Towards a unifying computational account of reference production and comprehension

Judith Degen

Feb 24, 2021

LingLangLunch, Brown University

Stanford University

ALPS interActive Language Processing Lab Stanford

Reference comprehension and production

Comprehension:

How do listeners predict the referent of (possibly ambiguous) referring expressions in real time?

Production:

How do speakers decide which features of an object to include in a referring expression?

Quantity-1: Make your contribution as informative as required.

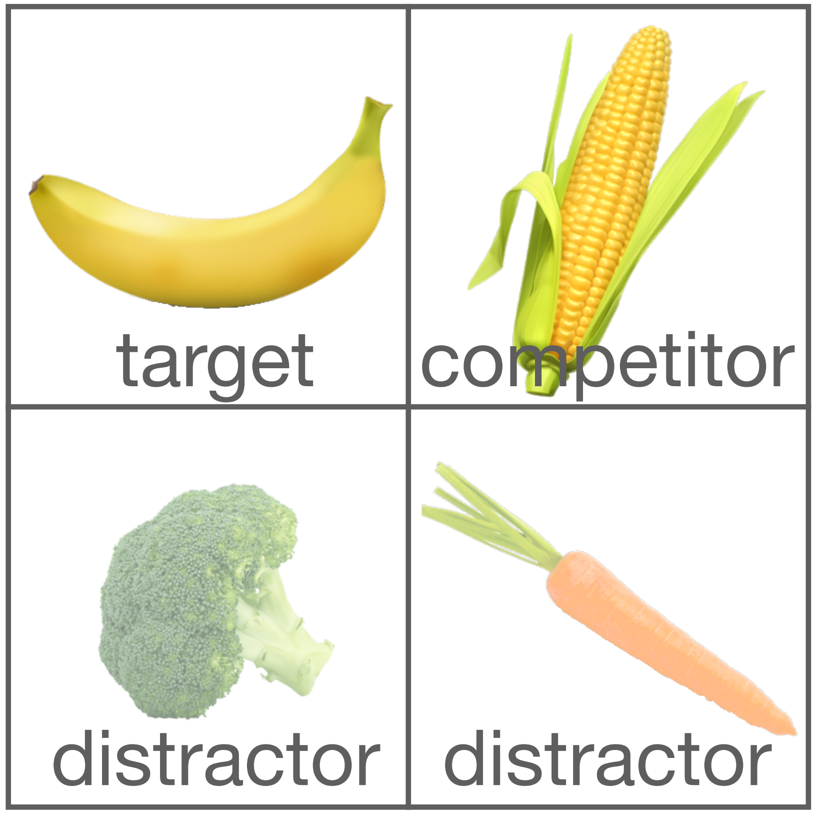
Quantity-2: Don't make your contribution more informative than necessary.

Manner: Be brief and orderly; avoid ambiguity and obscurity.

Grice 1975

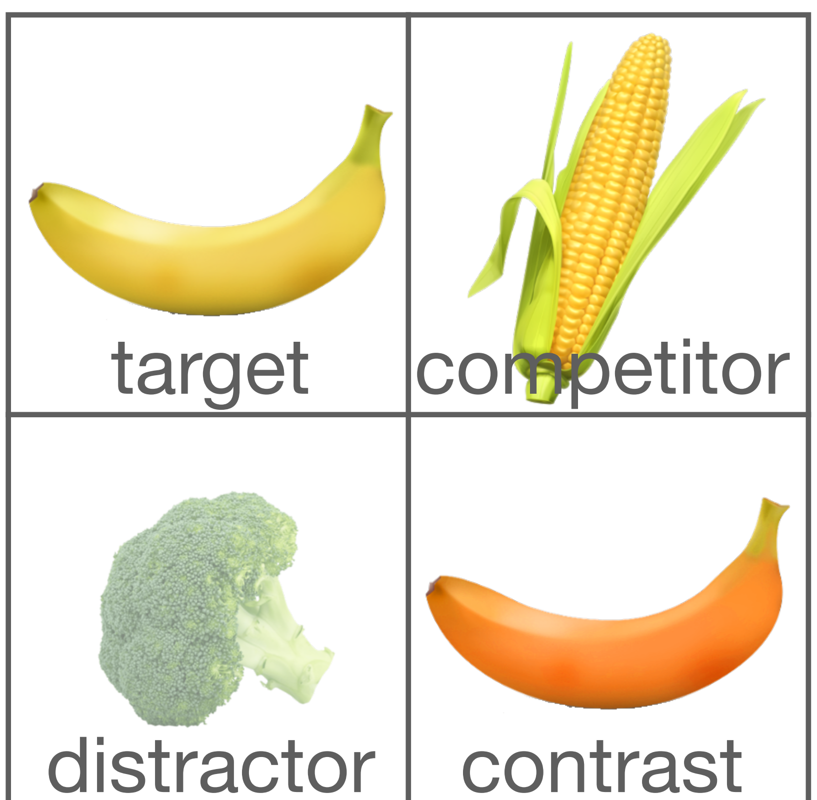
Comprehension: contrastive inferences

"Click on the yellow…"



