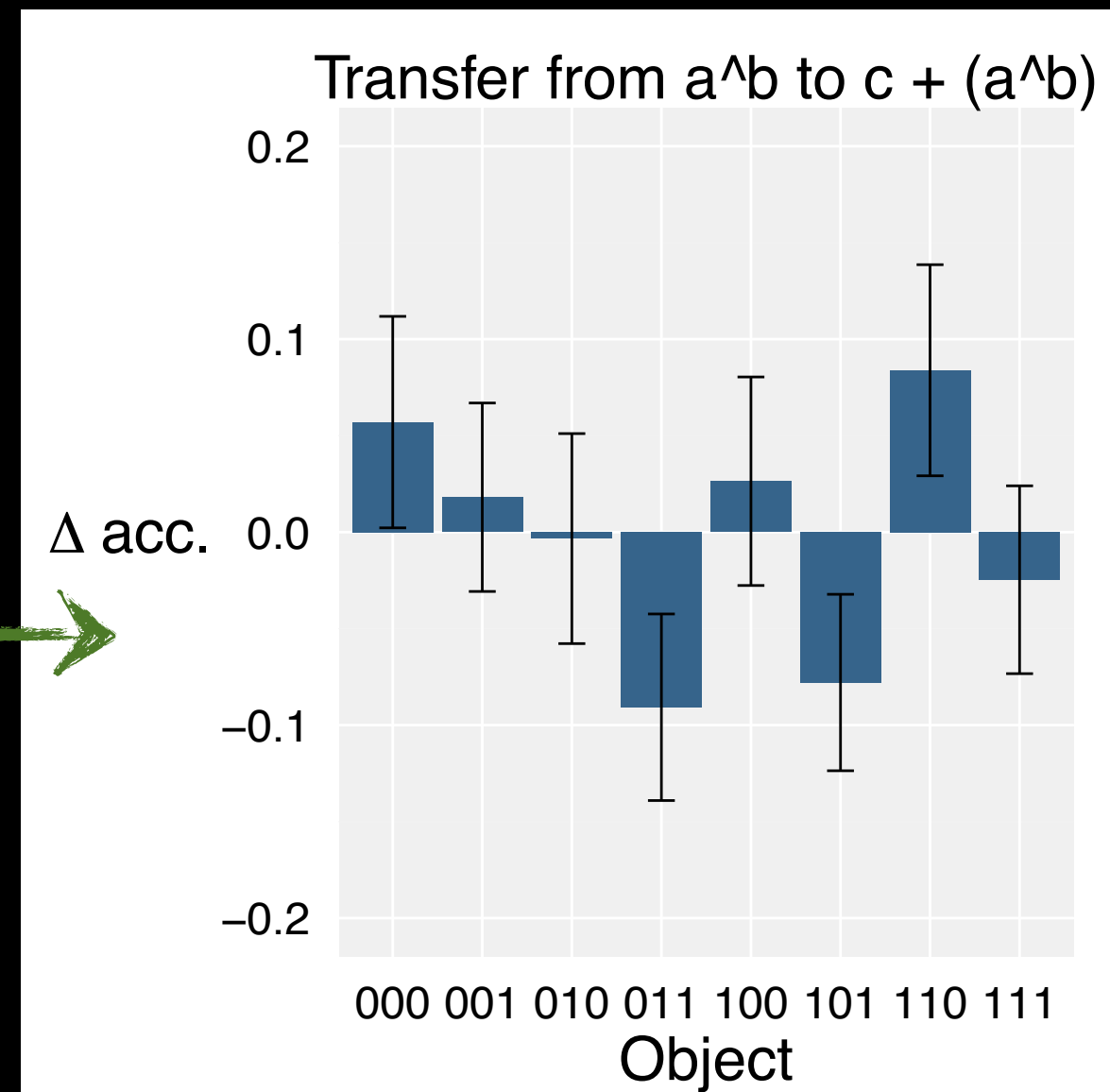
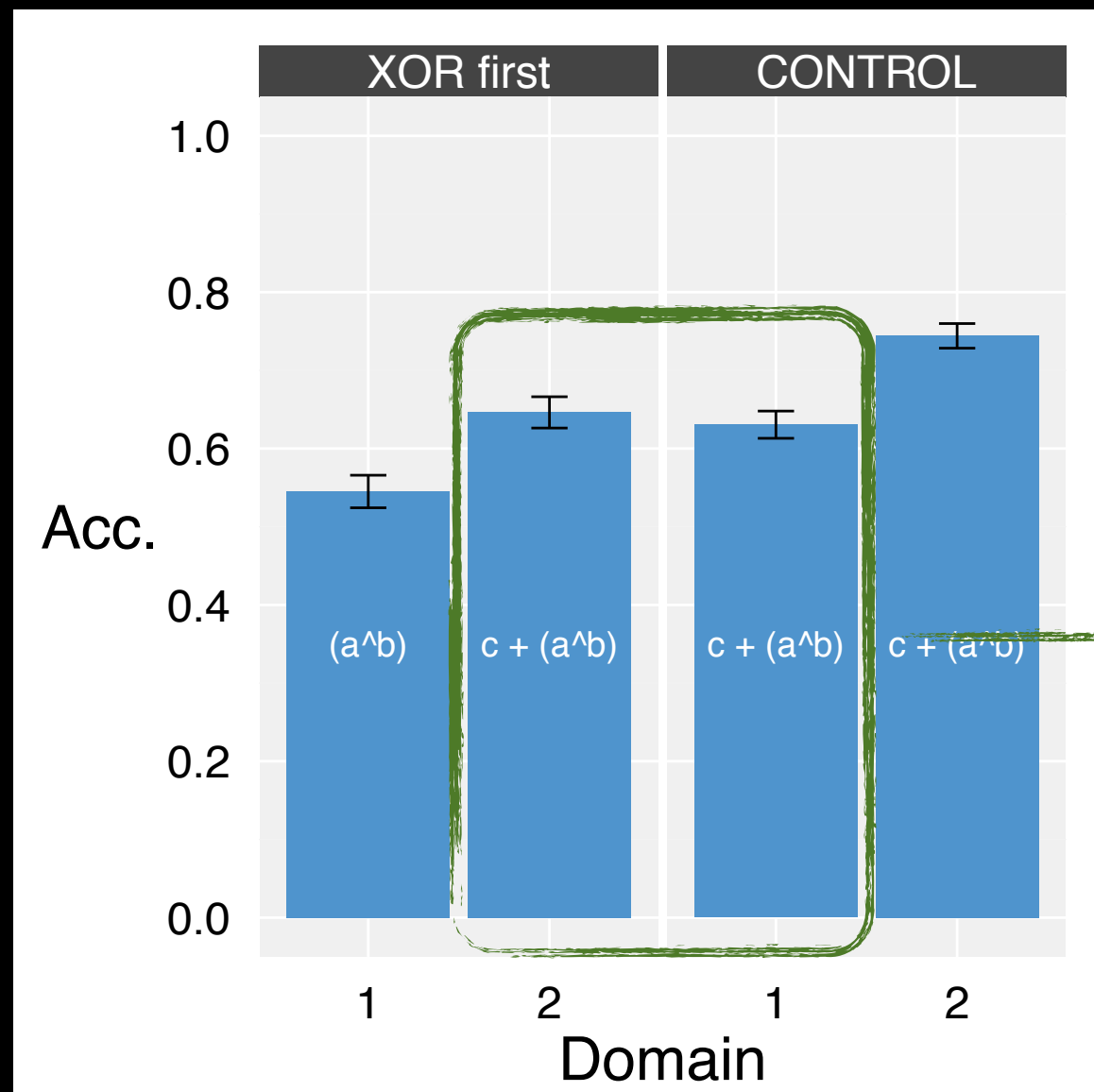
Results I: garden pathing

Does learning $a \oplus b$ transfer to $c + (a \oplus b)$?



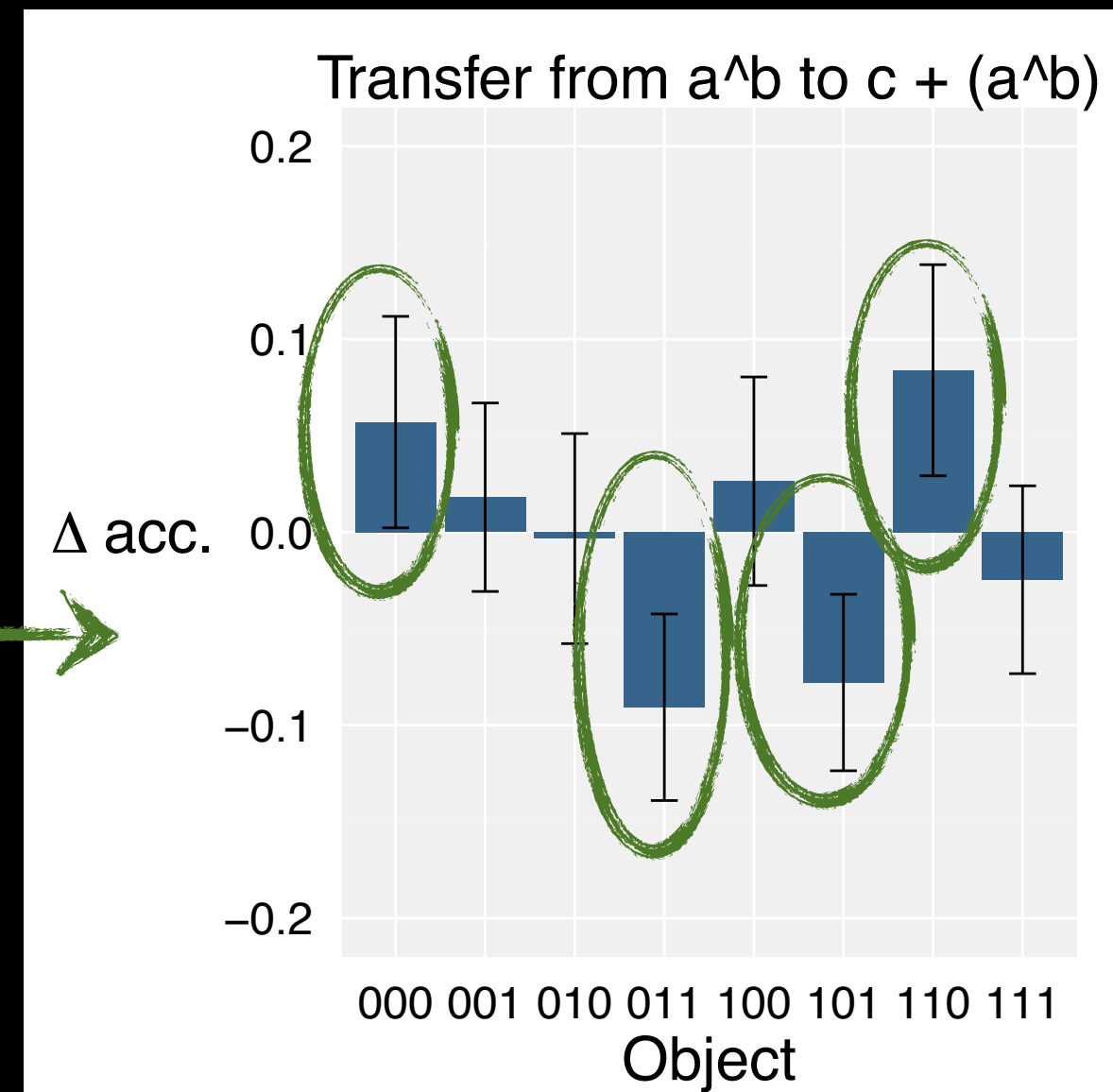
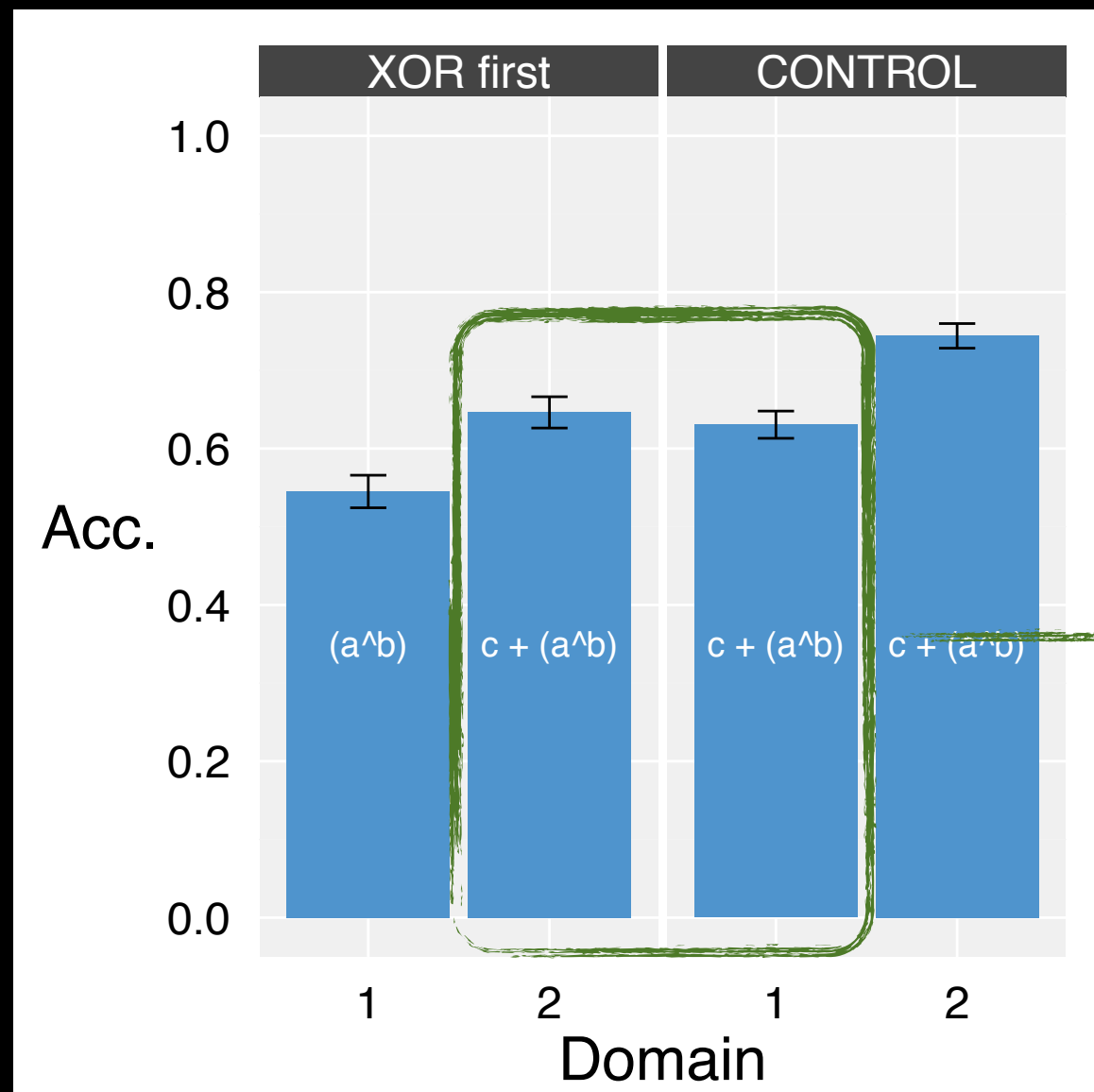
Results I: garden pathing

Does learning $a \wedge b$ transfer to $c + (a \wedge b)$?

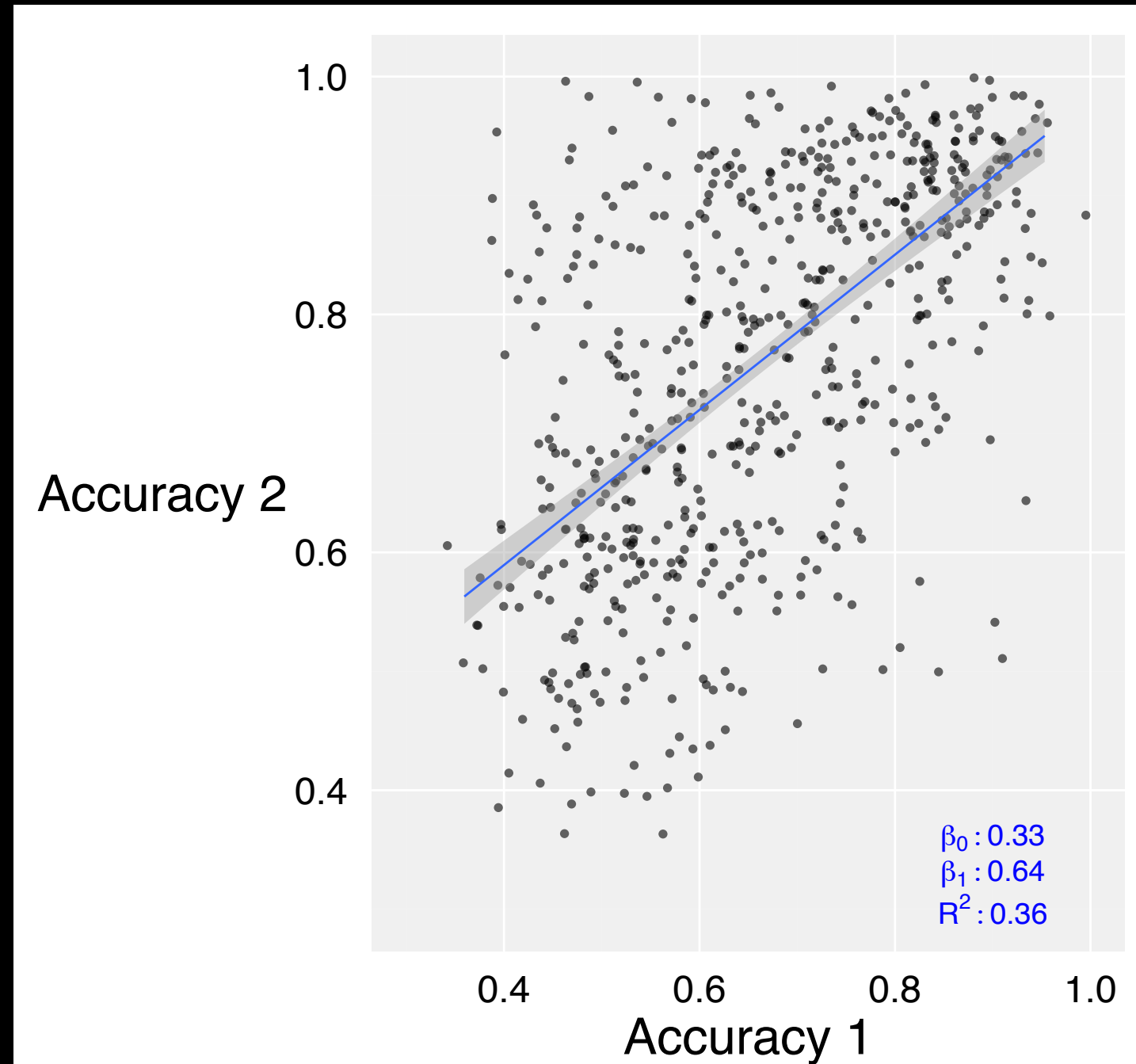


Results I: garden pathing

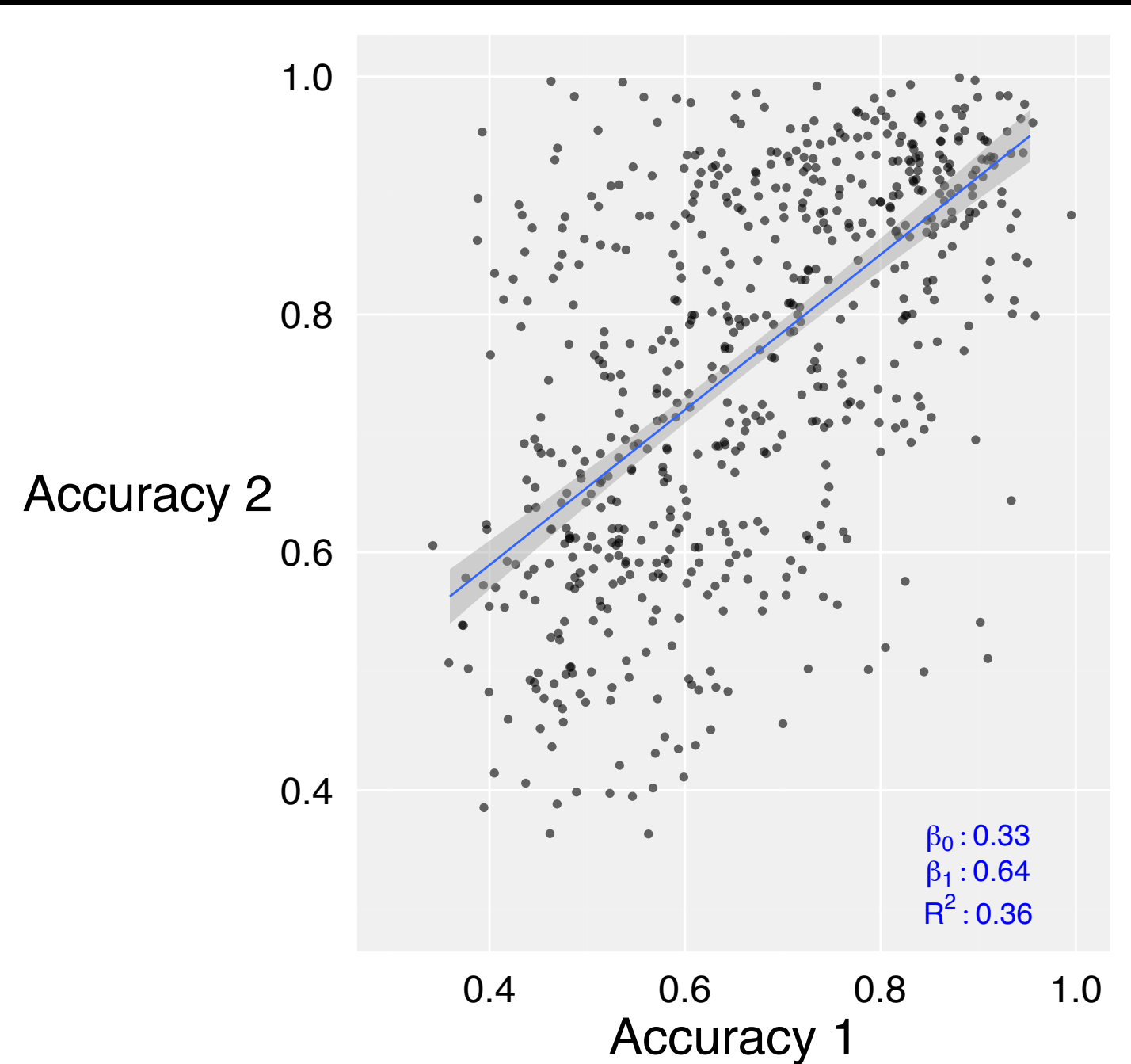
Does learning $a \wedge b$ transfer to $c + (a \wedge b)$?



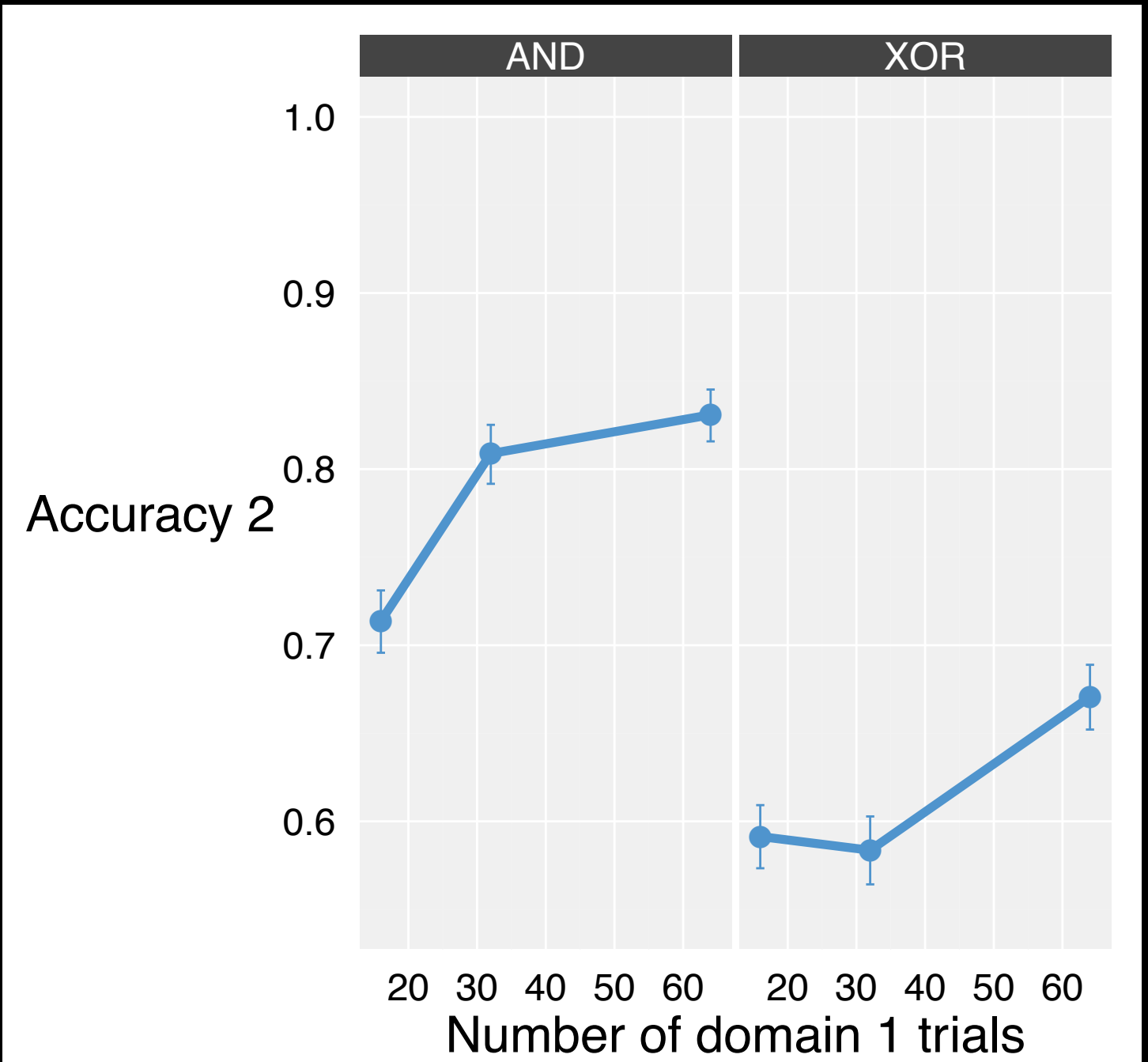
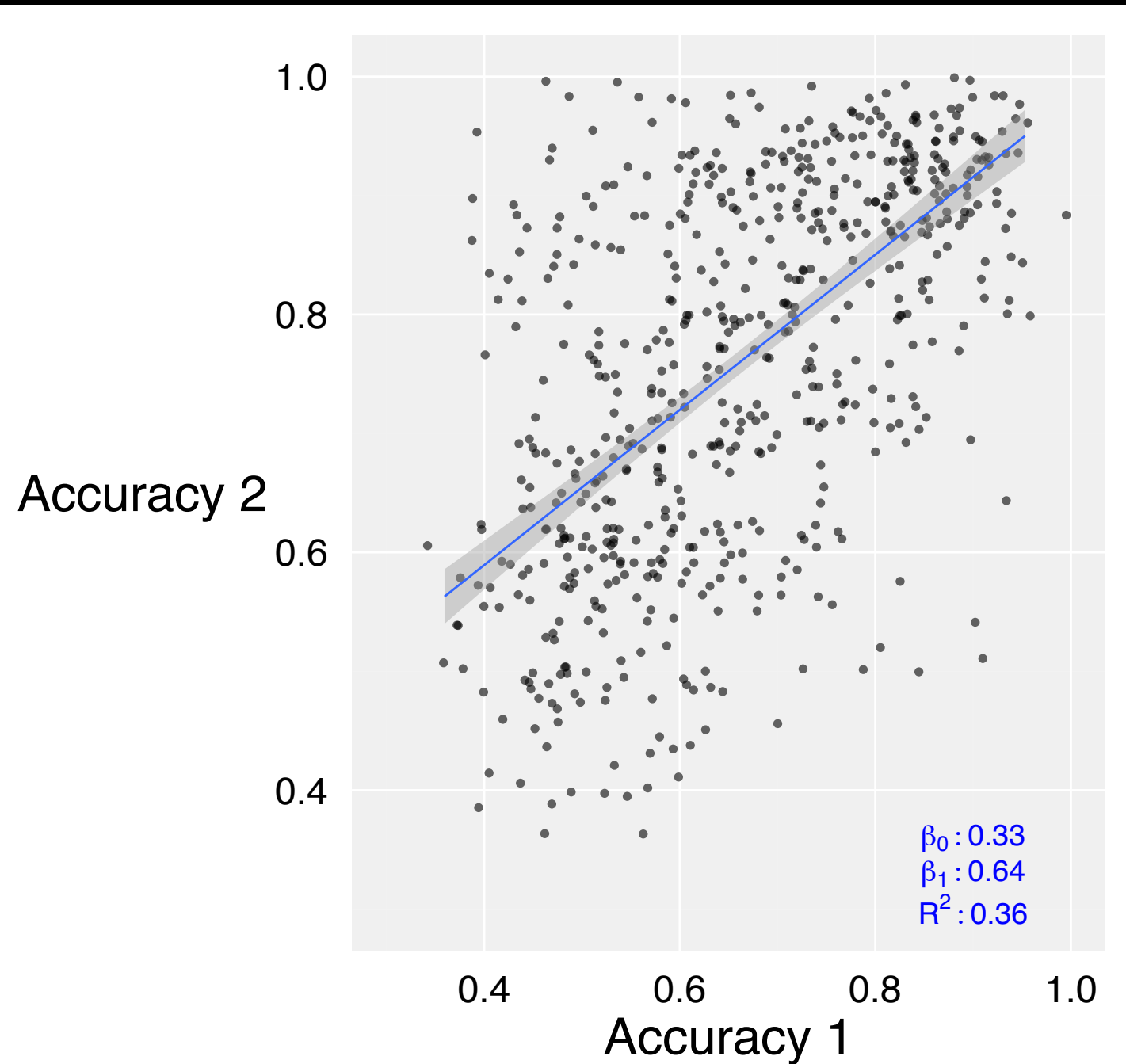
Results I: learning effect



Results I: learning effect



Results I: learning effect



Results I

- positive self-transfer
- garden pathing
- learning effect

Model: Rational Rules

Hypothesis space: DNF classifiers

Prior: Syntactic complexity, given by PCFG

Likelihood: Ability to generate correct labels

Reuse capability: Upweight production probs in PCFG

Goodman et al. (2008)

DNF Grammar: hypotheses and prior

DNF Grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

DNF Grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

DNF Grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

hook | hōok |

noun

- 1 a piece of metal or other material, curved or
• (also **fishhook**) a bent piece of metal, typical
- 2 a thing designed to catch people's attention:
• a chorus or repeated instrumental passage
- 3 a curved cutting instrument, esp. as used for 1
- 4 a short swinging punch made with the elbow
• Golf a stroke that makes the ball deviate in :
- 5 a curved stroke in handwriting, esp. as made
• Music an added stroke transverse to the stem
- 6 [usu. in place names] a curved promontory

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

hook | hōok |

noun

- 1 a piece of metal or other material, curved or
• (also **fishhook**) a bent piece of metal, typical
- 2 a thing designed to catch people's attention:
• a chorus or repeated instrumental passage
- 3 a curved cutting instrument, esp. as used for 1
- 4 a short swinging punch made with the elbow
• Golf a stroke that makes the ball deviate in :
- 5 a curved stroke in handwriting, esp. as made
• Music an added stroke transverse to the stem
- 6 [usu. in place names] a curved promontory

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

hook | hōok |

noun

- 1 a piece of metal or other material, curved or
 - (also **fishhook**) a bent piece of metal, typical
- 2 a thing designed to catch people's attention:
 - a chorus or repeated instrumental passage
- 3 a curved cutting instrument, esp. as used for 1
- 4 a short swinging punch made with the elbow
 - Golf a stroke that makes the ball deviate in :
- 5 a curved stroke in handwriting, esp. as made
 - Music an added stroke transverse to the ster
- 6 [usu. in place names] a curved promontory

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

e.g., curved \wedge used for shearing

hook | hōok |

noun

- 1 a piece of metal or other material, curved or
 - (also **fishhook**) a bent piece of metal, typical
- 2 a thing designed to catch people's attention:
 - a chorus or repeated instrumental passage
- 3 a curved cutting instrument, esp. as used for 1
- 4 a short swinging punch made with the elbow
 - Golf a stroke that makes the ball deviate in :
- 5 a curved stroke in handwriting, esp. as made
 - Music an added stroke transverse to the ster
- 6 [usu. in place names] a curved promontory

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

Feature \rightarrow F_a | F_b | F_c

F_a \rightarrow $a = 0$ | $a = 1$

F_b \rightarrow $b = 0$ | $b = 1$

F_c \rightarrow $c = 0$ | $c = 1$

hook | hōok |

noun

- 1 a piece of metal or other material, curved or
• (also **fishhook**) a bent piece of metal, typical
- 2 a thing designed to catch people's attention:
• a chorus or repeated instrumental passage
- 3 a curved cutting instrument, esp. as used for 1
- 4 a short swinging punch made with the elbow
• Golf a stroke that makes the ball deviate in :
- 5 a curved stroke in handwriting, esp. as made
• Music an added stroke transverse to the stem
- 6 [usu. in place names] a curved promontory

DNF grammar: hypotheses and prior

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label $\rightarrow \mathbf{p} \mid \mathbf{q}$ $p \sim \text{Dir}(\alpha_1, \alpha_1)$

Rule \rightarrow *Sense* \mid *Sense* \vee *Rule* $p \sim \text{Dir}(\alpha_2, \alpha_3)$

Sense \rightarrow *Feature* \mid *Feature* \wedge *Sense* $p \sim \text{Dir}(\alpha_4, \alpha_5)$

Feature $\rightarrow F_a \mid F_b \mid F_c$ $p \sim \text{Dir}(\alpha_6, \alpha_7, \alpha_8)$

$F_a \rightarrow a = 0 \mid a = 1$

$F_b \rightarrow b = 0 \mid b = 1$

$F_c \rightarrow c = 0 \mid c = 1$

Dirichlet-Multinomial

Dirichlet-Multinomial

$$\textit{Label} \rightarrow \mathbf{p} \mid \mathbf{q} \qquad p \sim \text{Dir}(0.5, 0.5)$$

Dirichlet-Multinomial

Label

$\rightarrow \mathbf{p} \mid \mathbf{q}$

$p \sim \text{Dir}(0.5, 0.5)$

Dirichlet-Multinomial

$$\textit{Label} \rightarrow \mathbf{p} \mid \mathbf{q} \qquad p \sim \text{Dir}(0.5, 0.5)$$

$$p(2 \text{ } \mathbf{p}'\text{s}) = 0.375$$

Dirichlet-Multinomial

$$\textit{Label} \rightarrow \mathbf{p} \mid \mathbf{q} \qquad p \sim \text{Dir}(0.5, 0.5)$$

$$p(2 \text{ } \mathbf{p}\text{'s}) = 0.375$$

$$p(2 \text{ } \mathbf{q}\text{'s}) = 0.375$$

Dirichlet-Multinomial

$$\textit{Label} \rightarrow \mathbf{p} \mid \mathbf{q} \qquad p \sim \text{Dir}(0.5, 0.5)$$

$$p(2 \text{ } \mathbf{p}\text{'s}) = 0.375$$

$$p(2 \text{ } \mathbf{q}\text{'s}) = 0.375$$

$$p(1 \text{ } \mathbf{p}, 1 \text{ } \mathbf{q}) = 0.25$$

Dirichlet-Multinomial

$$\textit{Label} \rightarrow \mathbf{p} \mid \mathbf{q} \qquad p \sim \text{Dir}(0.5, 0.5)$$

$$p(2 \text{ } \mathbf{p}\text{'s}) = 0.375$$

$$p(2 \text{ } \mathbf{q}\text{'s}) = 0.375$$

$$p(1 \text{ } \mathbf{p}, 1 \text{ } \mathbf{q}) = 0.25$$

Dirichlet-Multinomial

$$\textit{Label} \quad \rightarrow \mathbf{p} \mid \mathbf{q} \quad \quad p \sim \text{Dir}(0.5, 0.5)$$

$$p(2 \text{ } \mathbf{p}\text{'s}) = 0.375$$

$$p(2 \text{ } \mathbf{q}\text{'s}) = 0.375$$

$$p(1 \text{ } \mathbf{p}, 1 \text{ } \mathbf{q}) = 0.25$$

$$p(2 \text{ } \mathbf{q}\text{'s} \mid 1 \text{ } \mathbf{p}) = 0.75$$

Likelihood, $p(d \mid h)$

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Given x_i 's, correctly generate all $(x_i, h(x_i))$ pairs

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Given x_i 's, correctly generate all $(x_i, h(x_i))$ pairs

With probability η , can generate “outlier”

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Given x_i 's, correctly generate all $(x_i, h(x_i))$ pairs

With probability η , can generate “outlier”

$$p(d \mid h) \propto \eta^{(\# \text{ wrong})}$$

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Given x_i 's, correctly generate all $(x_i, h(x_i))$ pairs

With probability η , can generate “outlier”

$$p(d \mid h) \propto \eta^{(\# \text{ wrong})}$$

Decay: past data points matter less (power law)

Likelihood, $p(d \mid h)$

d is labeled data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Given x_i 's, correctly generate all $(x_i, h(x_i))$ pairs

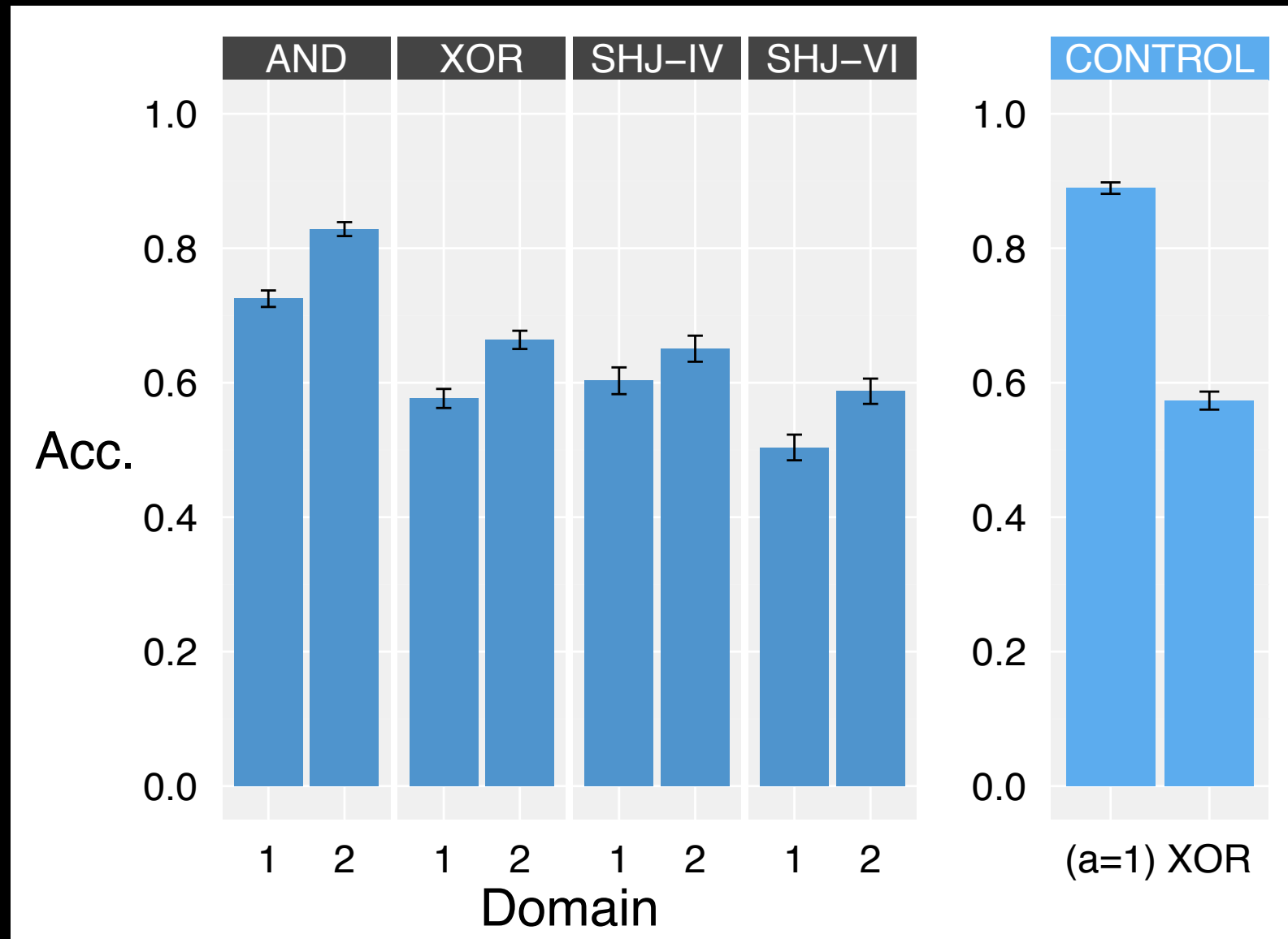
With probability η , can generate “outlier”