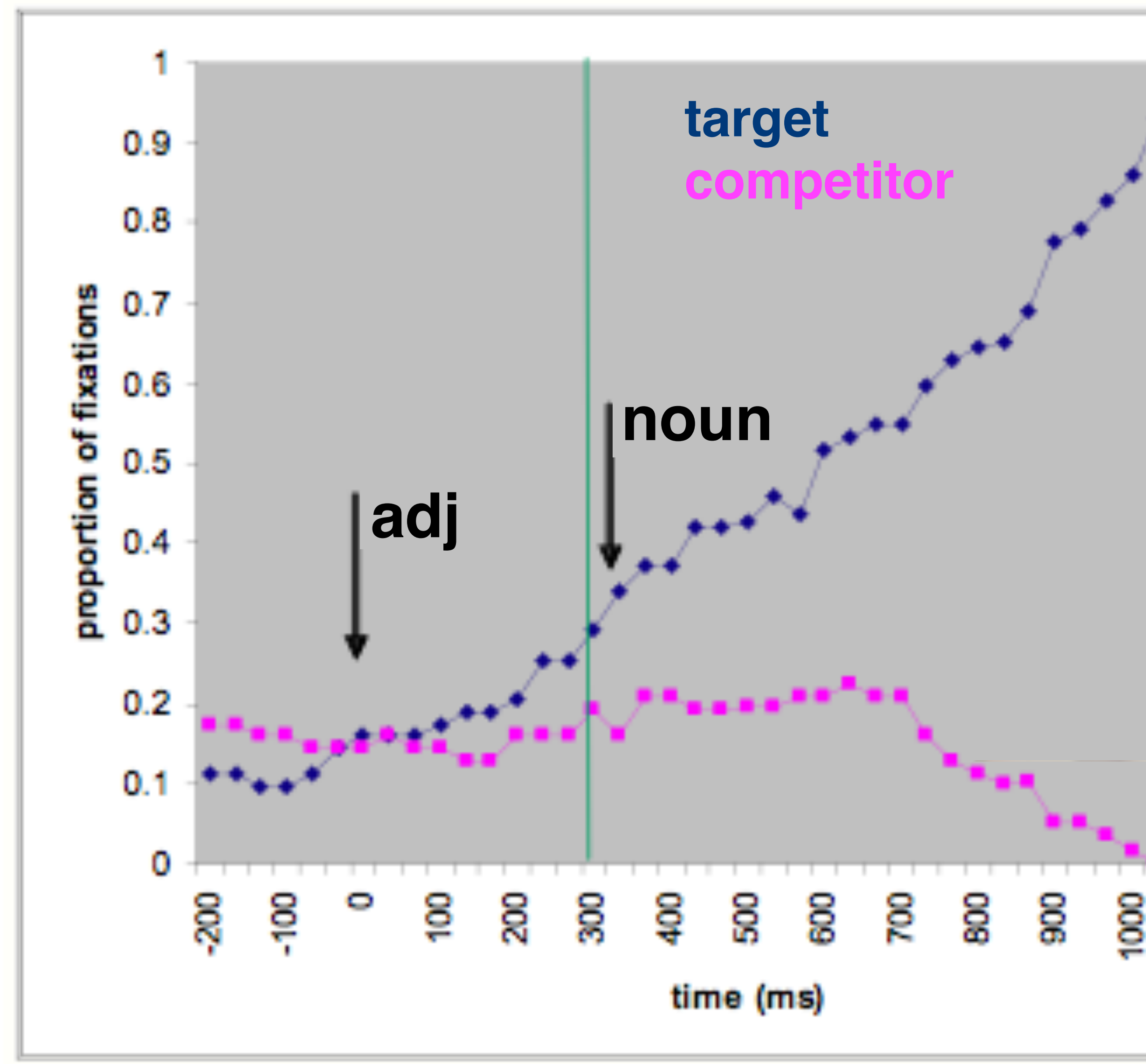
no contrast

Listeners expect speakers not to be over-informative.



contrast

Quantity-2



Sedivy et al 1999; Sedivy 2003; Grodner & Sedivy 2011; Heller et al 2008; Ryskin et al 2019; Rubio-Fernández & Jara-Ettinger 2018; Aparicio et al 2018; Alsop et al 2018

Production: redundant referring expressions

size sufficient

color sufficient

the small pin **Quantity-2** *the blue pin*

75-80% ***the small blue pin*** **8-10%**

Speakers produce seemingly overinformative referring expressions.

Deutsch 1976; Pechmann 1989; Sedivy 2003; Gatt et al. 2011; Koolen et al 2013; Rubio-Fernández 2016; Westerbeek et al 2015; Davies & Katsos 2013; van Gompel et al 2019

Reference comprehension and production

Comprehension:

Listeners draw inferences based on the expectation that speakers not be over-informative.