$$p(d \mid h) \propto \eta^{(\# \text{ wrong})}$$

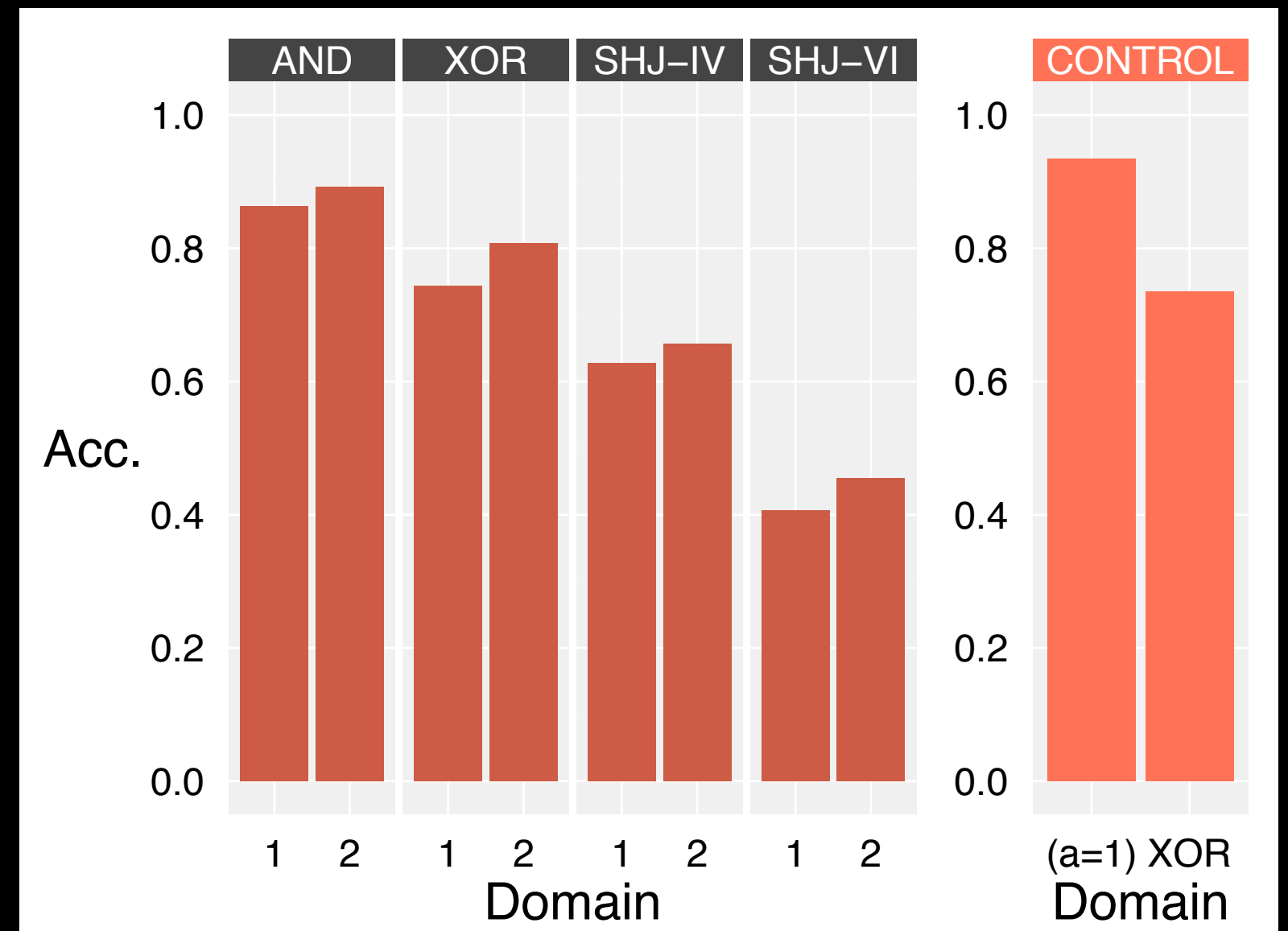
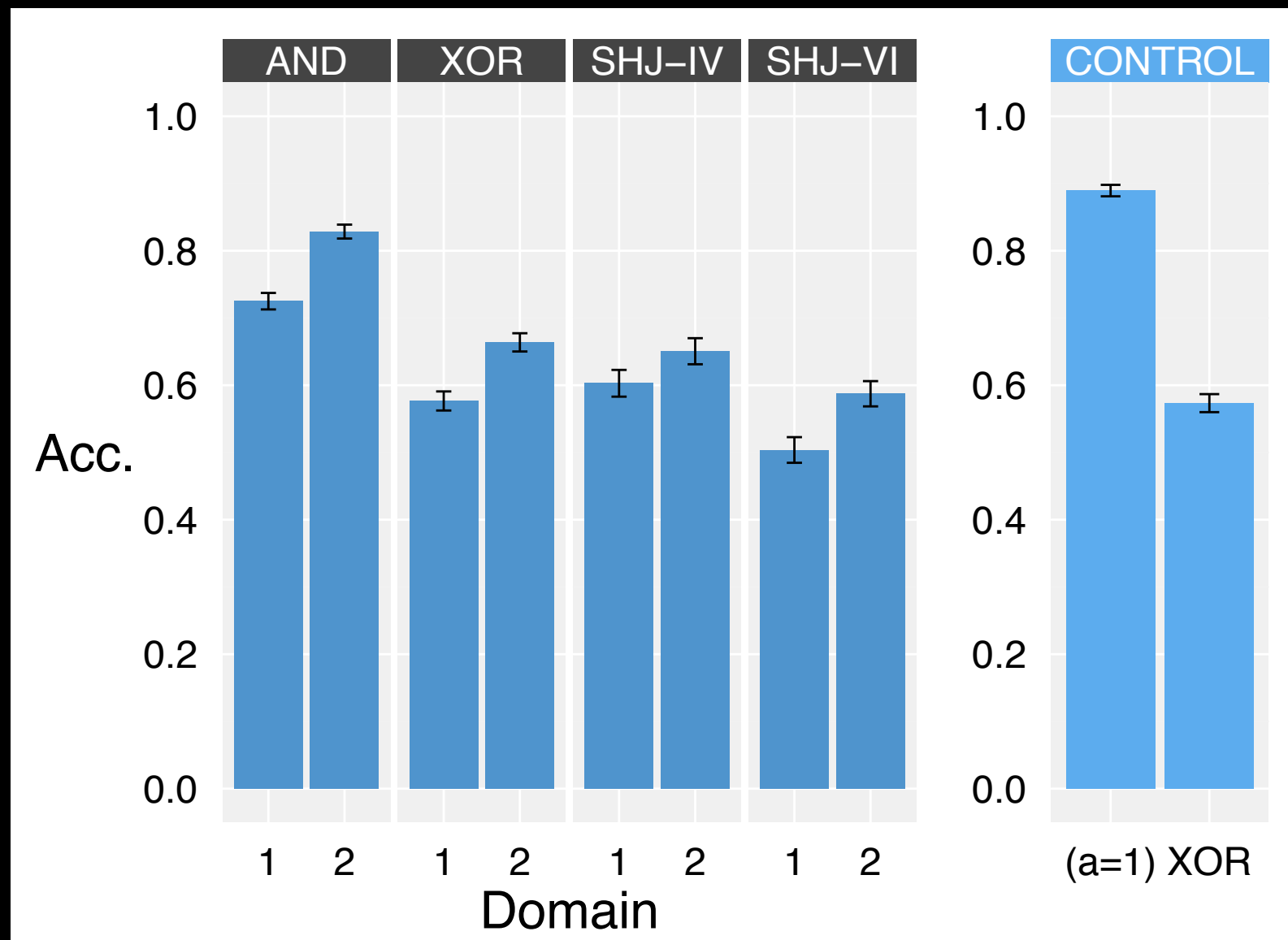
Decay: past data points matter less (power law)

Stochasticity: h can noisily generate the “wrong” label

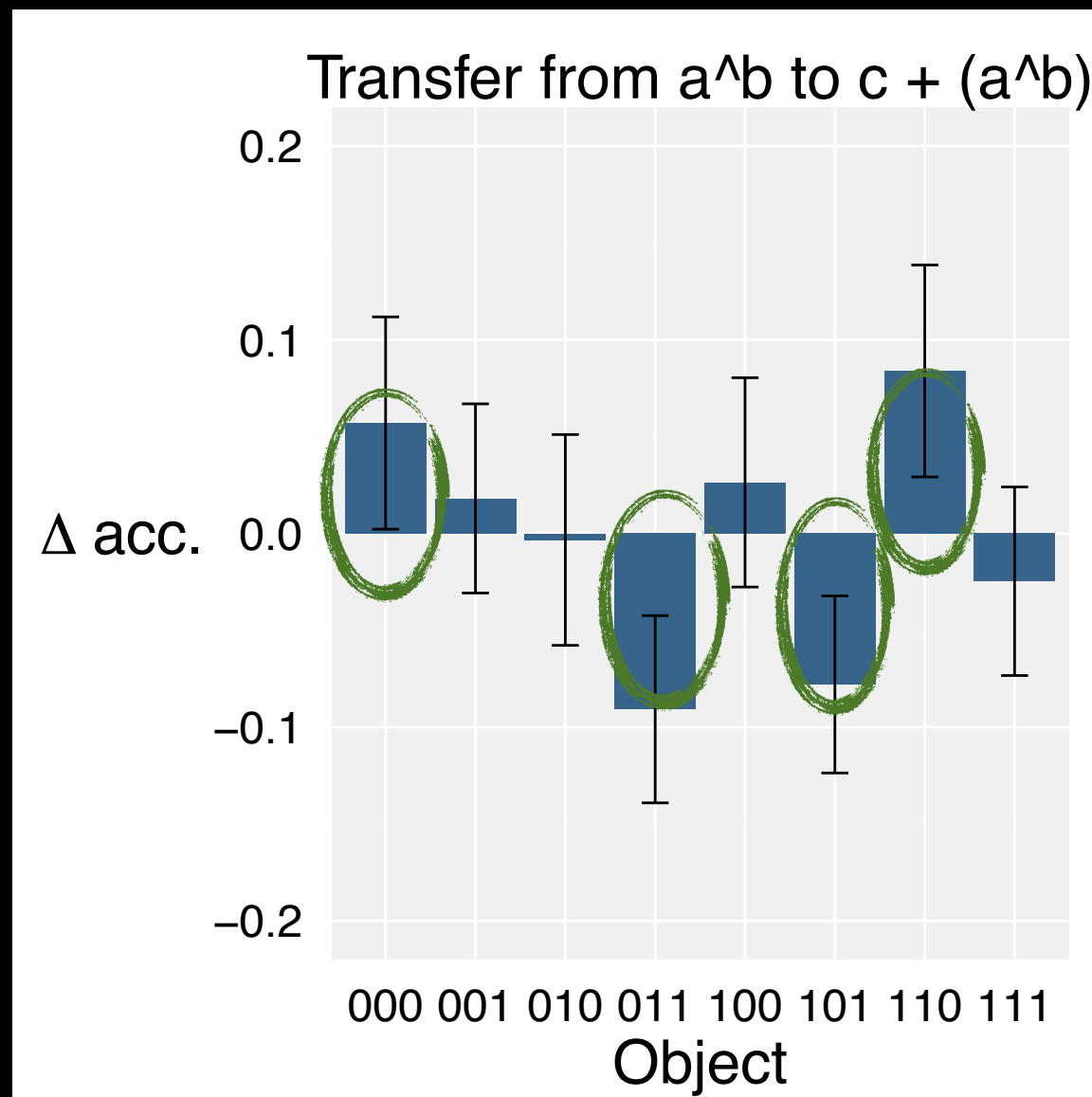
Model results: + self-transfer



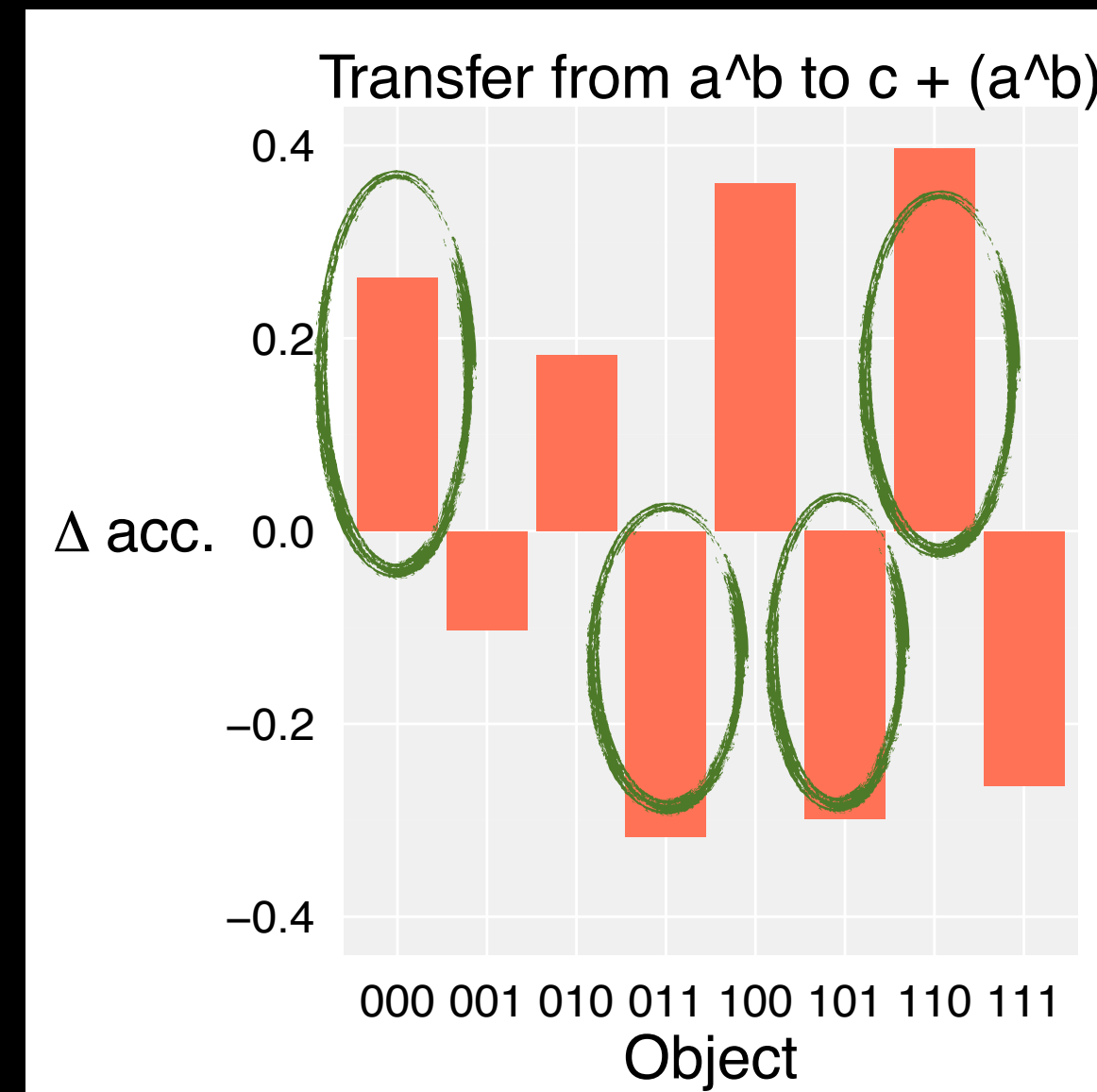
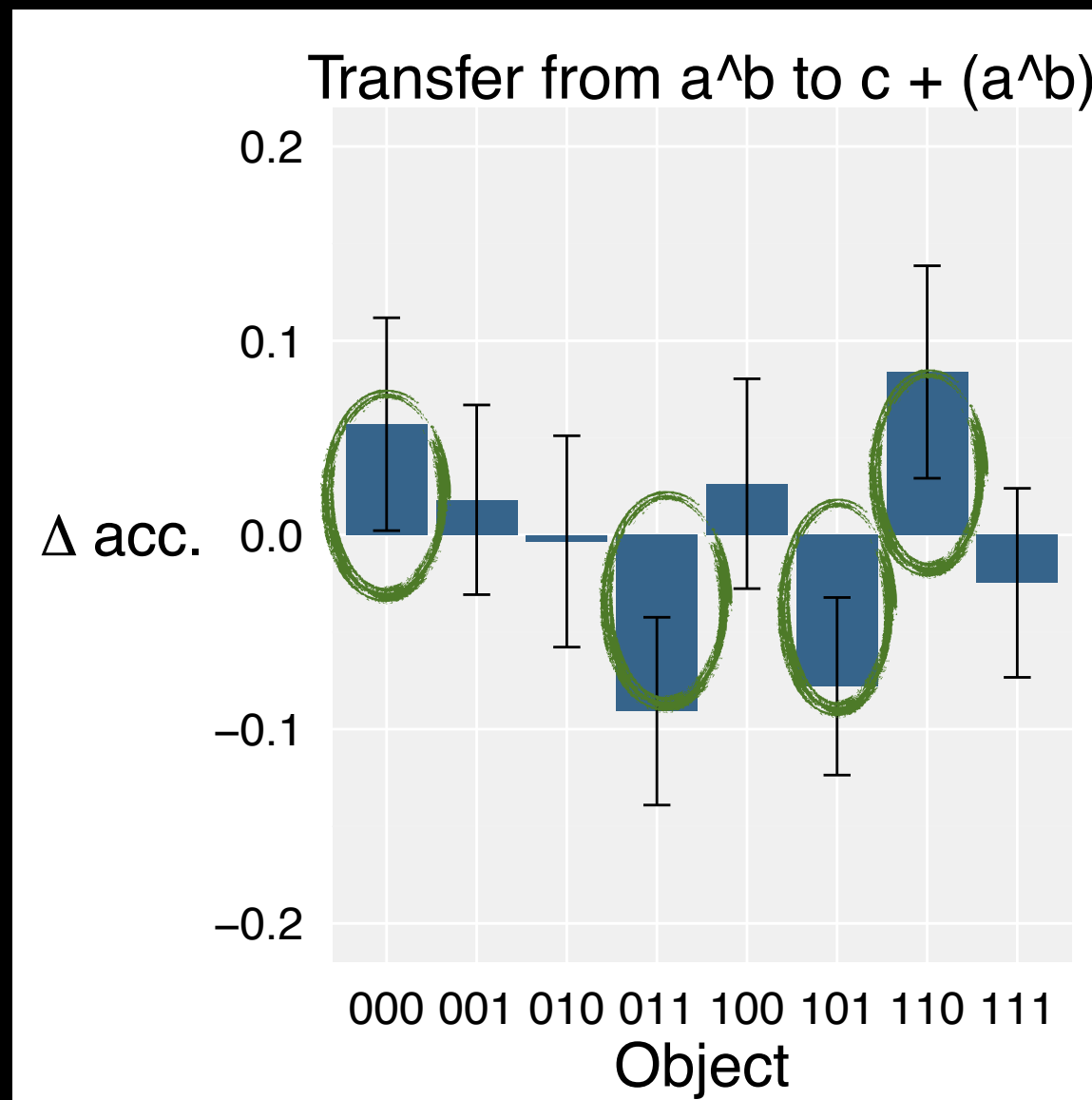
Model results: + self-transfer



Model results: garden pathing

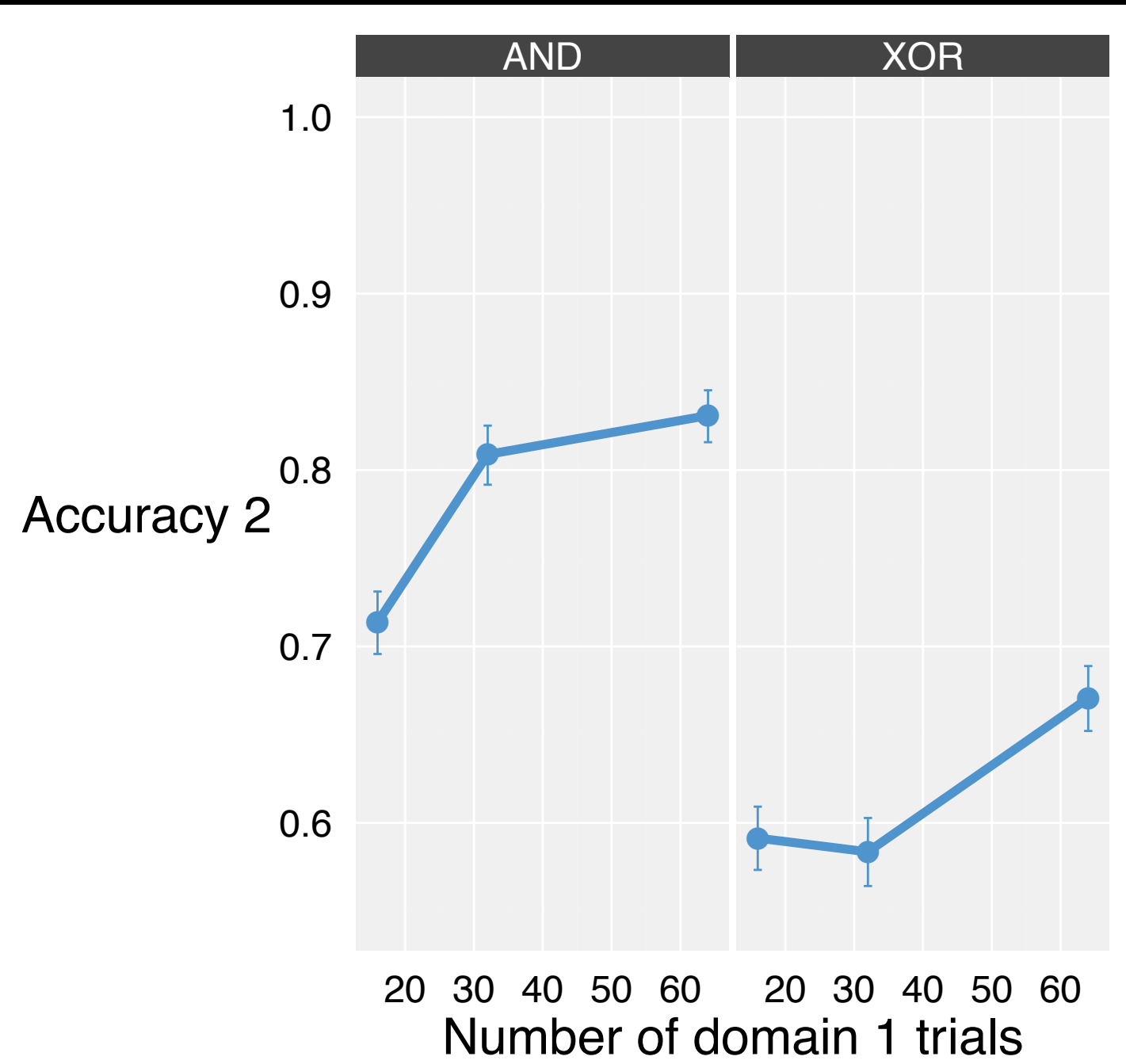


Model results: garden pathing

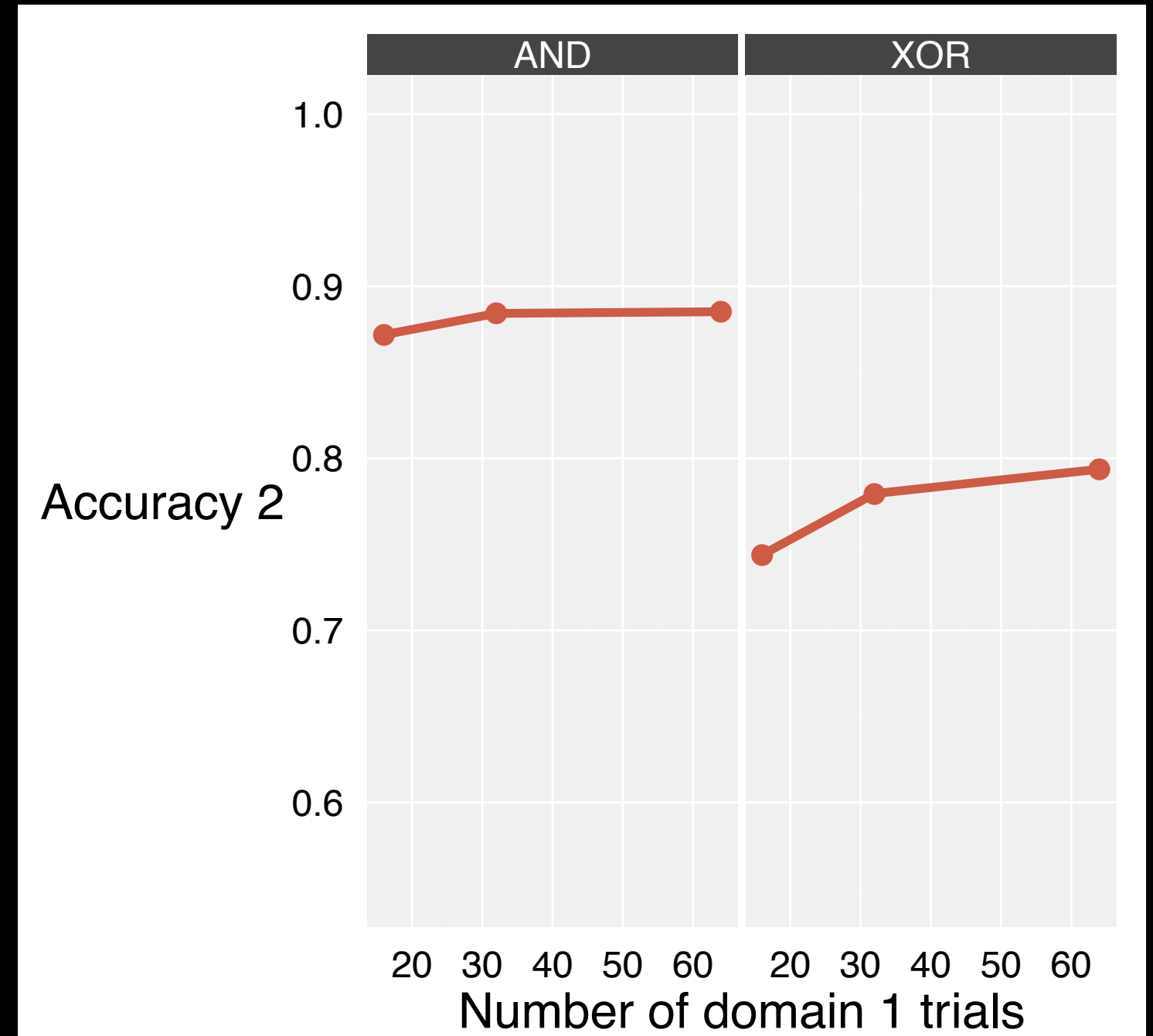
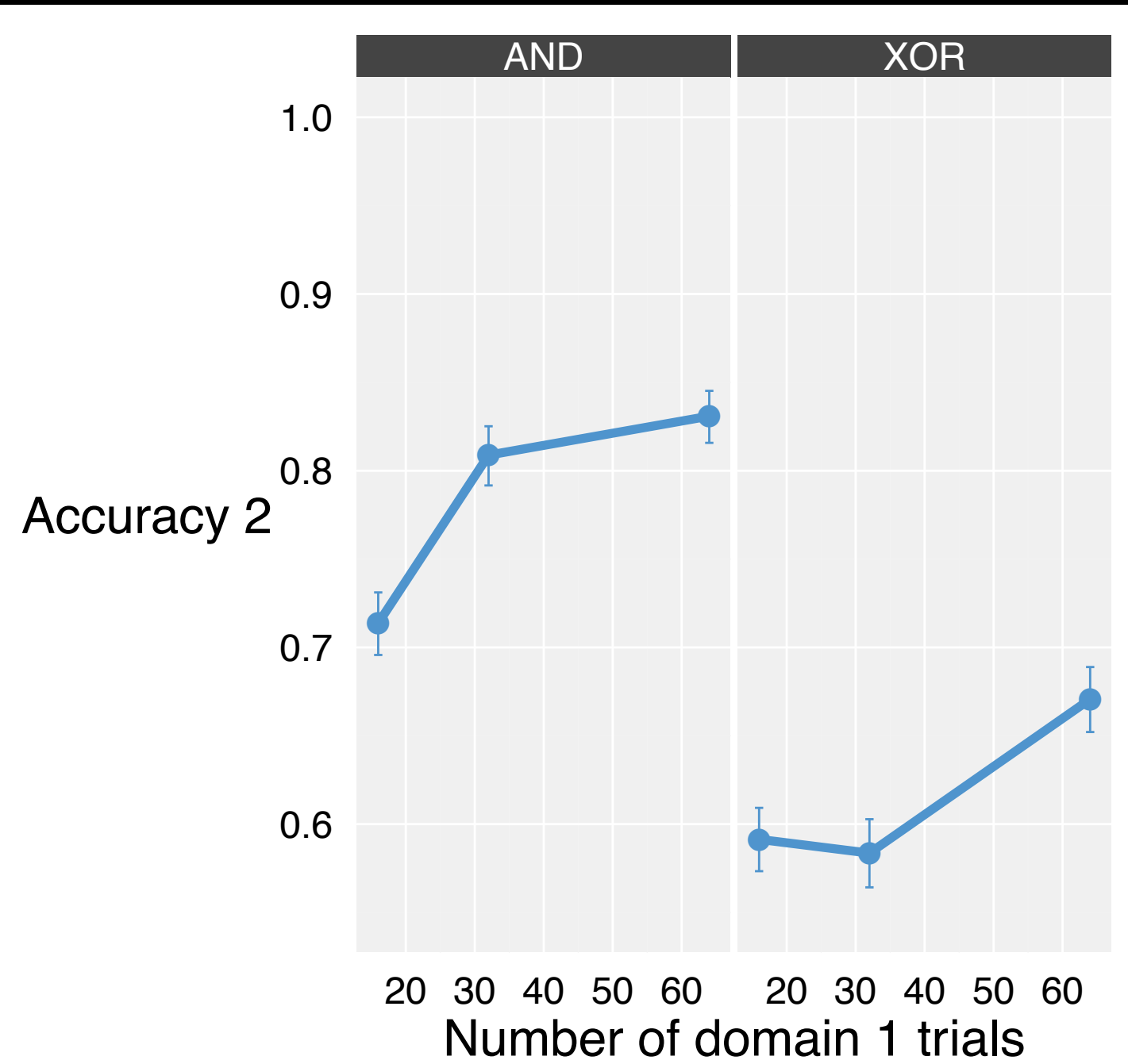


$$R^2 = 0.76, p < 0.01$$

Model results: learning effect



Model results: learning effect



Rational Rules so far

- ☑ positive self-transfer
- ☑ garden pathing
- ☑ learning effect
- ?? compositional transfer

Dax conditions

Dax conditions

Shorthand

Dax conditions

Shorthand	Formula I
-----------	-----------

Dax conditions

Shorthand	Formula I	Cube I
-----------	-----------	--------

Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2
-----------	-----------	--------	-----------

Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
-----------	-----------	--------	-----------	--------


Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax				


Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab			



Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax		



Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	



Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax




Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)				




Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab			





Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab			





Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		(abc) '	





Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		(abc)'	






Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		$(abc)'$	
Label flip				






Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		(abc)'	
Label flip	abc			







Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		$(abc)'$	
Label flip	abc			

Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		$(abc)'$	
Label flip	abc		$(abc)'$	

Dax conditions

Shorthand	Formula 1	Cube 1	Formula 2	Cube 2
Red dax	ab	 dax	abc	 red dax
Not (red dax)	ab		$(abc)'$	
Label flip	abc		$(abc)'$	

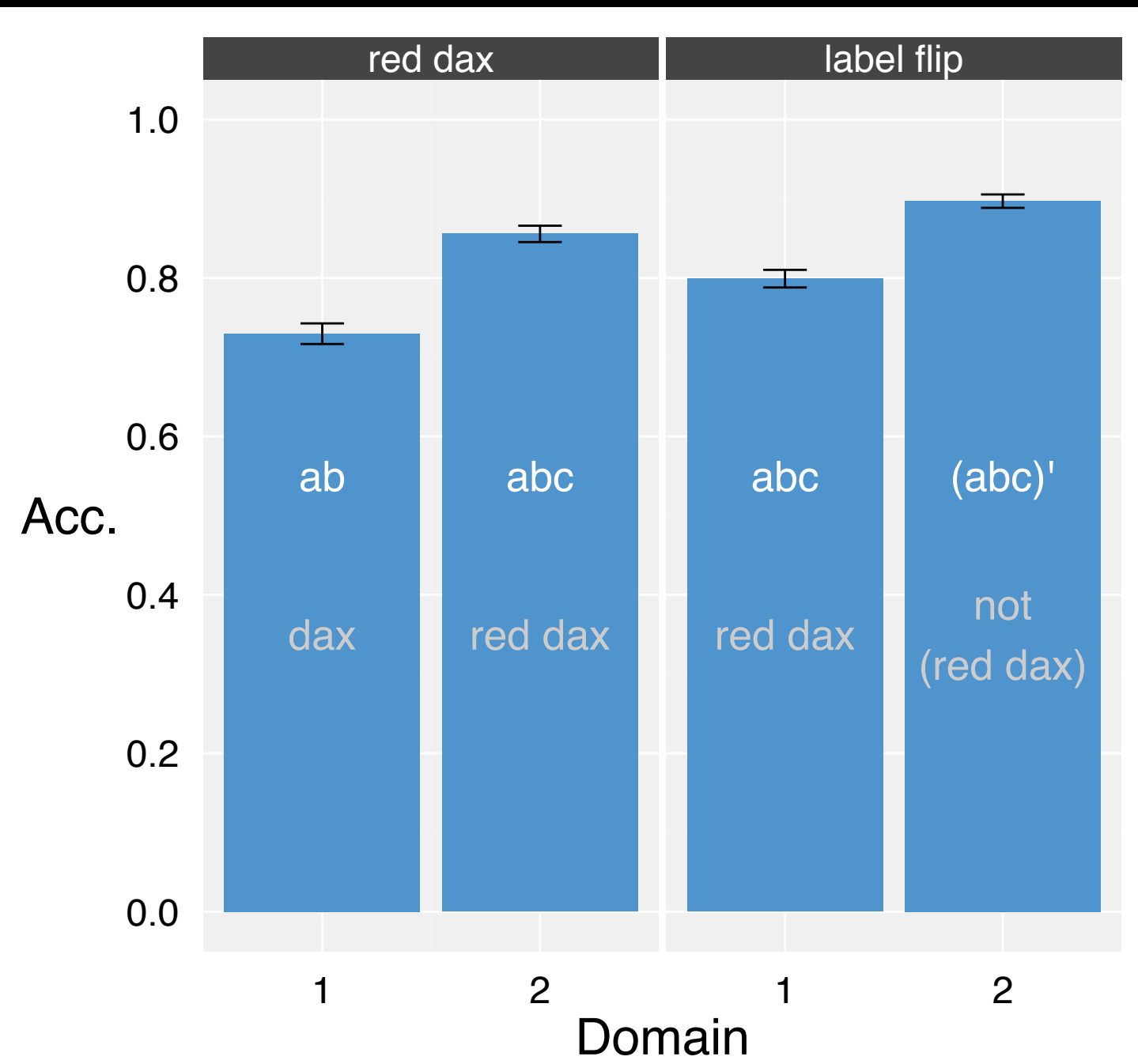
Dax results (human)

Dax results (human)

+ T from dax to red dax

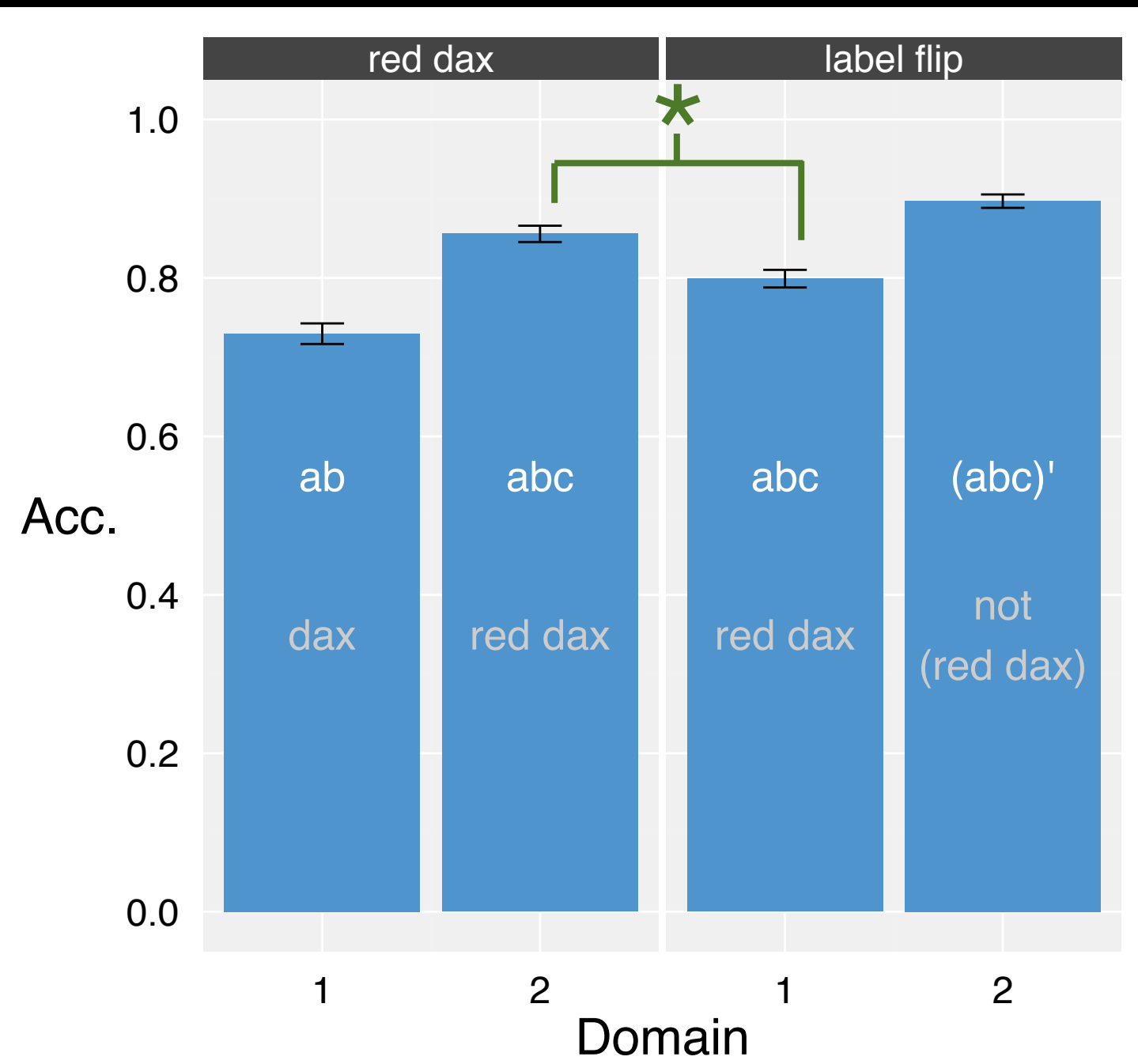
Dax results (human)

+ T from dax to red dax



Dax results (human)

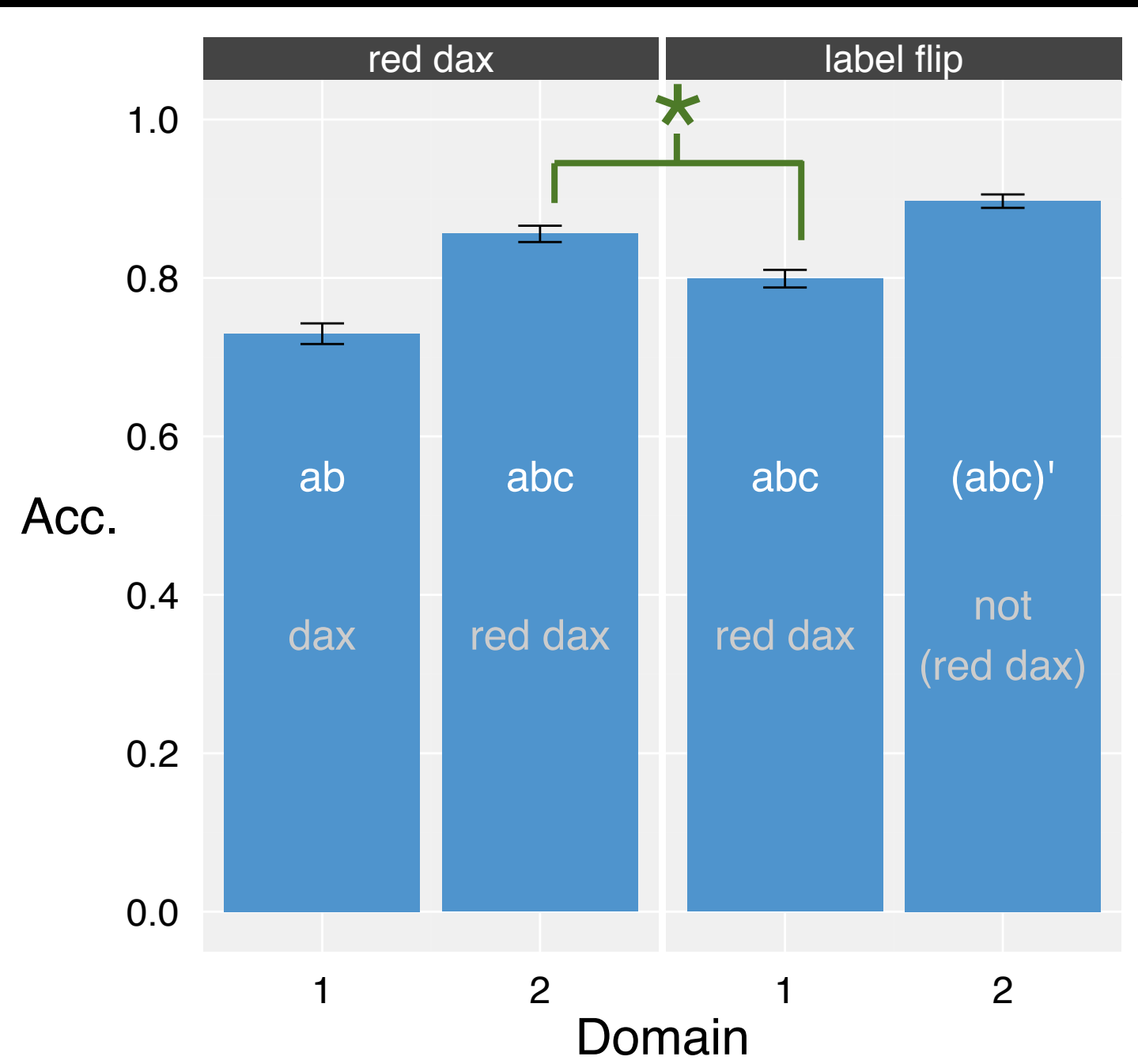
+ T from dax to red dax



Dax results (human)