Quantity-2

Production:

Speakers produce seemingly overinformative referring expressions.

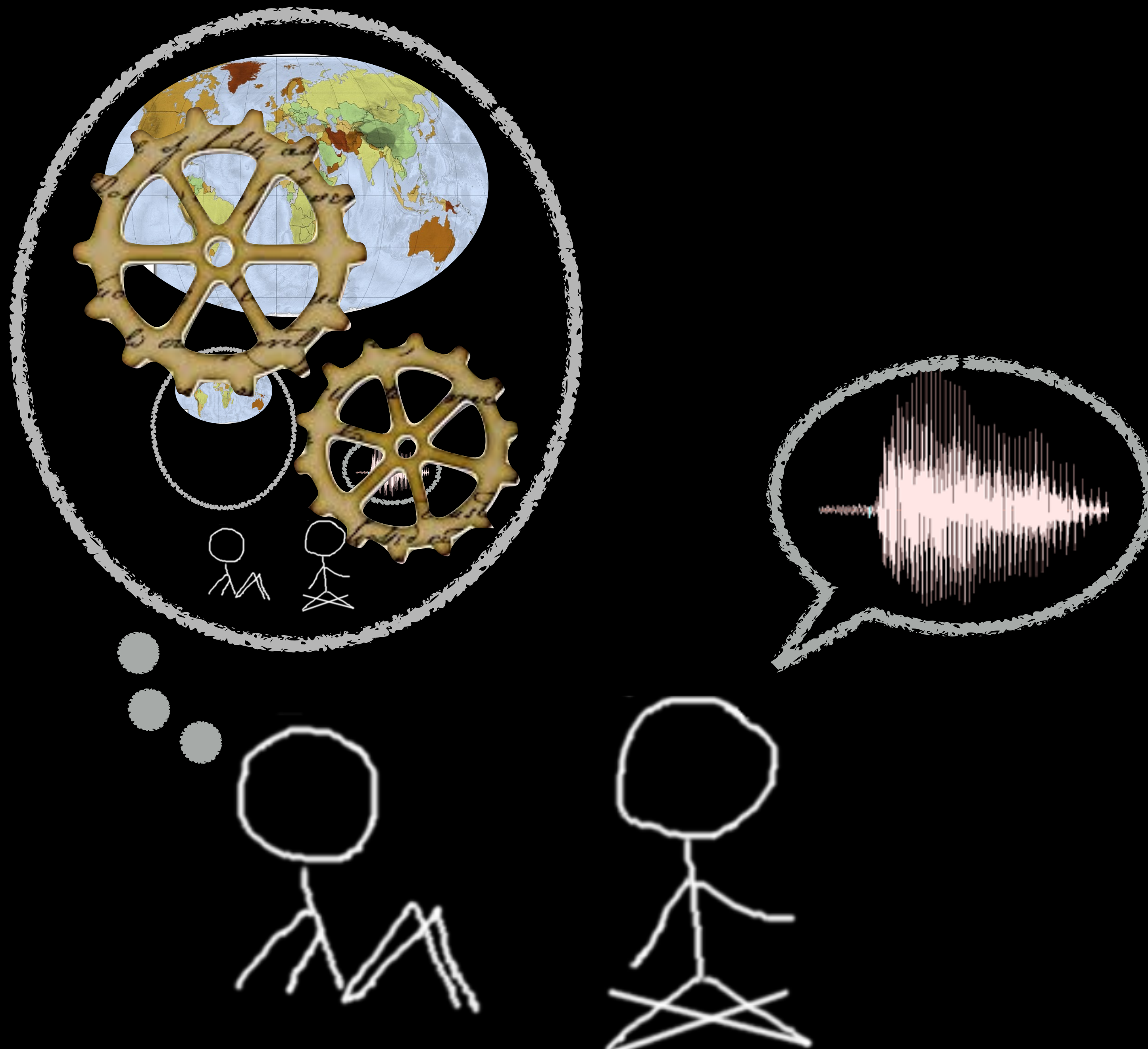
~~**Quantity-2**~~

HOW TO RESOLVE?

world
knowledge

reasoning

context



linguistic
signal

PRAGMATICS

Probabilistic pragmatics

Franke & Jäger, 2016; Goodman & Frank, 2016; Scontras, Tessler, & Franke 2018

Reference

Frank & Goodman, 2012; Qing & Franke, 2015; Degen & Franke, 2012; Stiller et al., 2015; Franke & Degen, 2015; Degen et al, 2020

Cost-based Quantity implicatures

Degen et al., 2013; Rohde et al., 2012

Scalar implicatures

Goodman & Stuhlmüller, 2013; Degen et al., 2015

Embedded implicatures

Potts et al., 2016; Bergen et al., 2016

M-implicatures

Bergen et al., 2012

Figurative meaning

Kao et al., 2013; 2014; 2015; Cohn-Gordon & Bergen, under review

Gradable adjectives

Lassiter & Goodman, 2013; 2015; Qing & Franke, 2014

Adjective ordering

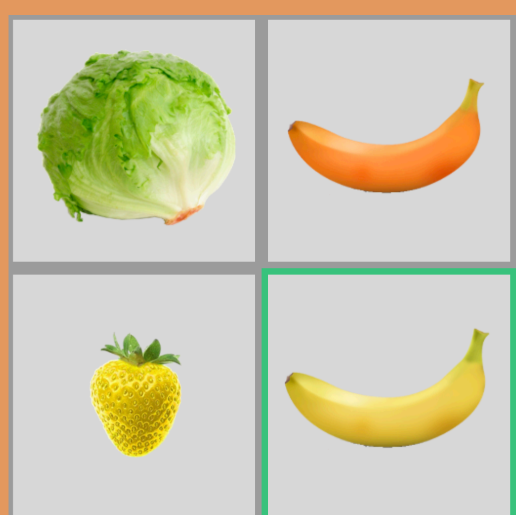
Hahn et al 2018; Scontras et al 2019

Other

plural predication Scontras & Goodman 2017
I-implicatures Poppels & Levy, 2016
generics Tessler & Goodman, 2019
modals Herbstritt & Franke, 2017
vague quantifiers Schöller & Franke, 2017
convention formation Hawkins et al 2018; 2019
questions Hawkins et al 2015
pragmatic adaptation Schuster & Degen, 2020
exhaustivity inferences Javangula & Degen in prep
atypicality inferences Kratvchenko & Demberg
social meaning Burnett 2017; 2019

The Rational Speech Act framework (RSA)

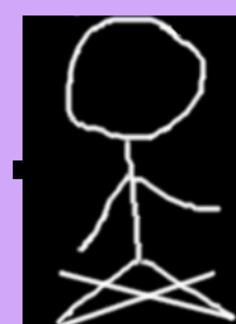
Comprehension:
contrastive inferences

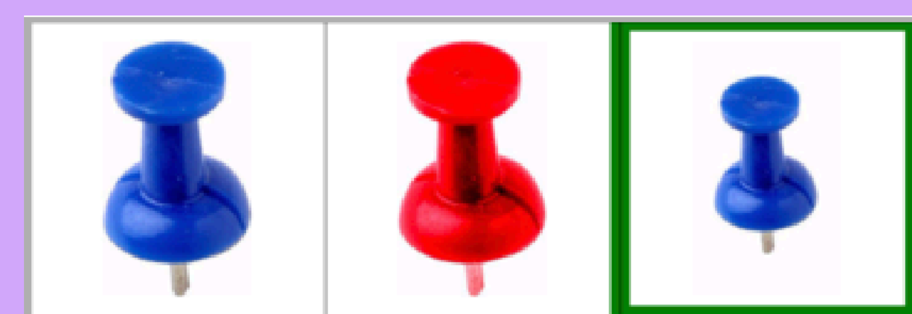


"Click on the yellow…"

$$P_{L_1}(r\,|\,u) \propto \boxed{P_{S_1}(u\,|\,r)} \cdot P(r)$$

Production:
redundant referring expressions





the small blue pin
la tachuela pequeña

$$\boxed{P_{S_1}(u\,|\,r)} = e^{\alpha(\ln \boxed{P_{L_0}(r|u)} - C(u))}$$

$$\boxed{P_{L_0}(r\,|\,u)} \propto [[u]](r) \cdot P(r)$$

Elisa Kreiss

PART I

—

Comprehension of referring expressions: contrastive inferences

Kreiss & Degen 2020, in prep

Contrastive Inference (CI)

….is elicited by the contrastive function of adjectives.
….is elicited by relative, but not absolute adjectives.