+ T from dax to red dax

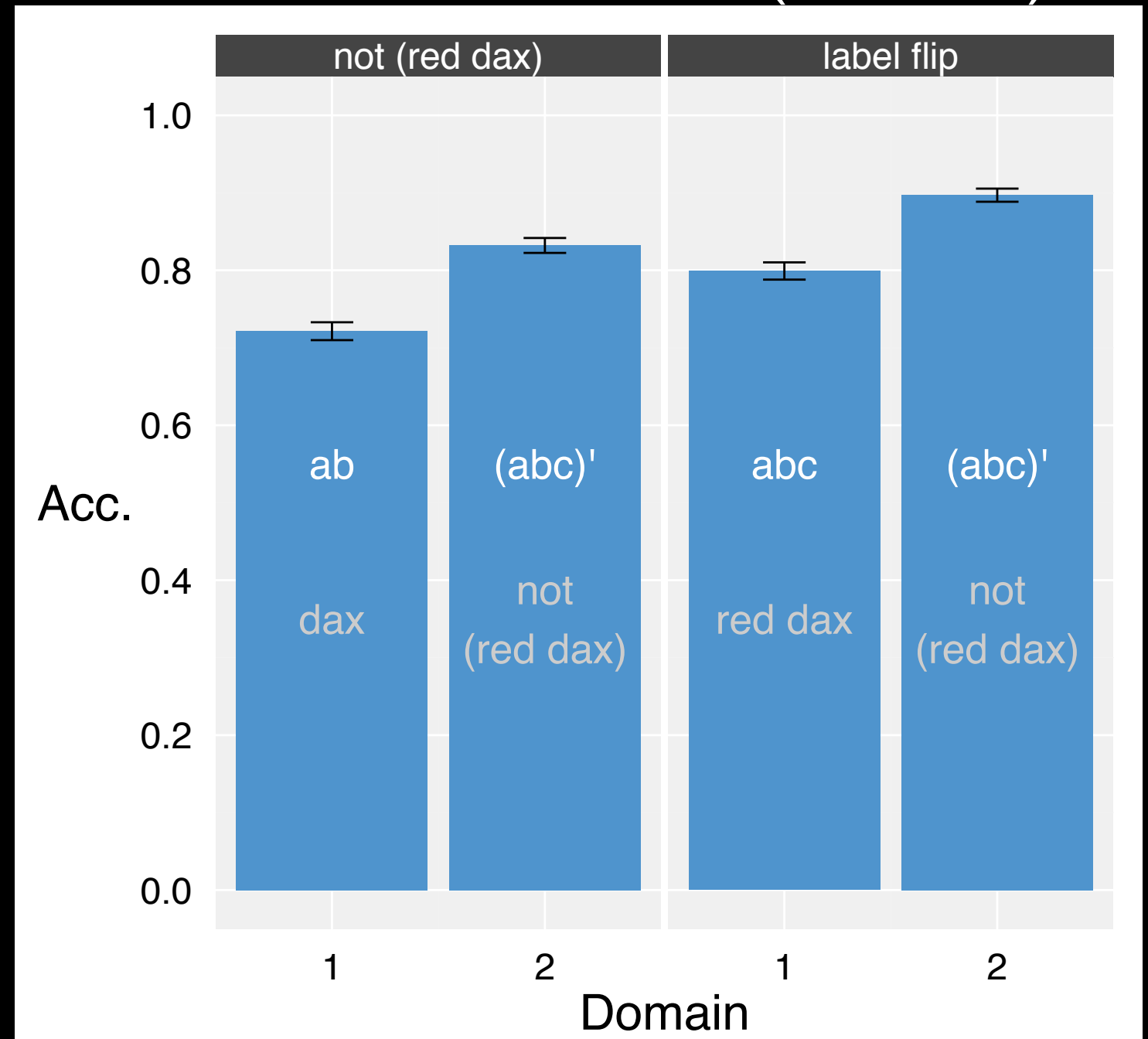
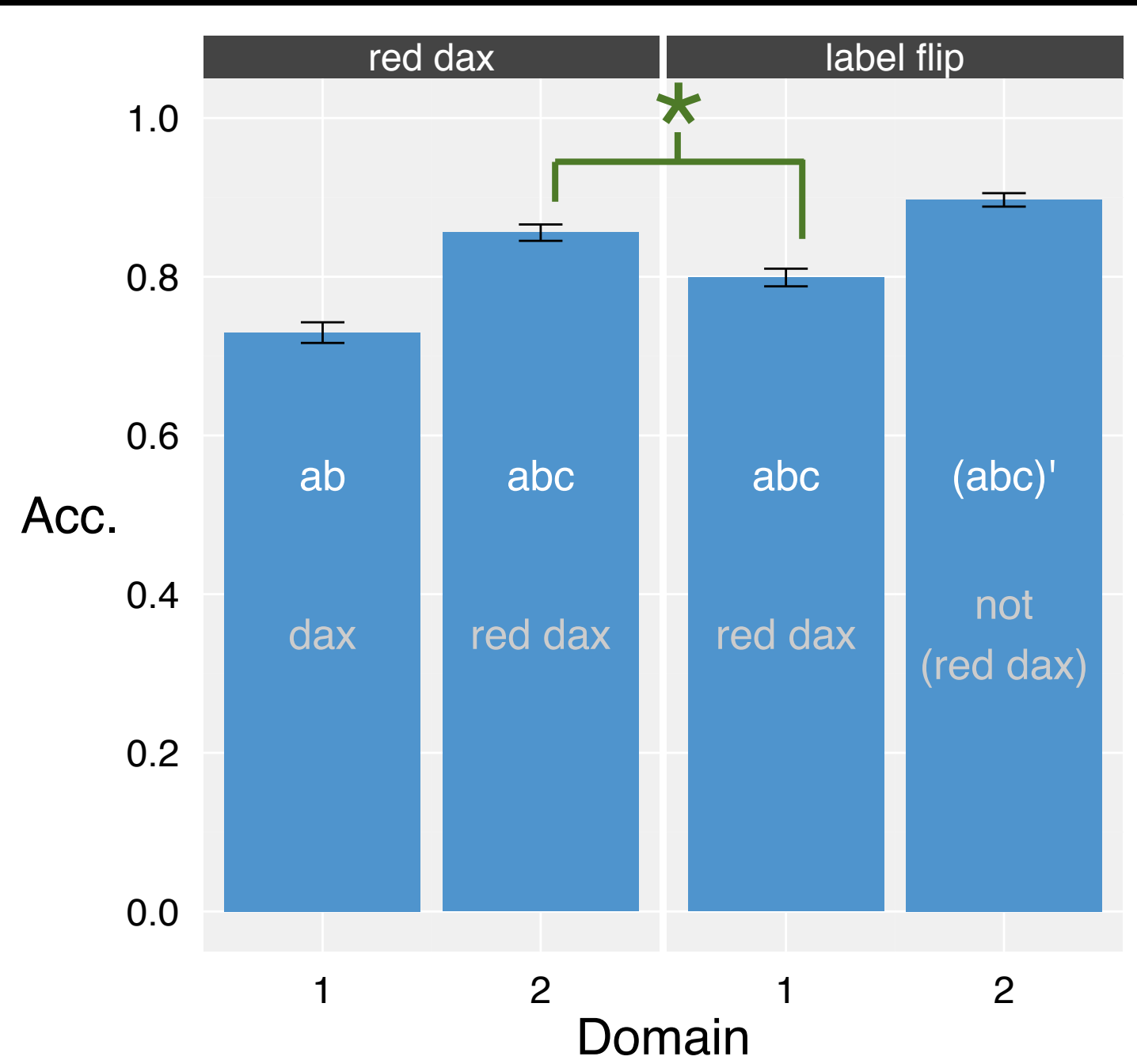
+ T from dax to not (red dax)



Dax results (human)

+ T from dax to red dax

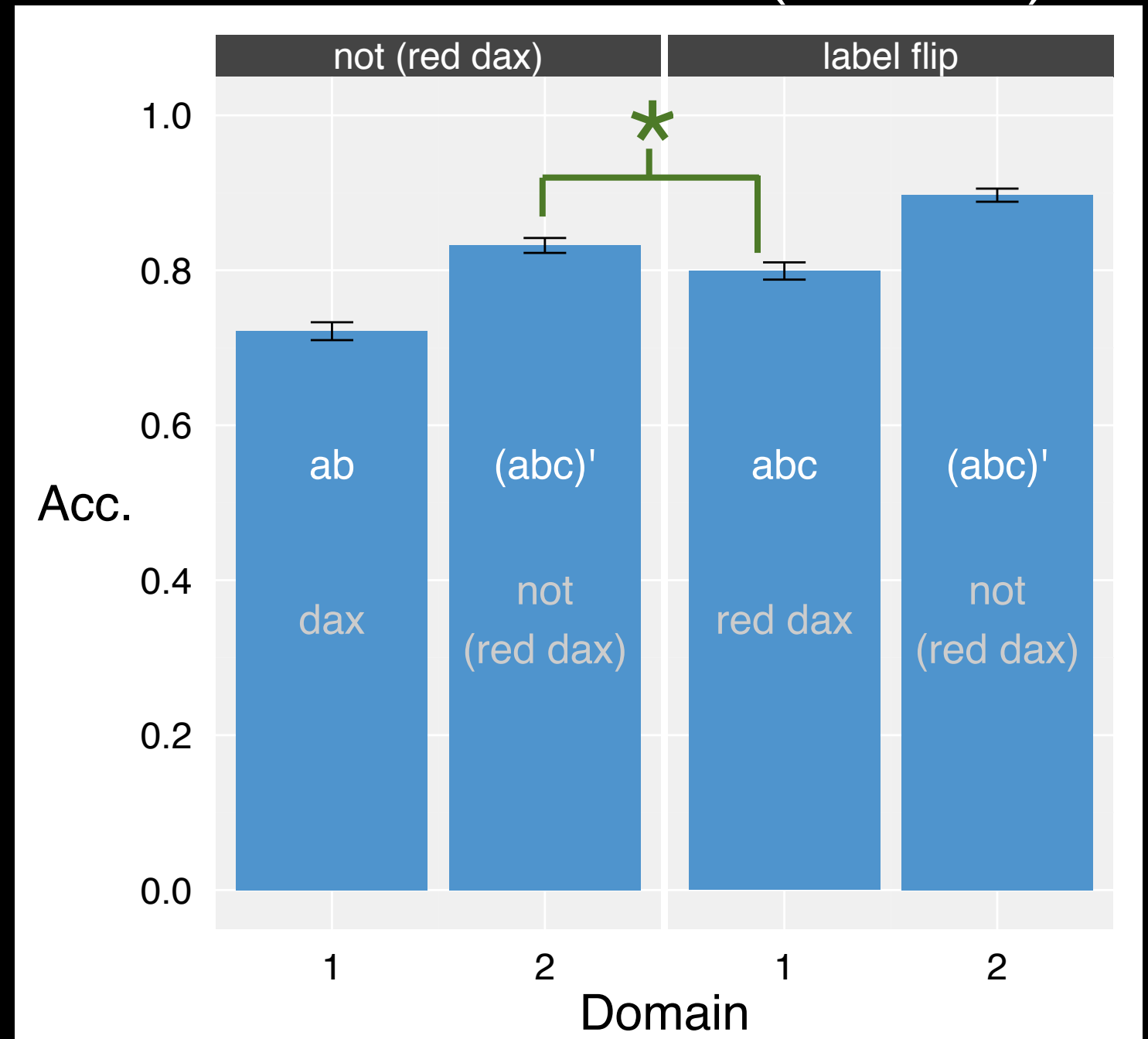
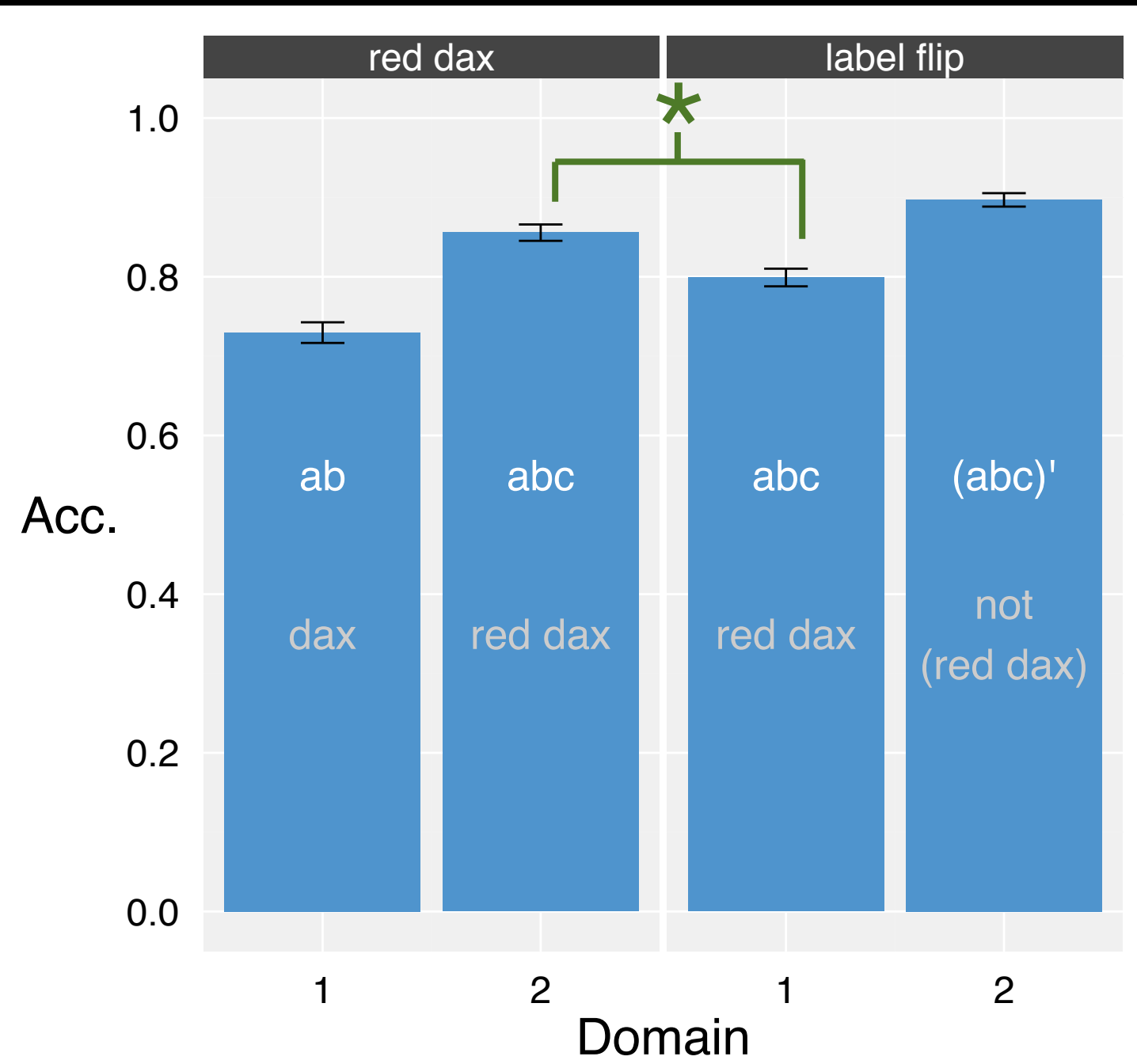
+ T from dax to not (red dax)



Dax results (human)

+ T from dax to red dax

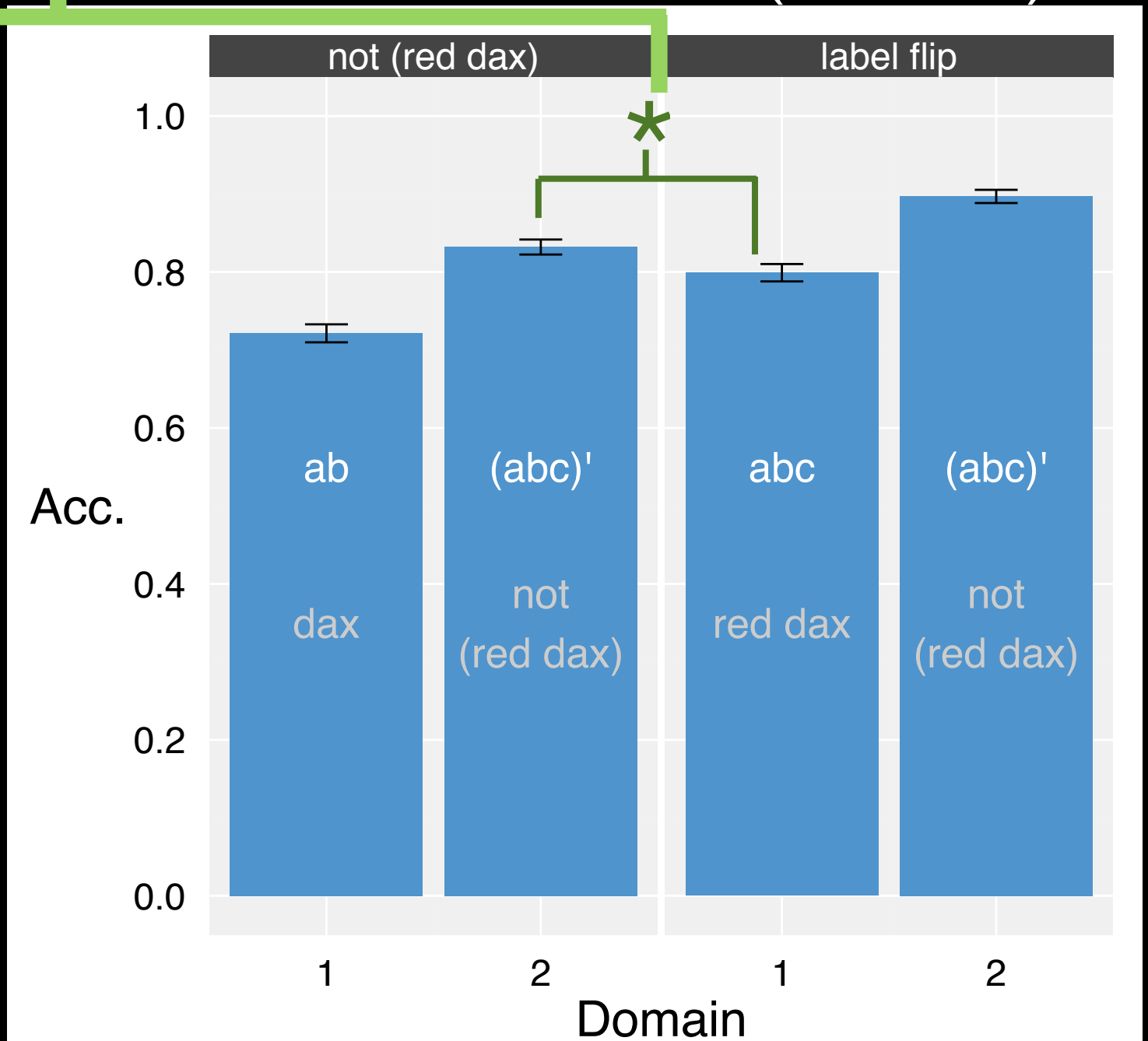
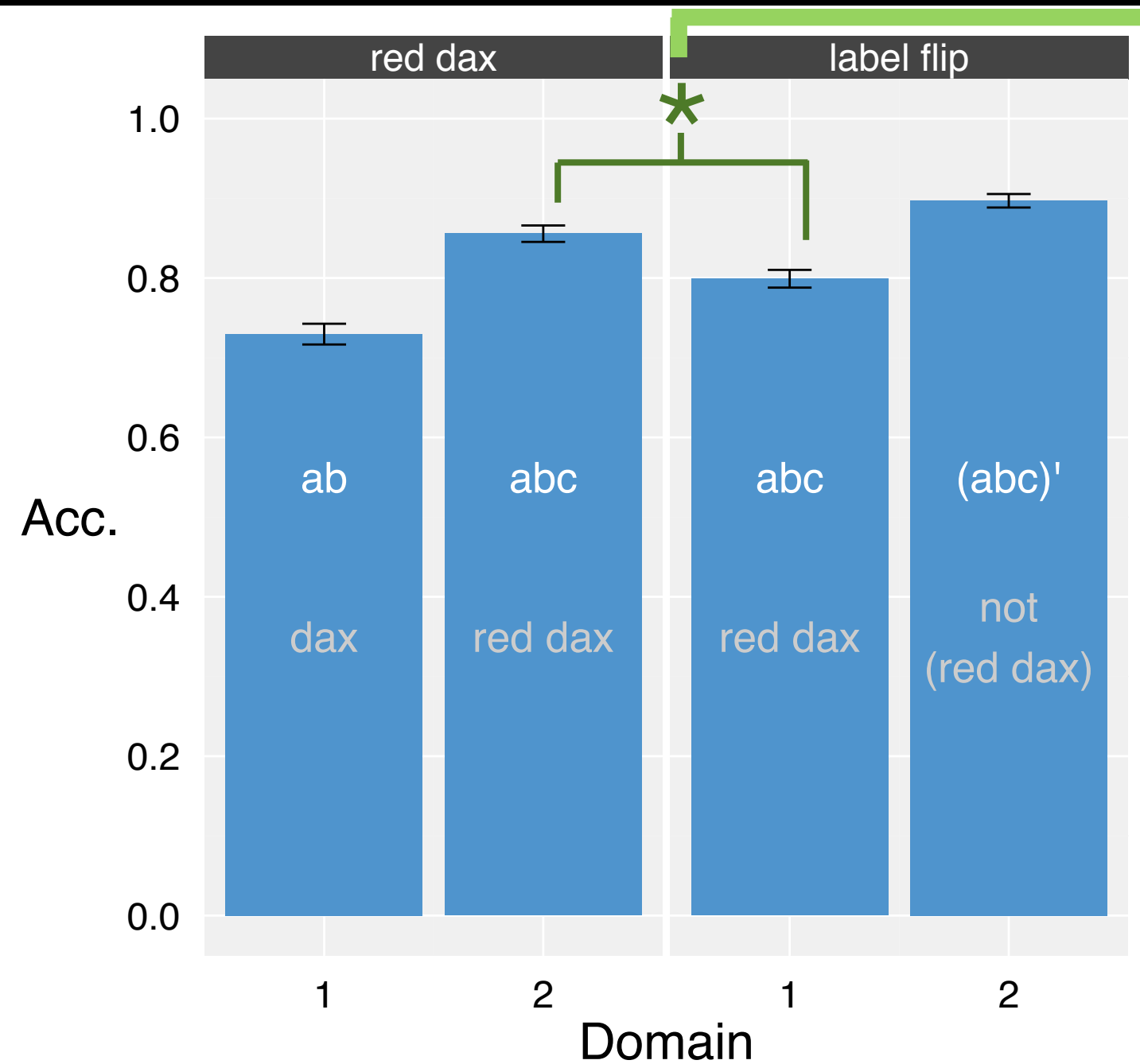
+ T from dax to not (red dax)



Dax results (human)

+ T from dax to red dax

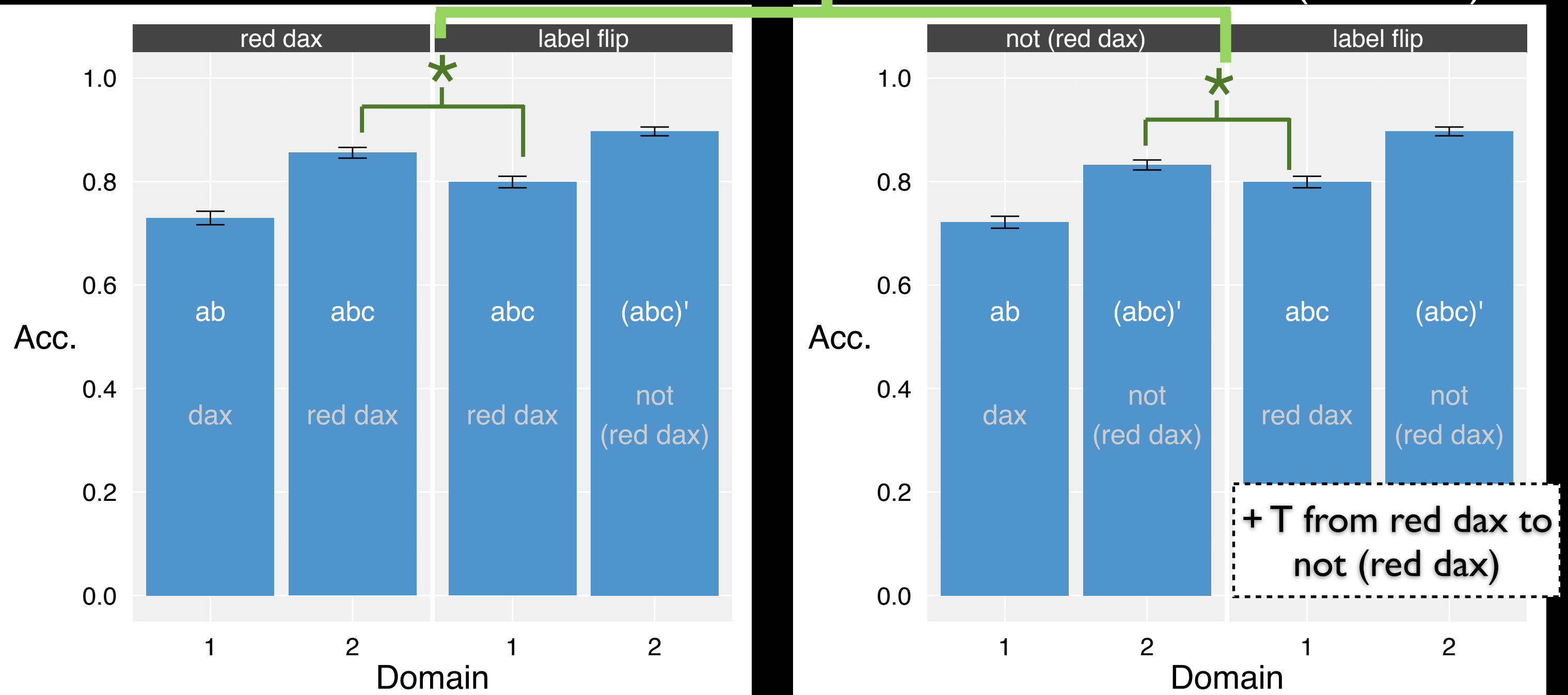
* + T from dax to not (red dax)