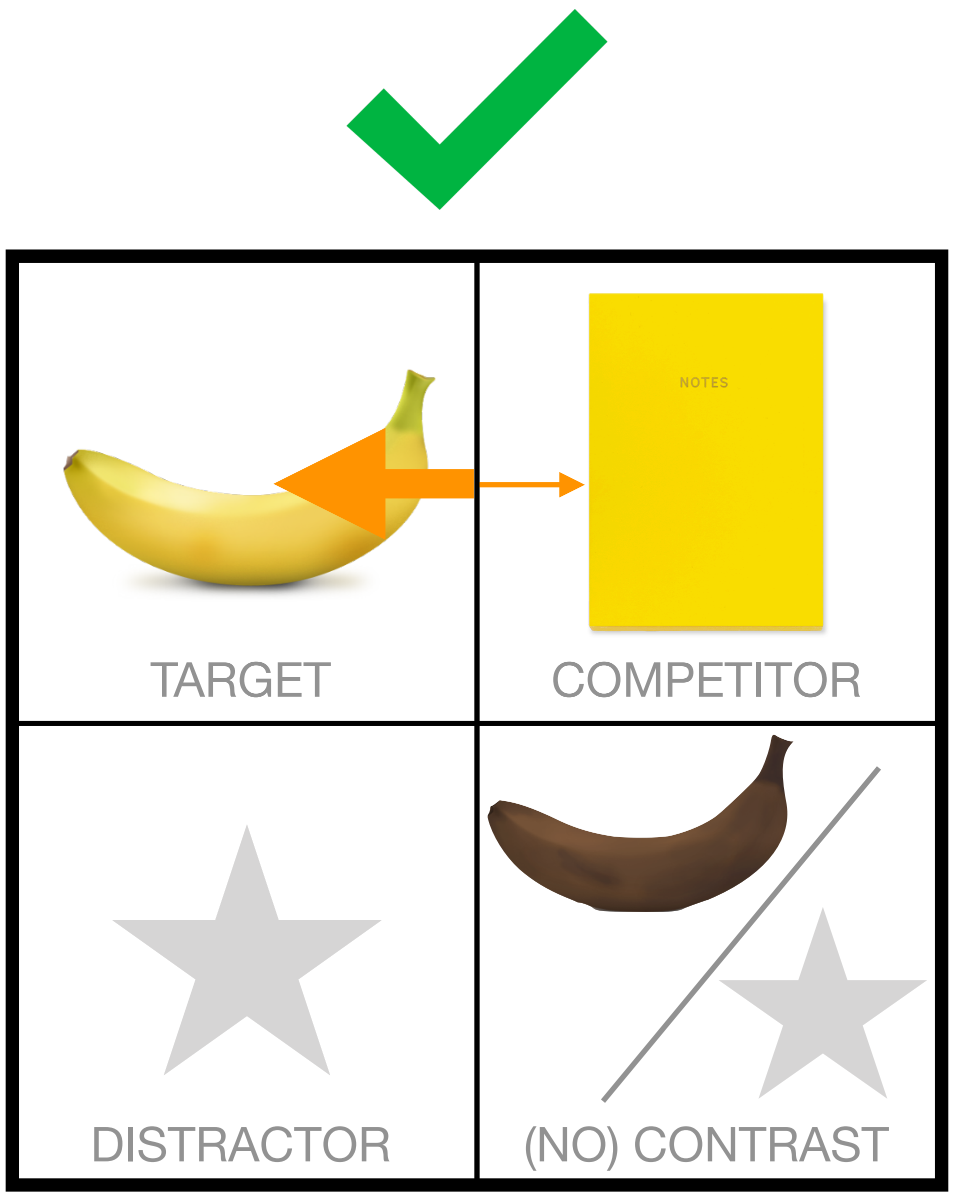
“Click on the (adj) … !”

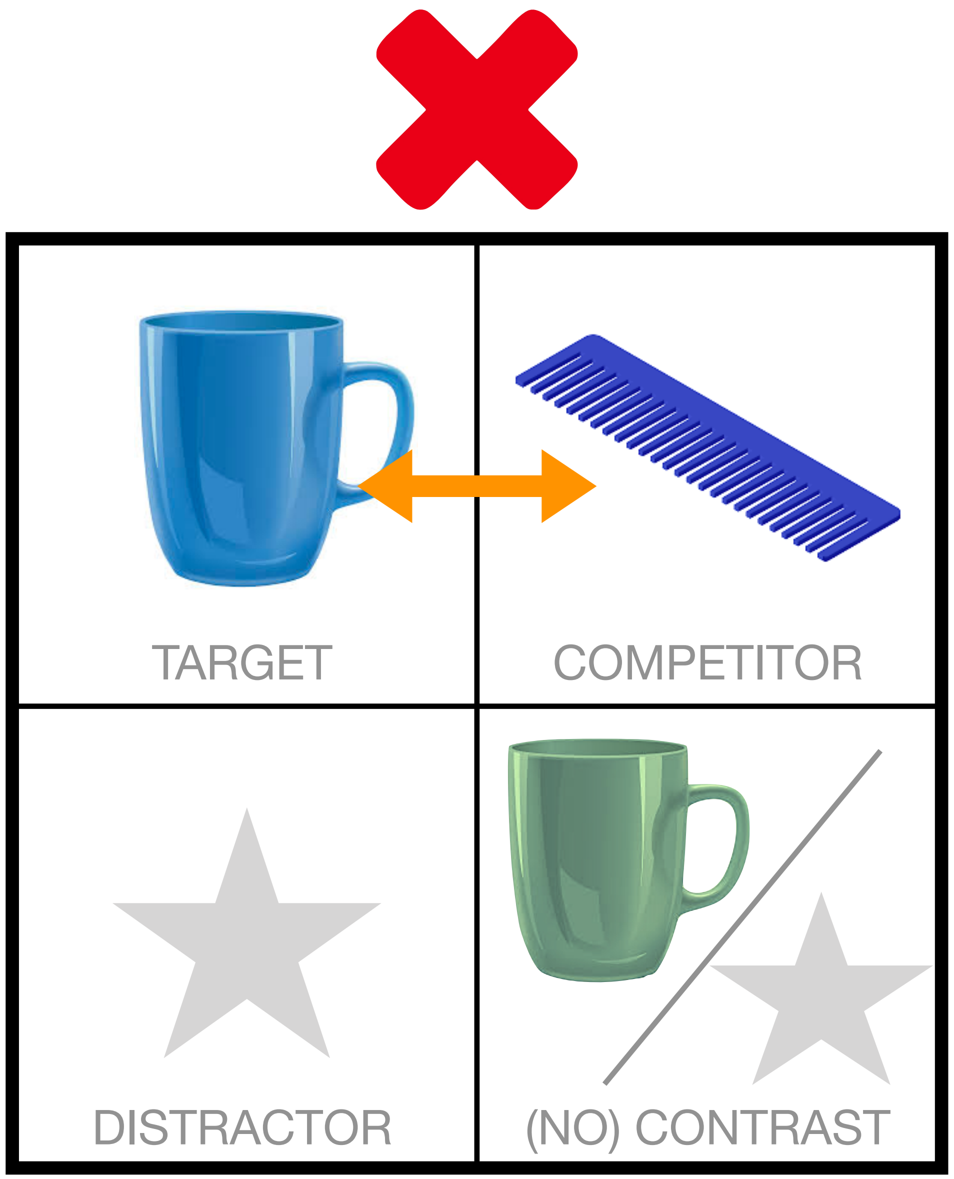
Sedivy et al (2003)