Dax results (human)

+ T from dax to red dax

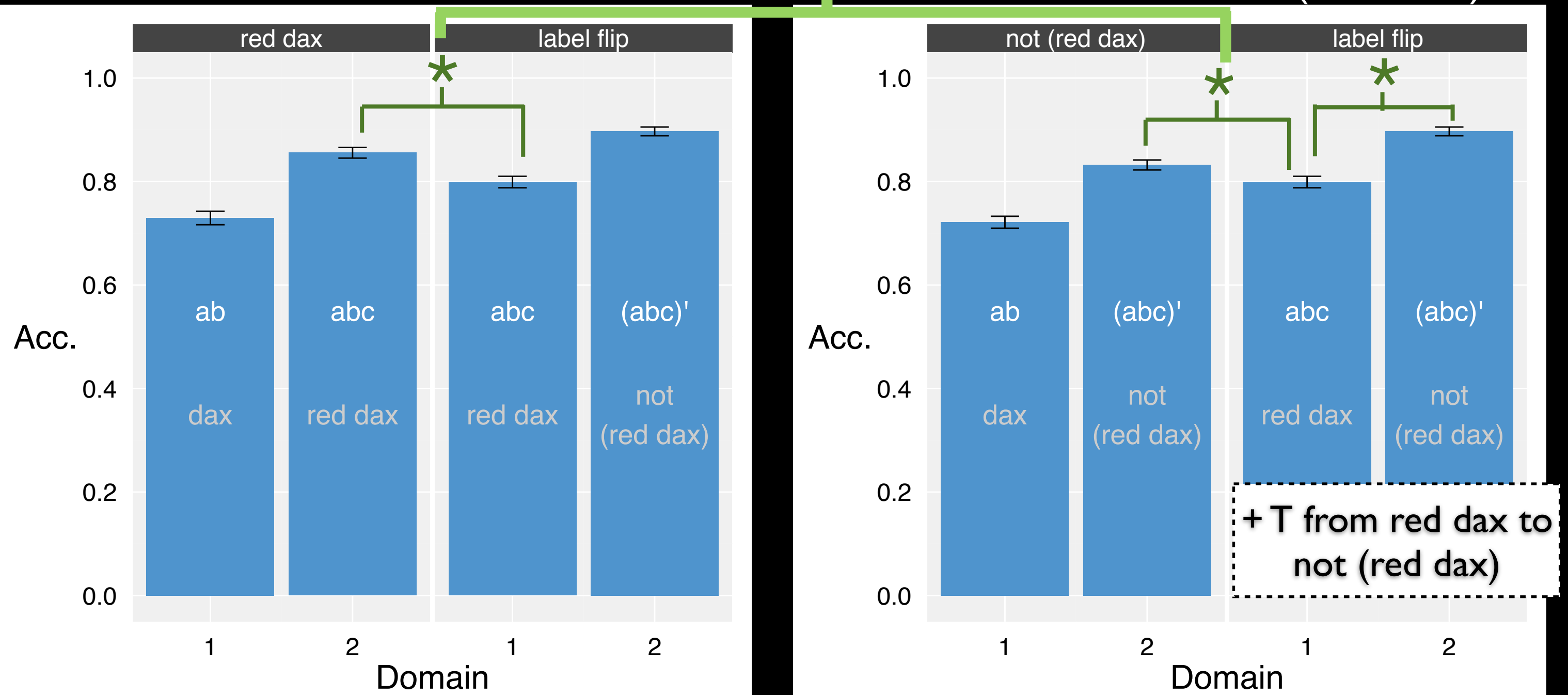
* + T from dax to not (red dax)



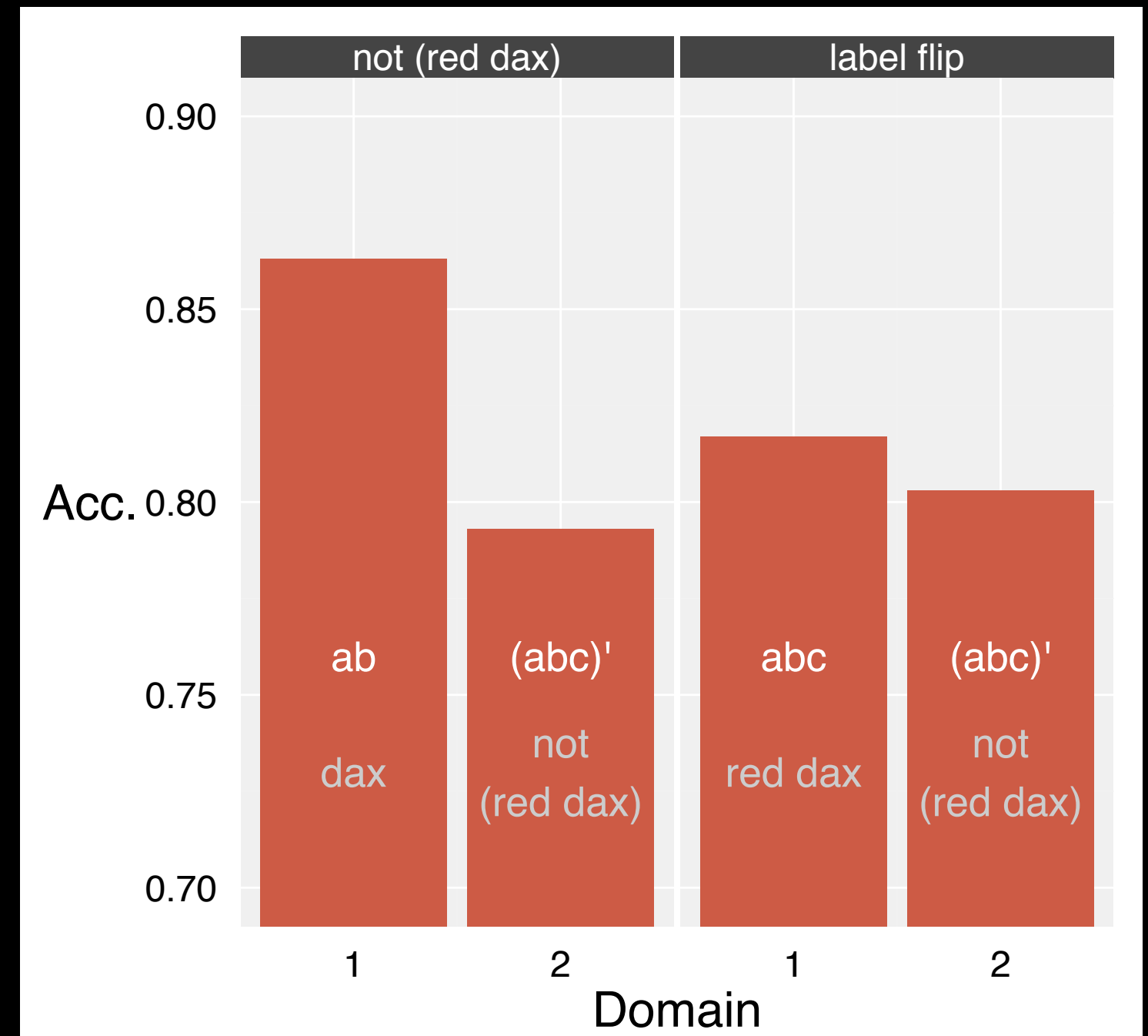
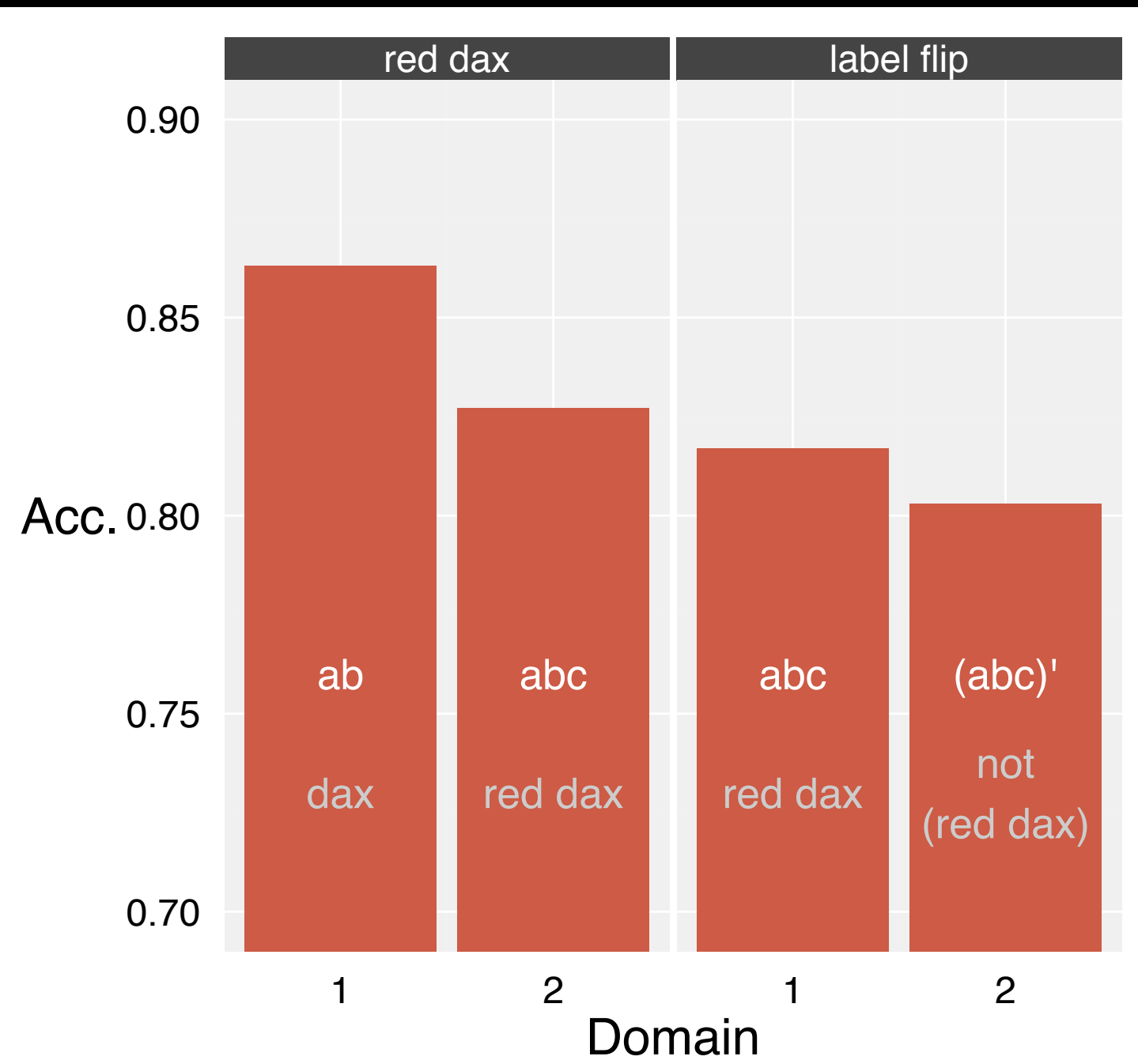
Dax results (human)

+ T from dax to red dax

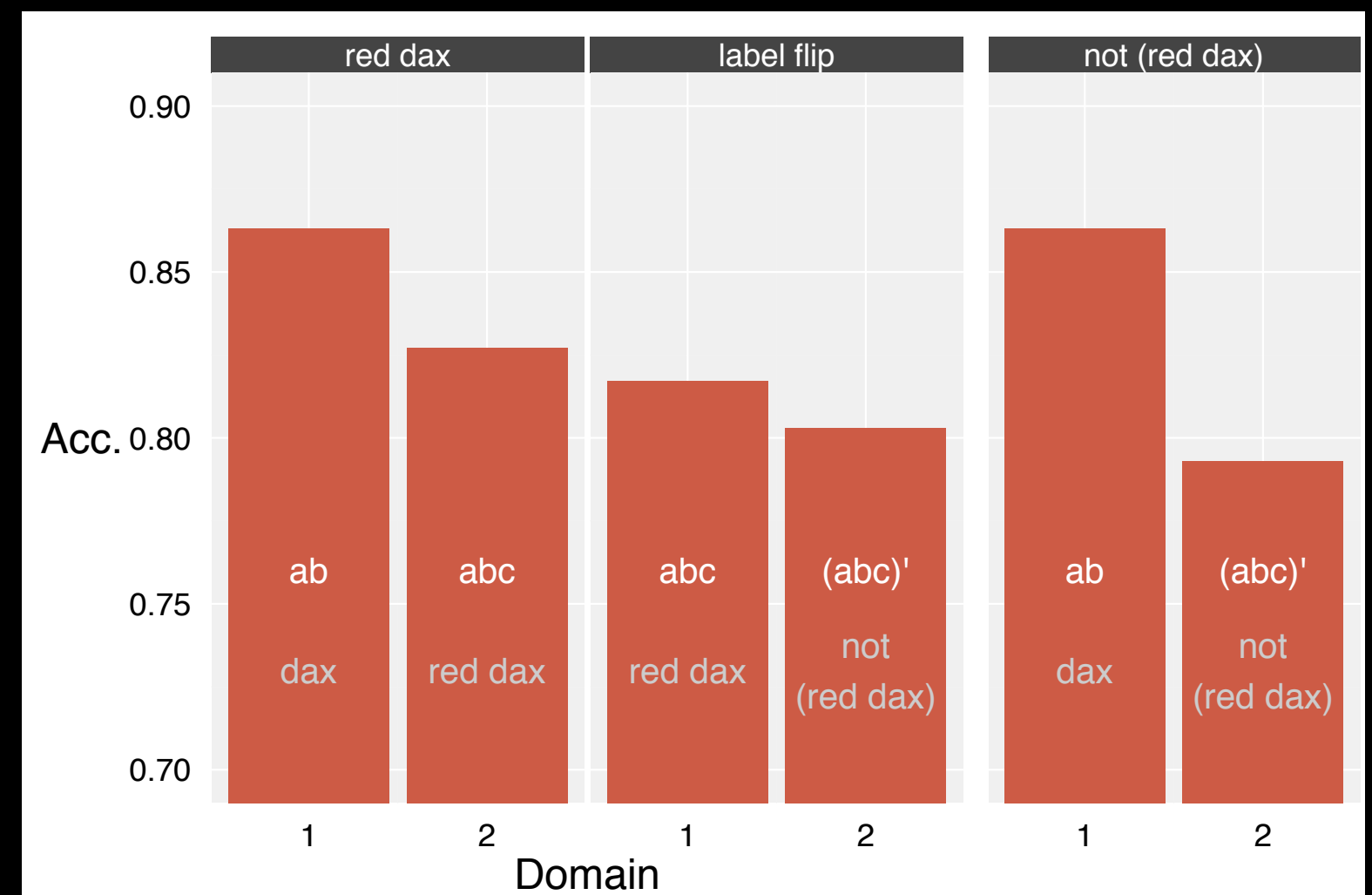
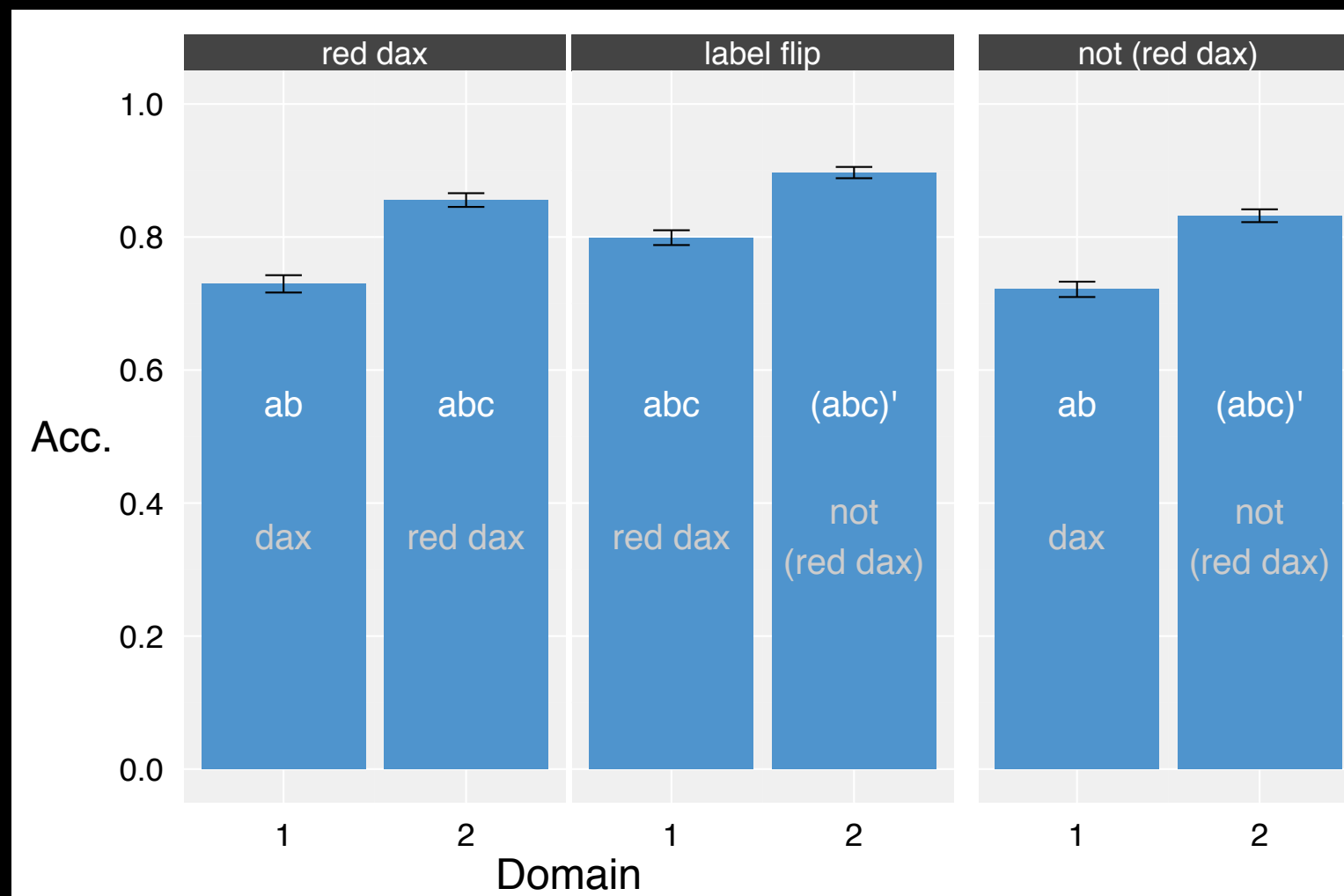
* + T from dax to not (red dax)



Dax results (model)

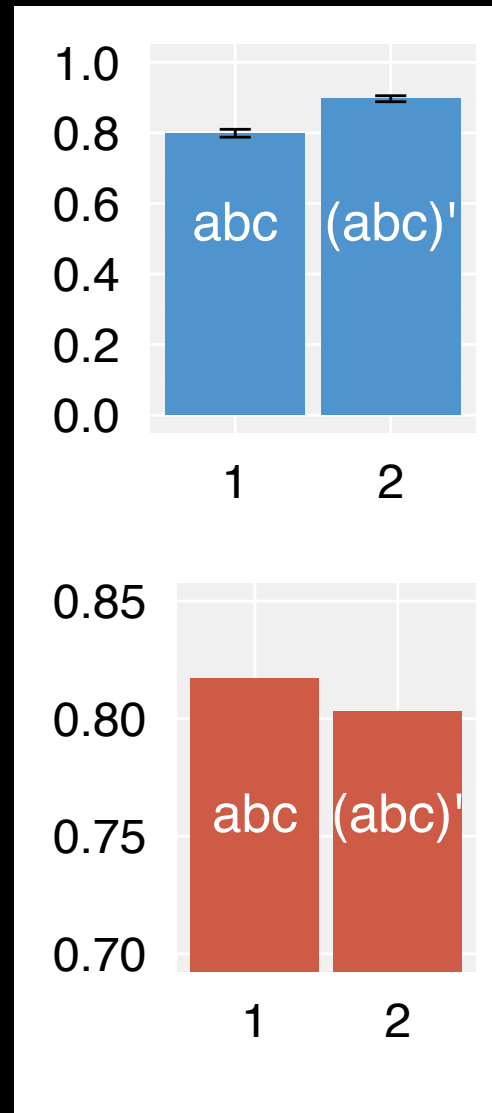


Dax results (comparison)



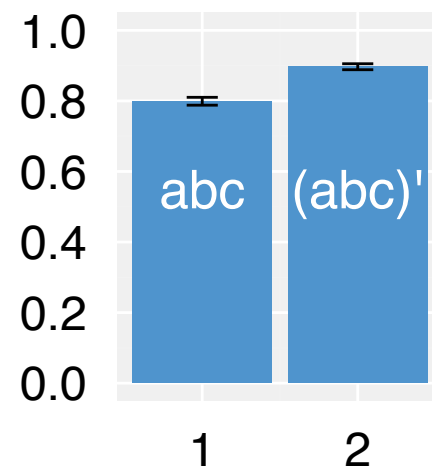
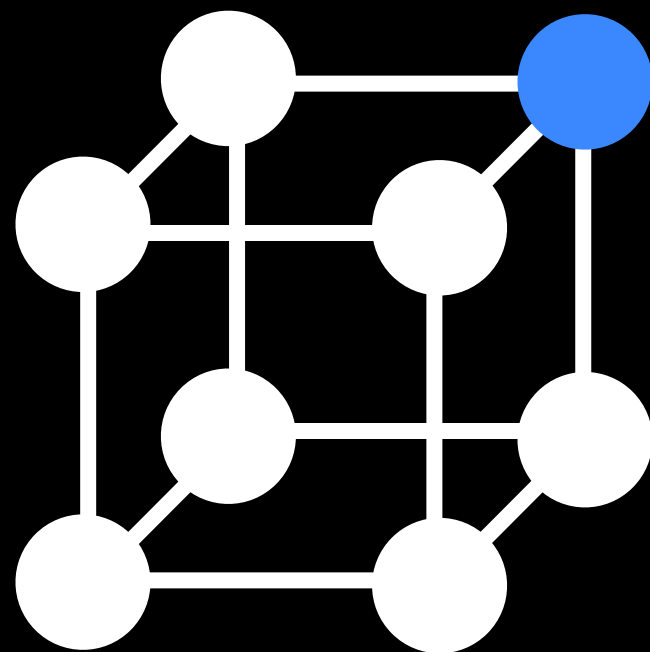
Label flip