Contrastive Inference (CI)

… is elicited by relative, but not (?) absolute adjectives.

"Click on the (adj) … !"





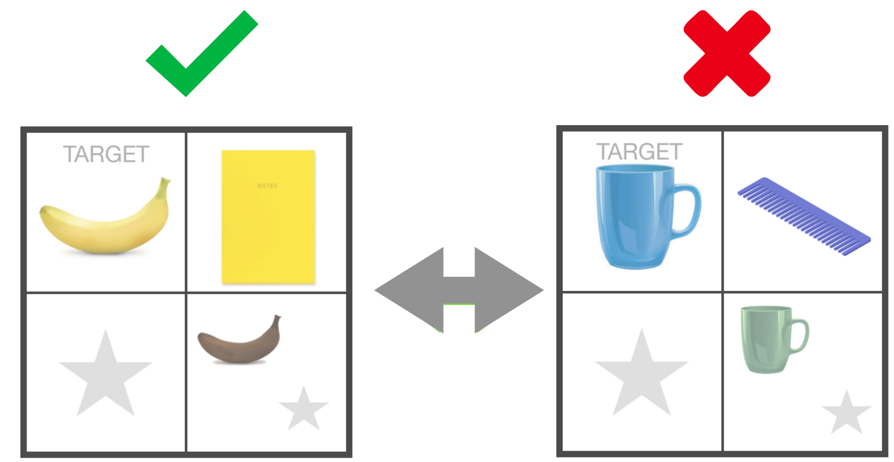
Sedivy et al (2003)