Label flip



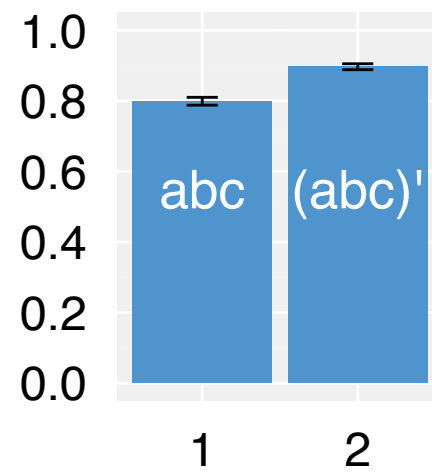
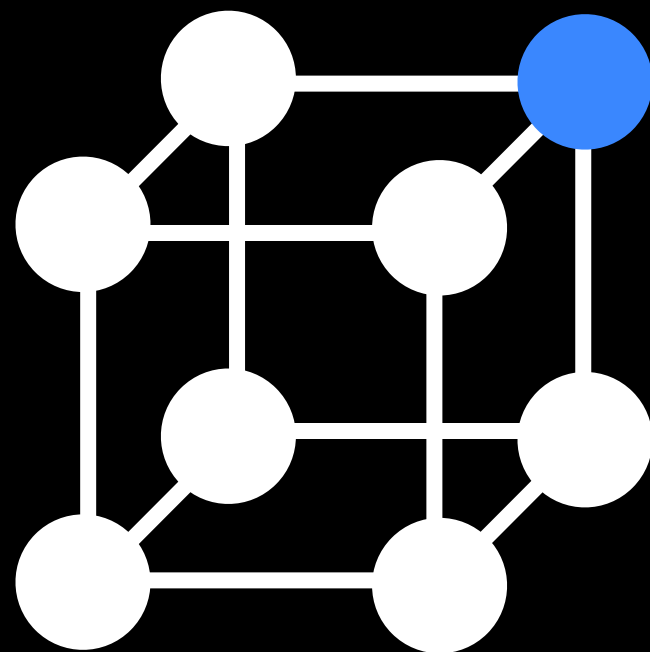
Label flip

Domain I

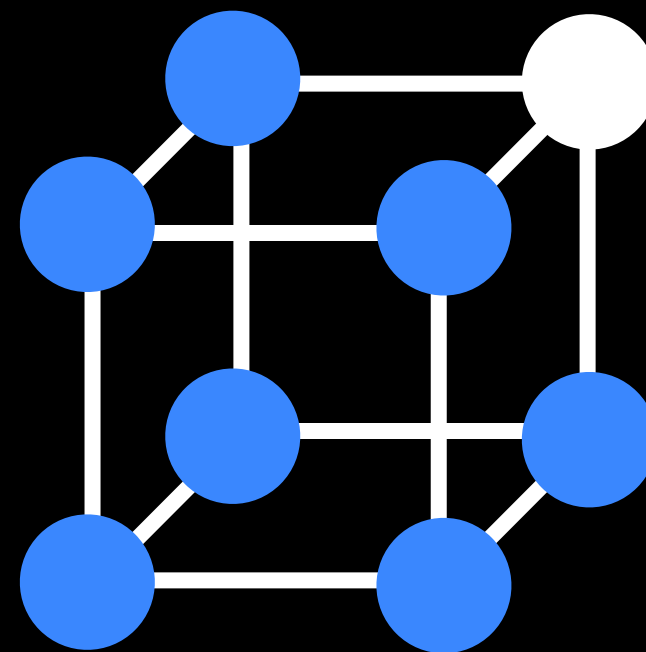


Label flip

Domain 1

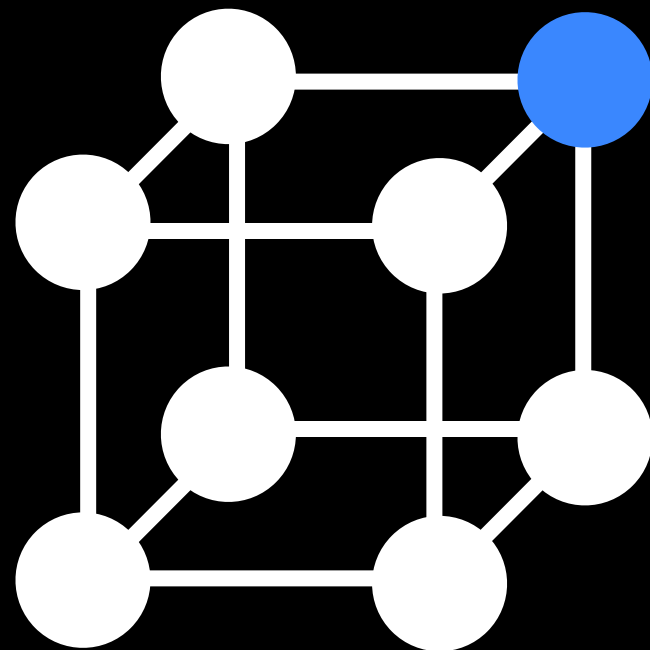


Domain 2

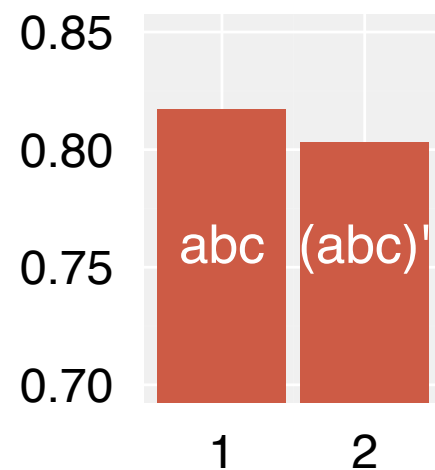
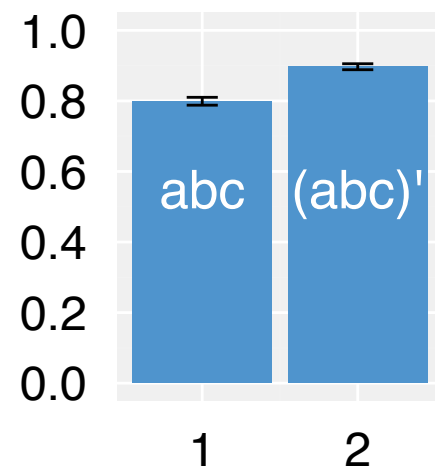


Label flip

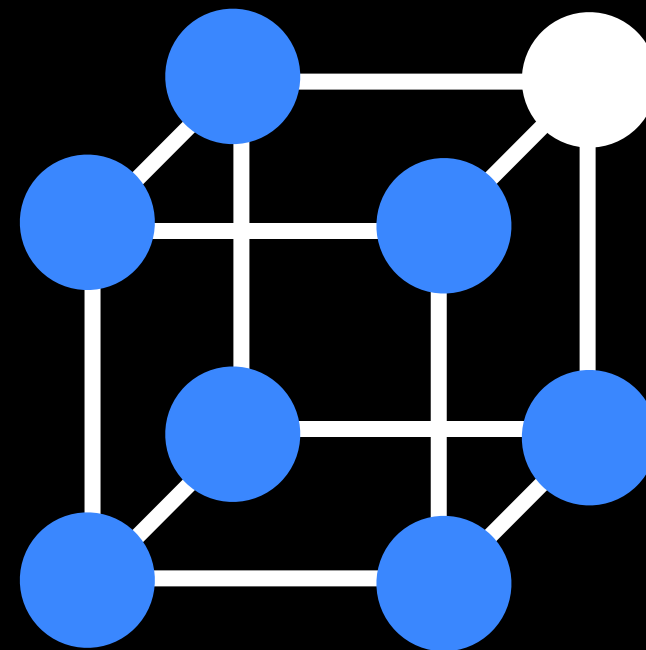
Domain 1



$$q \Leftrightarrow a \wedge b \wedge c$$

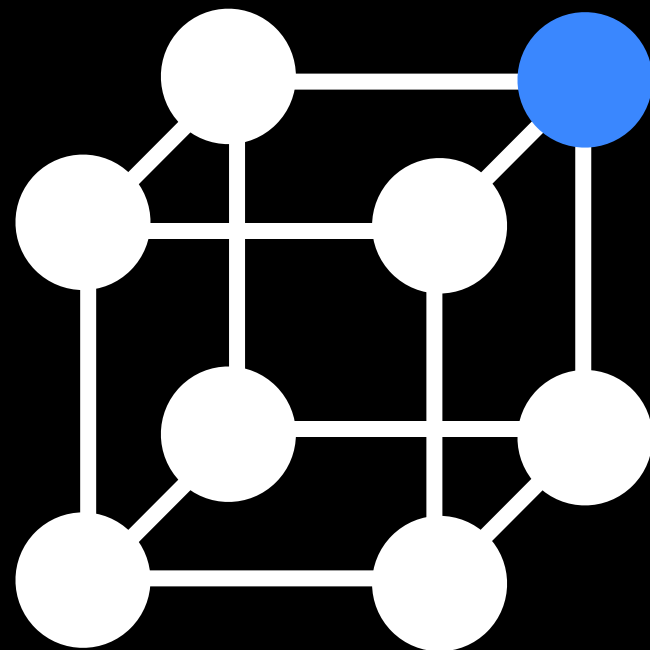


Domain 2

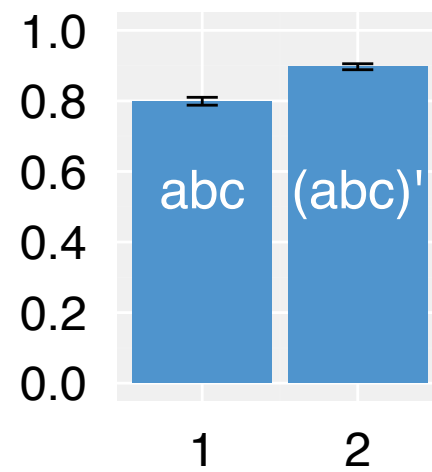


Label flip

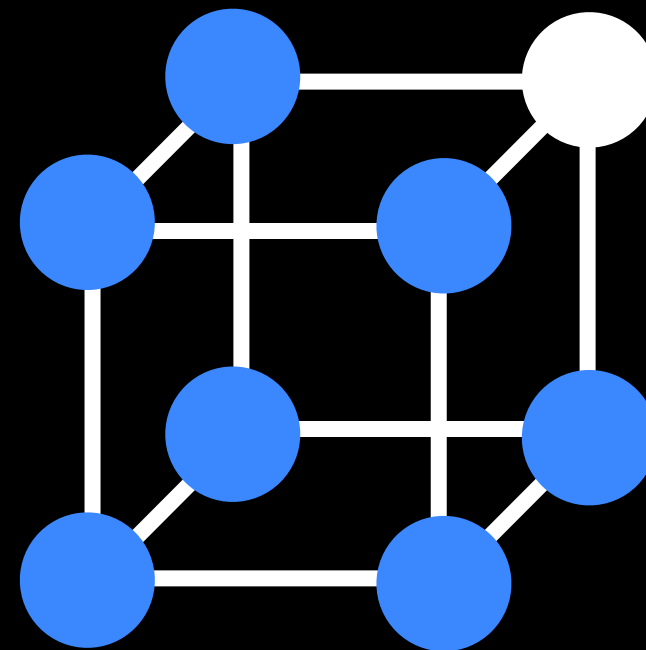
Domain 1



$$q \Leftrightarrow a \wedge b \wedge c$$

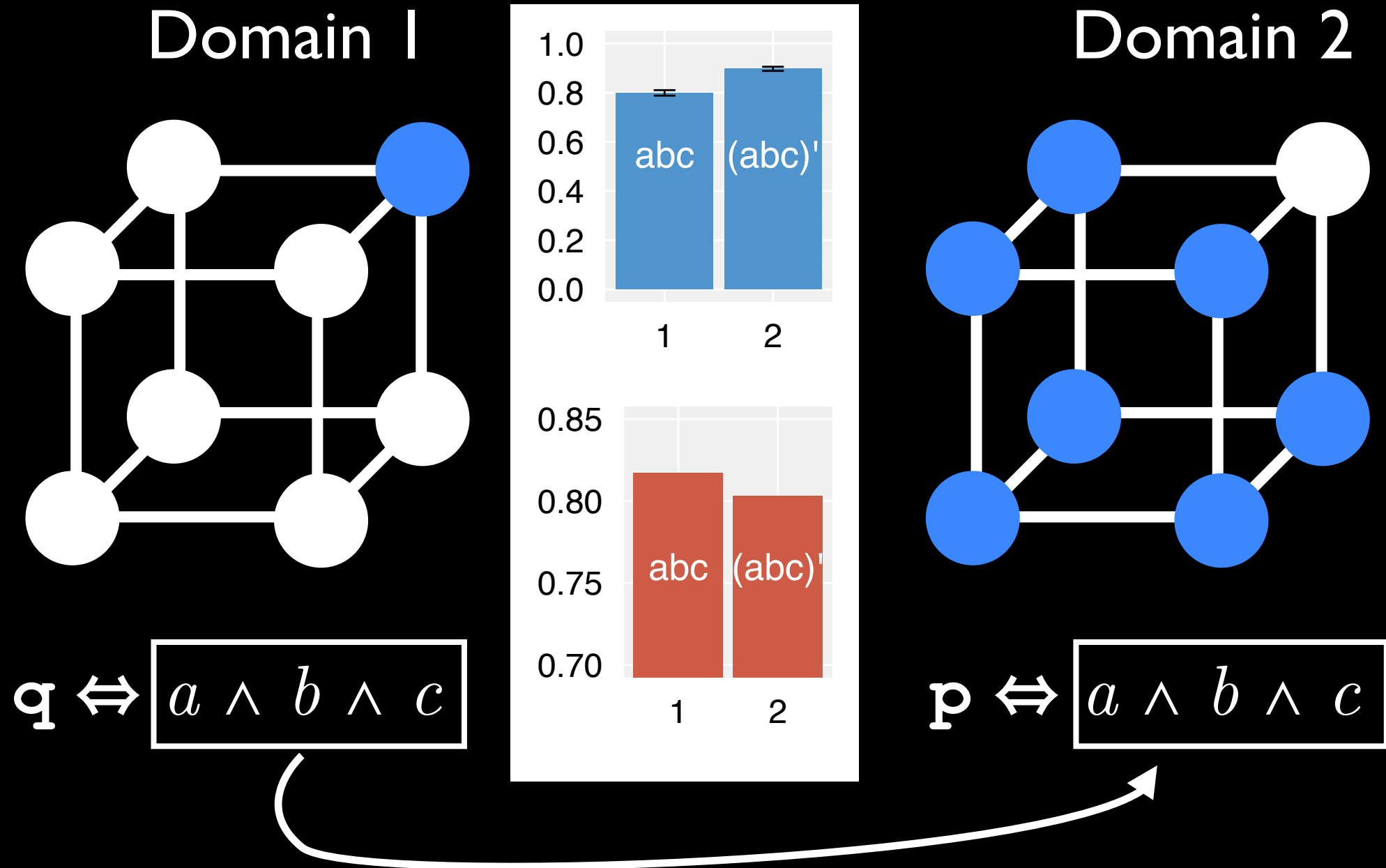


Domain 2



$$p \Leftrightarrow a \wedge b \wedge c$$

Label flip



Rational Rules summary

- ☒ positive self-transfer
- ☒ garden pathing
- ☒ learning effect
- ☐ compositional transfer

Rational Rules (short-circuited)

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

Feature \rightarrow F_a | F_b | F_c

F_a \rightarrow $a = 0$ | $a = 1$

F_b \rightarrow $b = 0$ | $b = 1$

F_c \rightarrow $c = 0$ | $c = 1$

Rational Rules (short-circuited)

At feature level of the PCFG, can short-circuit to previous rule

Concept \rightarrow *Label* \Leftrightarrow *Rule*

Label \rightarrow **p** | **q**

Rule \rightarrow *Sense* | *Sense* \vee *Rule*

Sense \rightarrow *Feature* | *Feature* \wedge *Sense*

Feature \rightarrow *Old* | *New*

NewFeat \rightarrow F_a | F_b | F_c

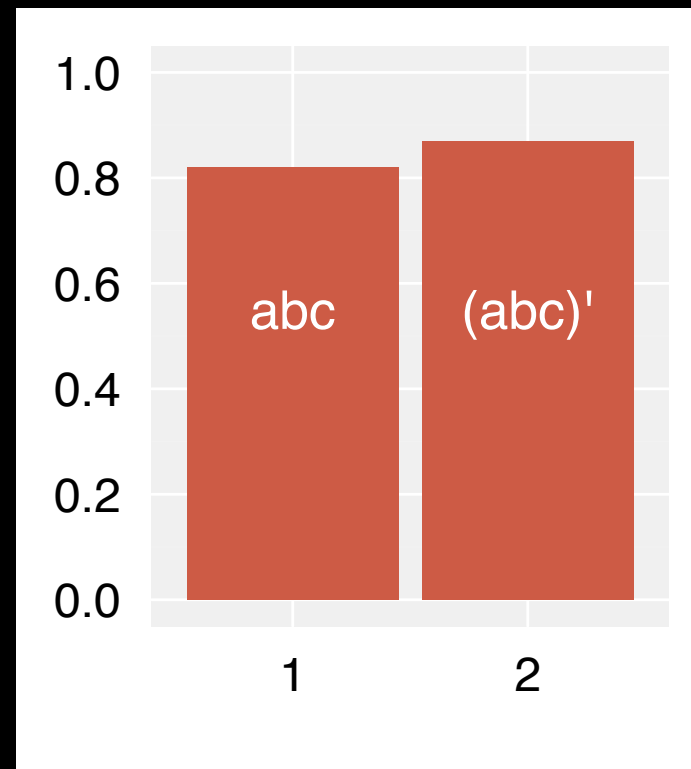
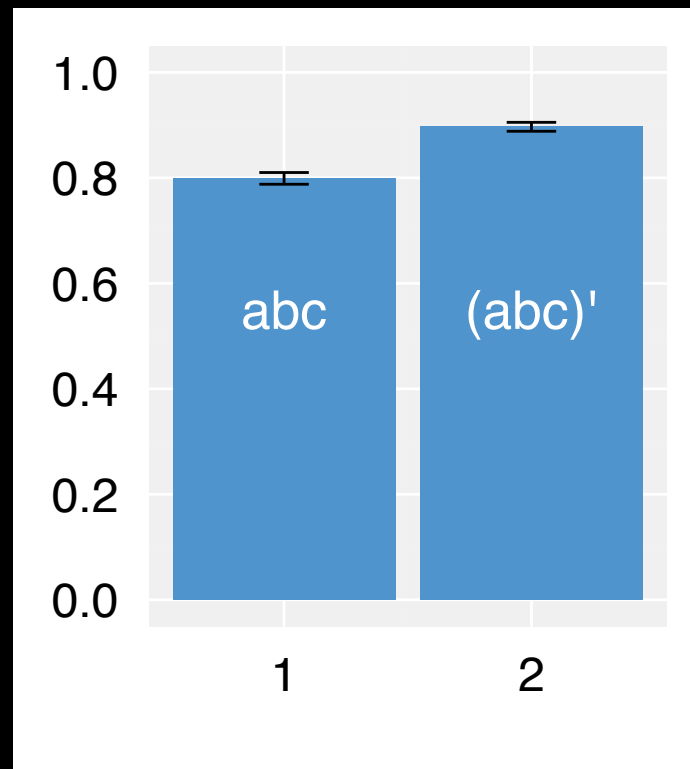
F_a \rightarrow $a = 0$ | $a = 1$

...