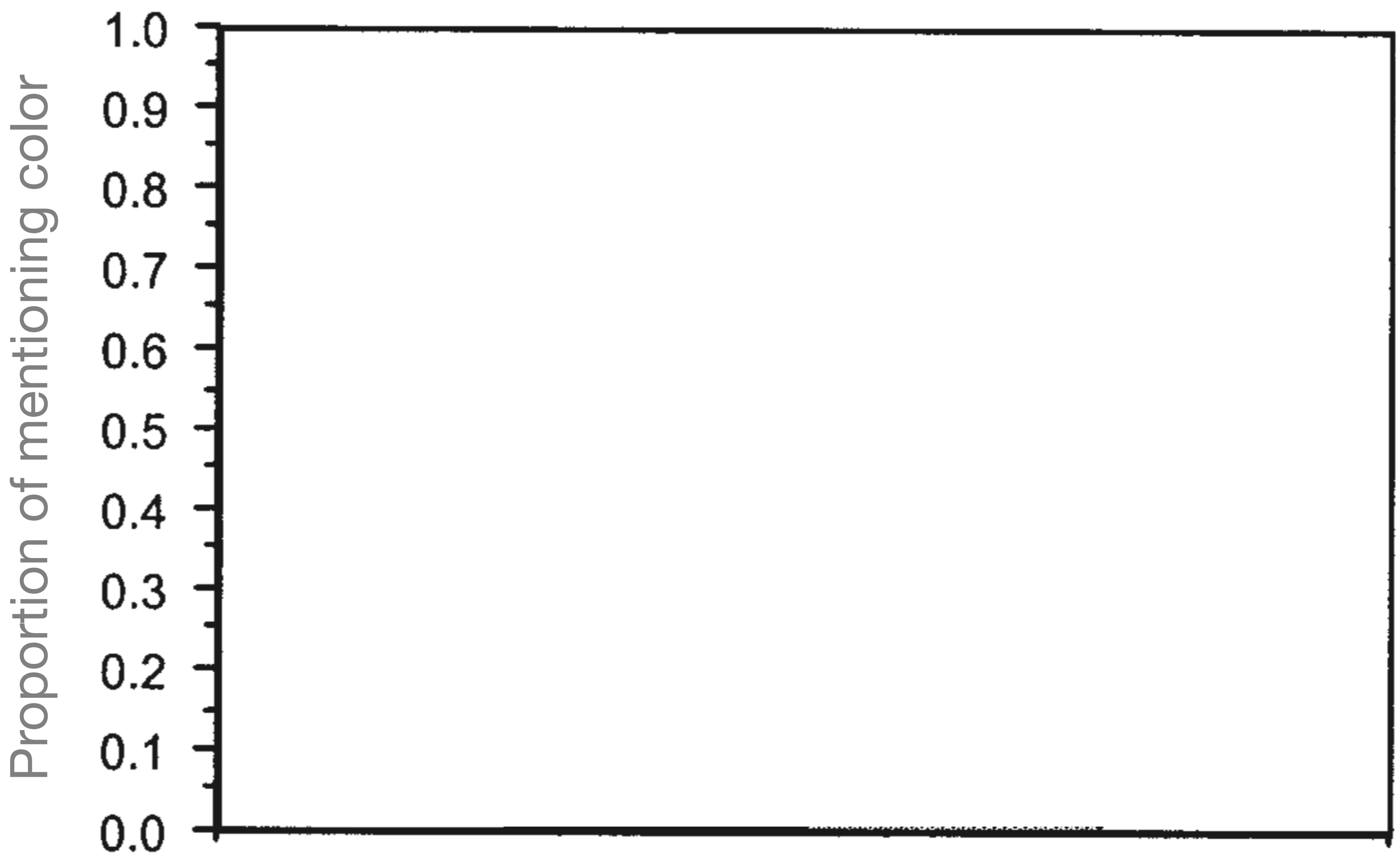
Contrastive Inference (CI)

… is elicited by relative, but not (?) absolute adjectives.

is elicited by all relative adjectives and absolute adjectives with a low production probability.