Short-circuit results

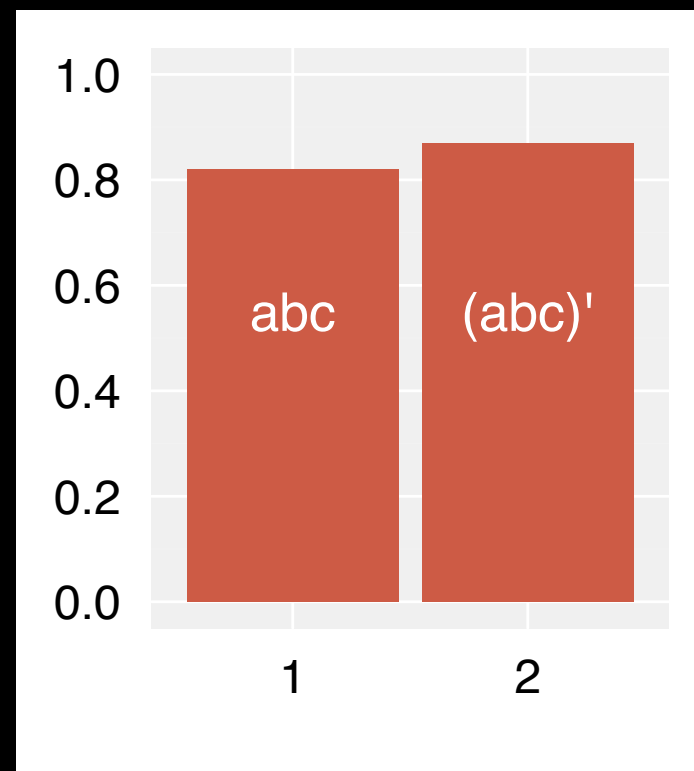
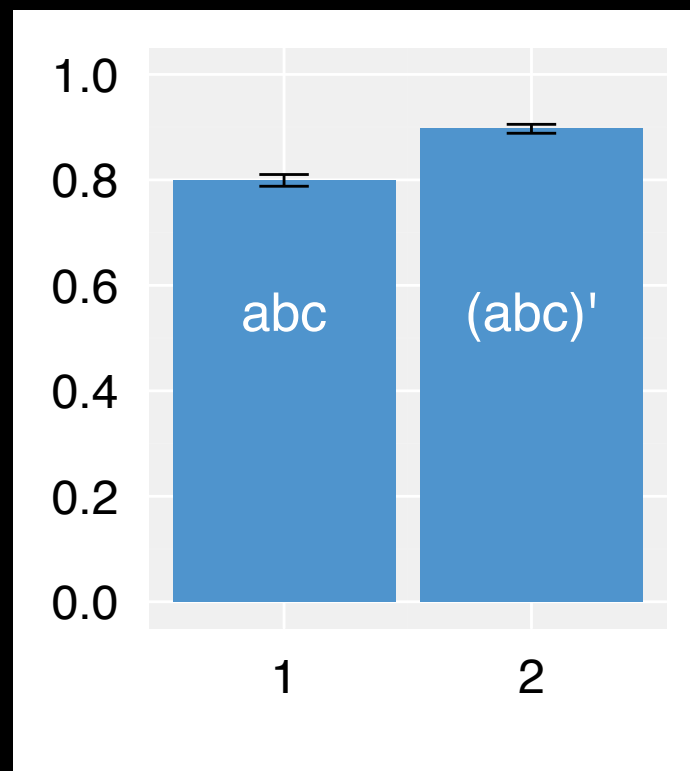
Short-circuit results

Label flip:

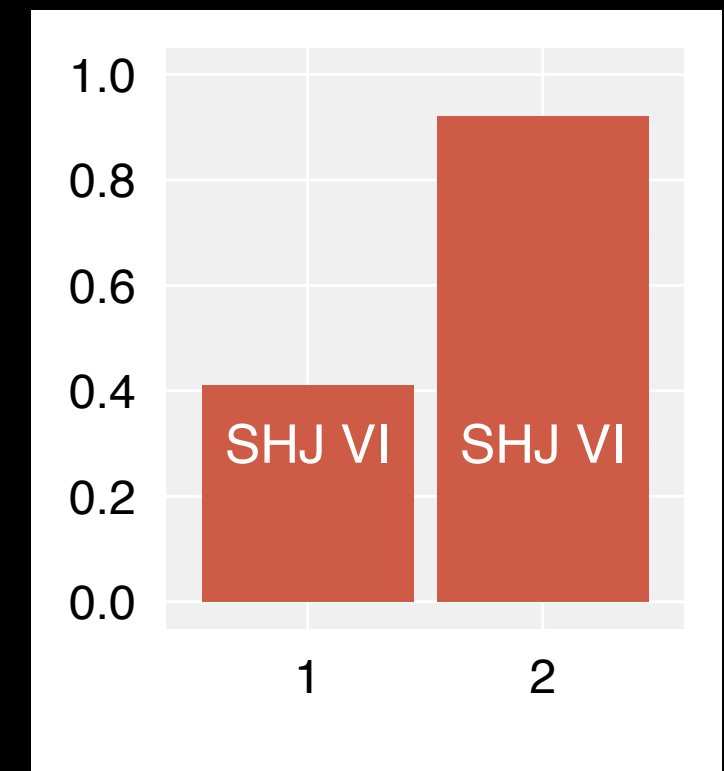
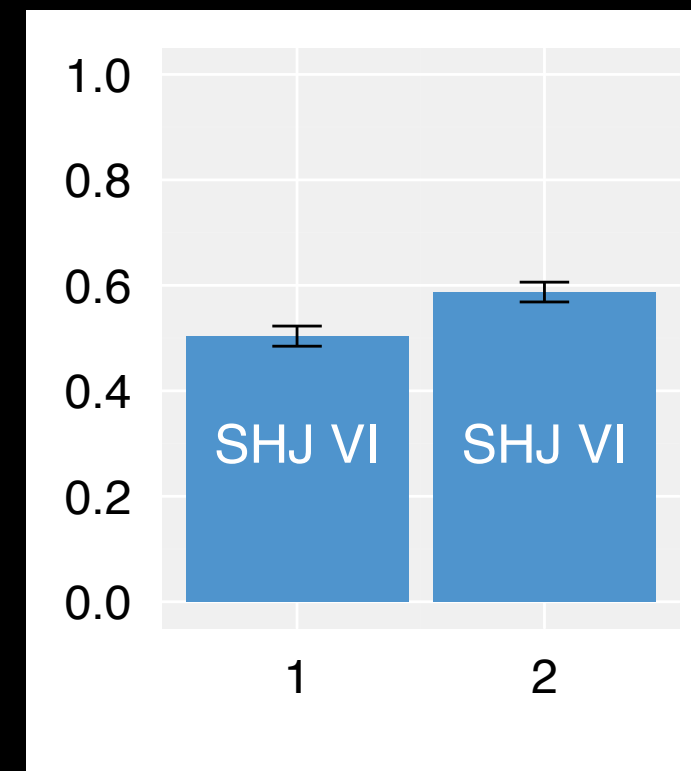


Short-circuit results

Label flip:



SHJ VI:



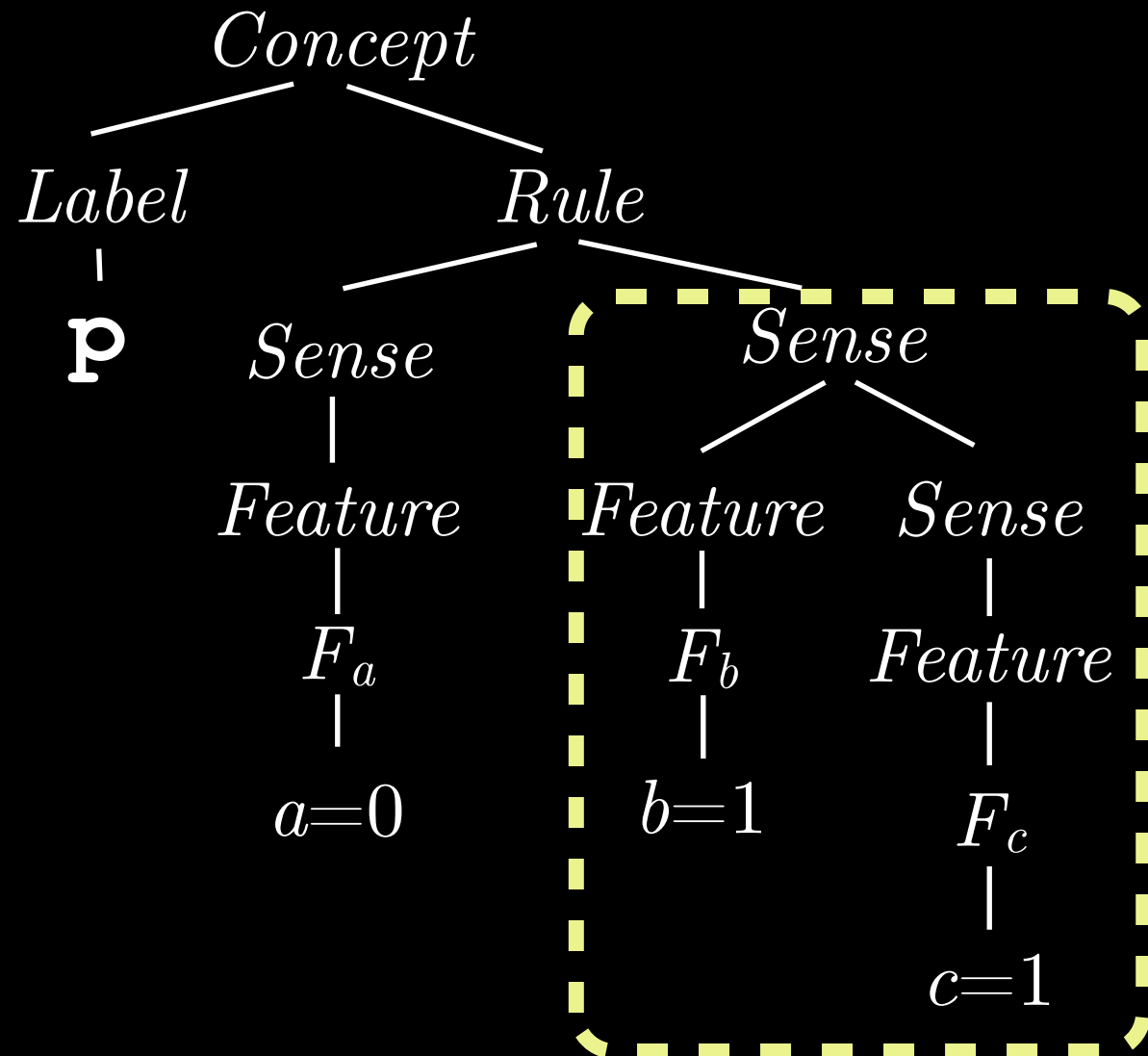
Possible fixes?

Possible fixes?

Adaptor grammar (Johnson,
2007); reuse arbitrary **subtrees**:

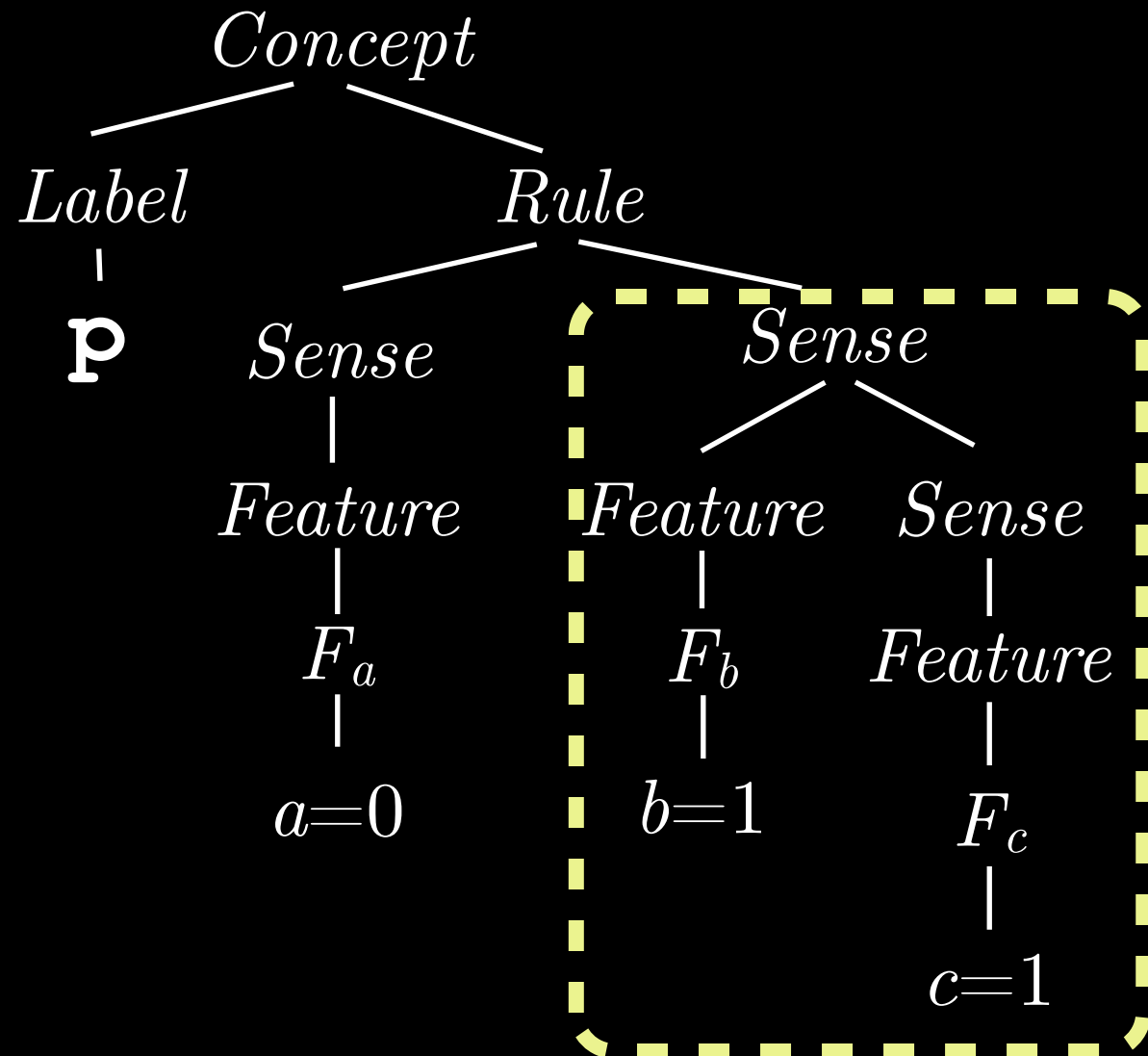
Possible fixes?

Adaptor grammar (Johnson, 2007); reuse arbitrary **subtrees**:



Possible fixes?

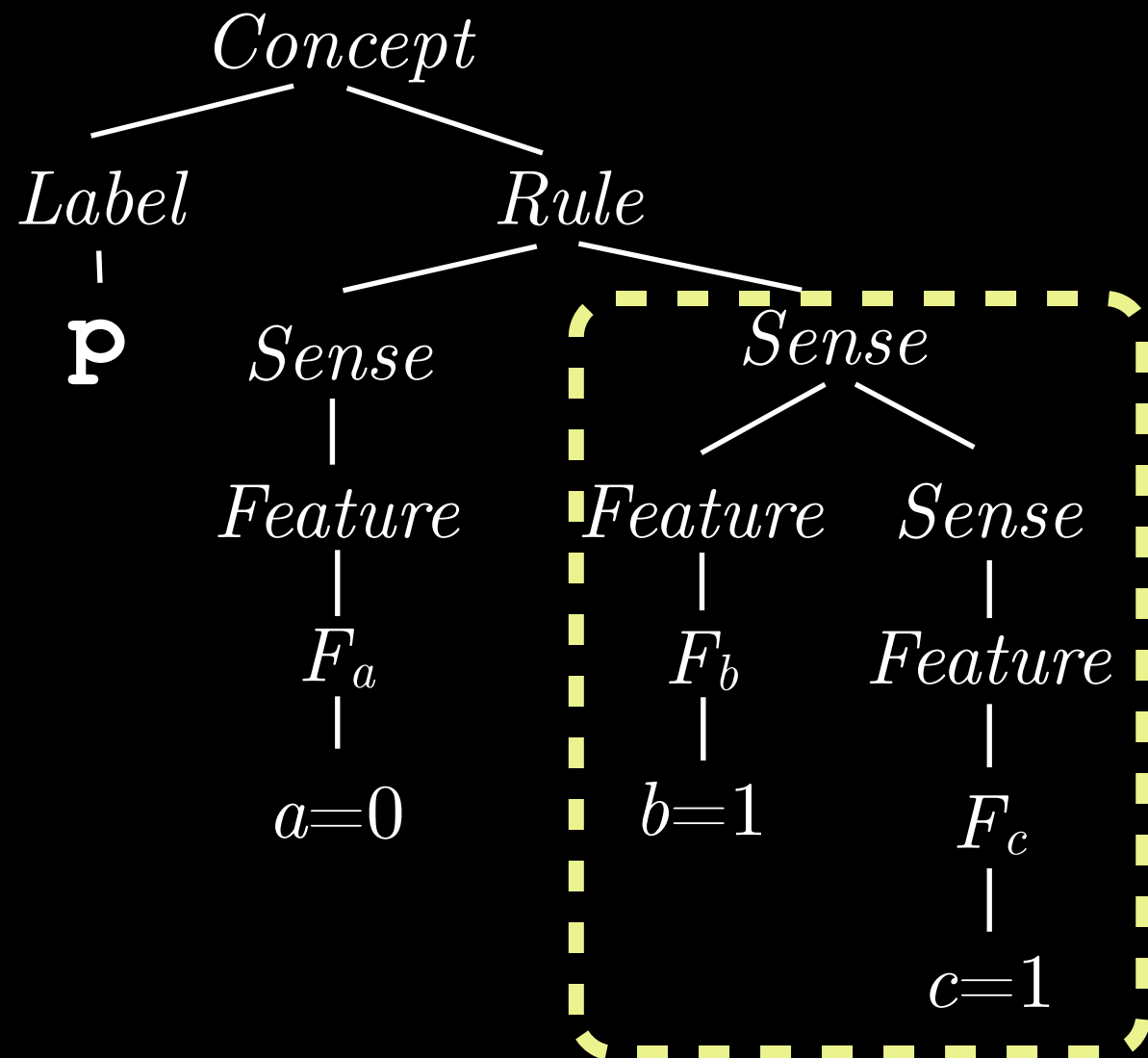
Adaptor grammar (Johnson, 2007); reuse arbitrary **subtrees**:



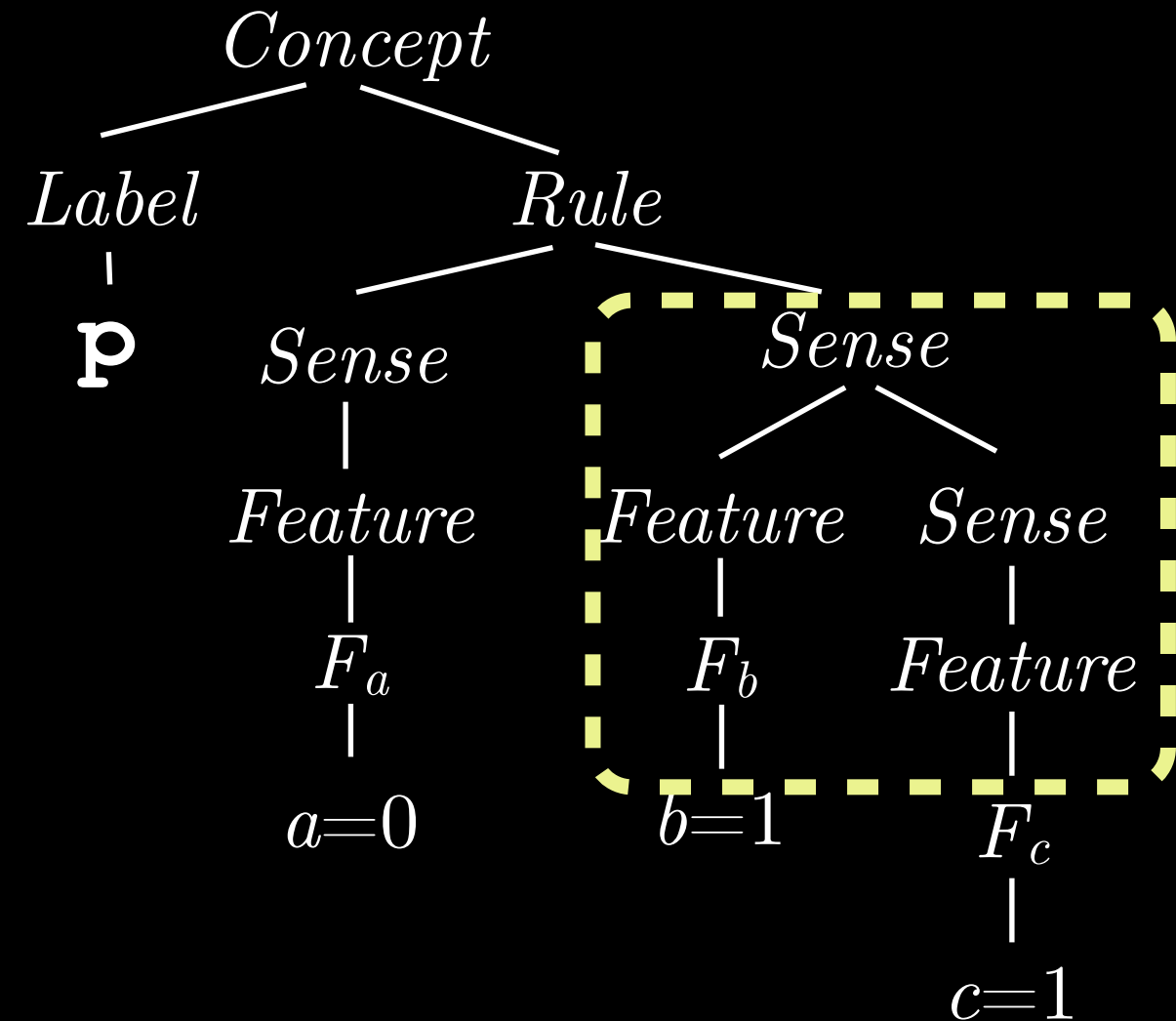
Fragment grammar (O'Donnell, 2011); reuse arbitrary **fragments**:

Possible fixes?

Adaptor grammar (Johnson, 2007); reuse arbitrary **subtrees**:



Fragment grammar (O'Donnell, 2011); reuse arbitrary **fragments**:



So far...

Experimental

Found many transfer effects in Boolean concept learning:

- i. positive self-transfer, garden pathing, learning effect
- ii. compositional (*dax*)

Computational

Theory where concepts and features are like in kind

RR accounts for (i) but not (ii)

RR-SC doesn't work for (ii), maybe RR-AG or RR-FG do

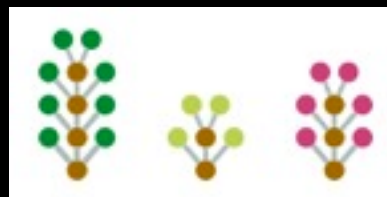
Ahead...

Within-paradigm variations:

- Have some items that re-appear in domain 2
- Different response modalities (e.g., mouse click in domain 2)
- Lexical labels (e.g., *wug*) versus keyboard responses
- More than 2 domains
- Change of dimensionality
- Simultaneous, rather than serial, presentation
- Vary inter-trial interval

Other paradigms:

- Non-Boolean:



Other domains:

- Causal systems
- Learning sequences of motor actions; problem solving (cf. Luchins, 1944)
- Relationship to analogy

Computational:

- Fine-tuning LOT (e.g., adding negation)
- Adaptor grammar / Fragment grammar
- RL / stochastic search approximations
- Metacognitive strategy: mitigate negative transfer (e.g., forgetting previous domains)
- Metrizability of Boolean concept space

Questions?

LOUYANG@STANFORD.EDU