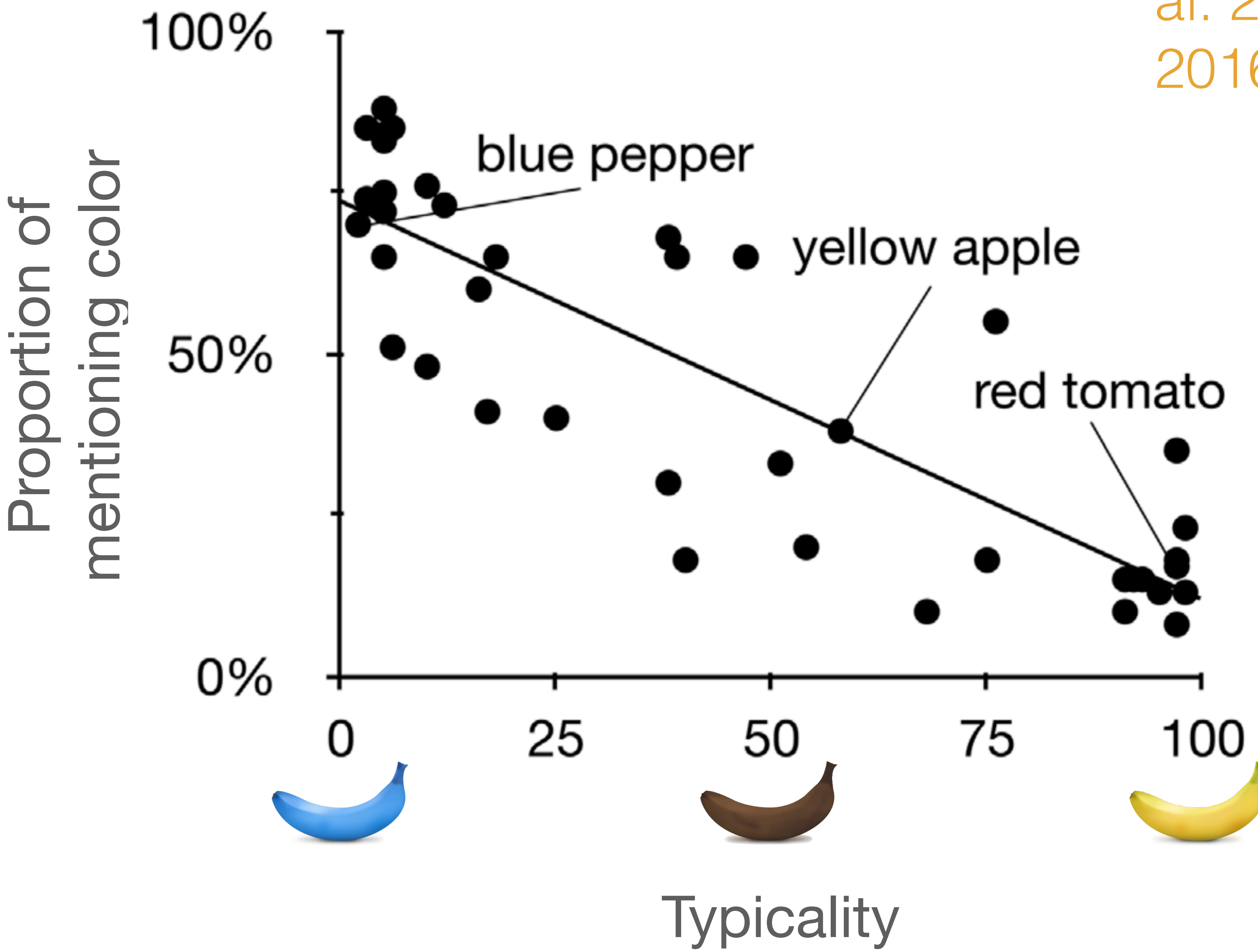
Speech production





Sedivy et al (2003)

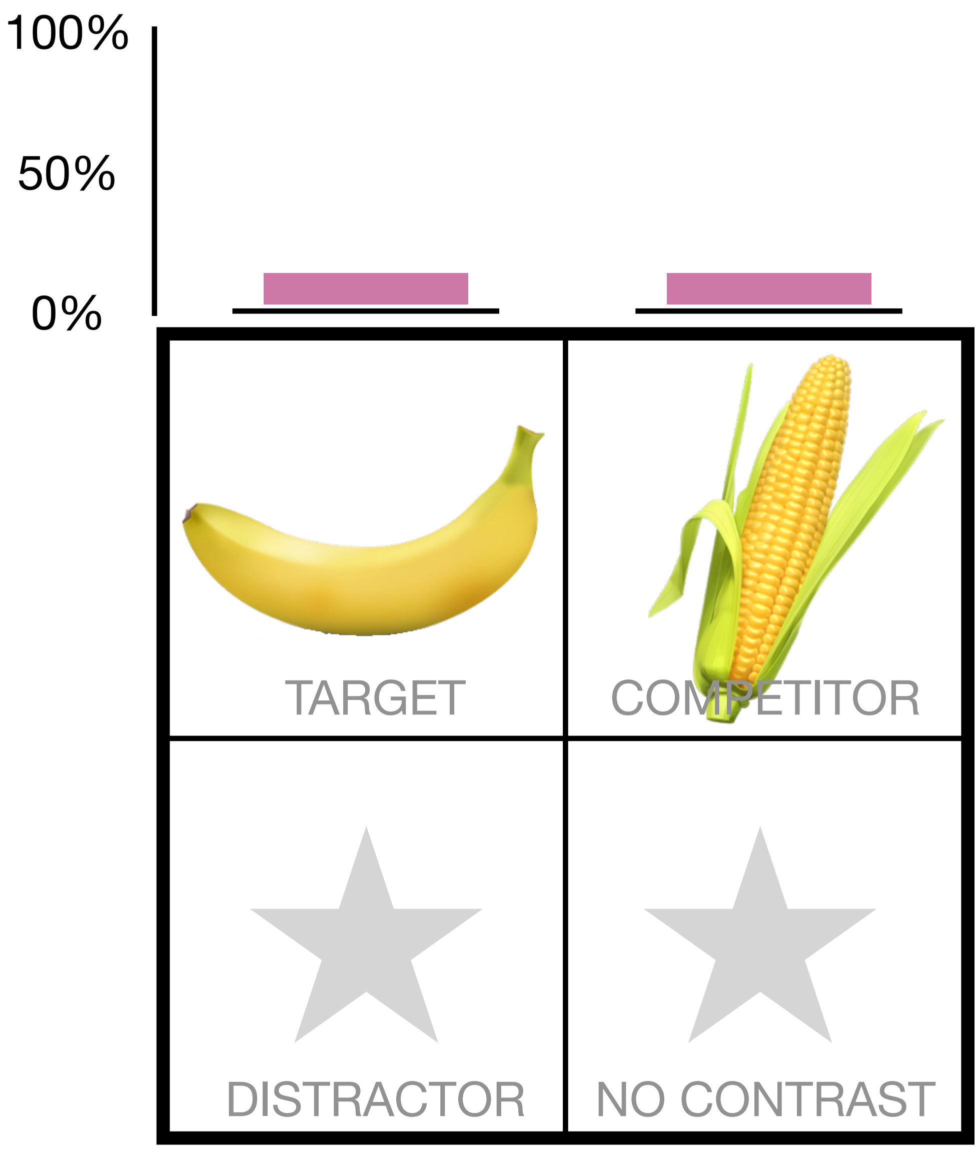
Sedivy 2003; Westerbeek et al. 2015; Rubio-Fernandez 2016; Mitchell et al. 2013

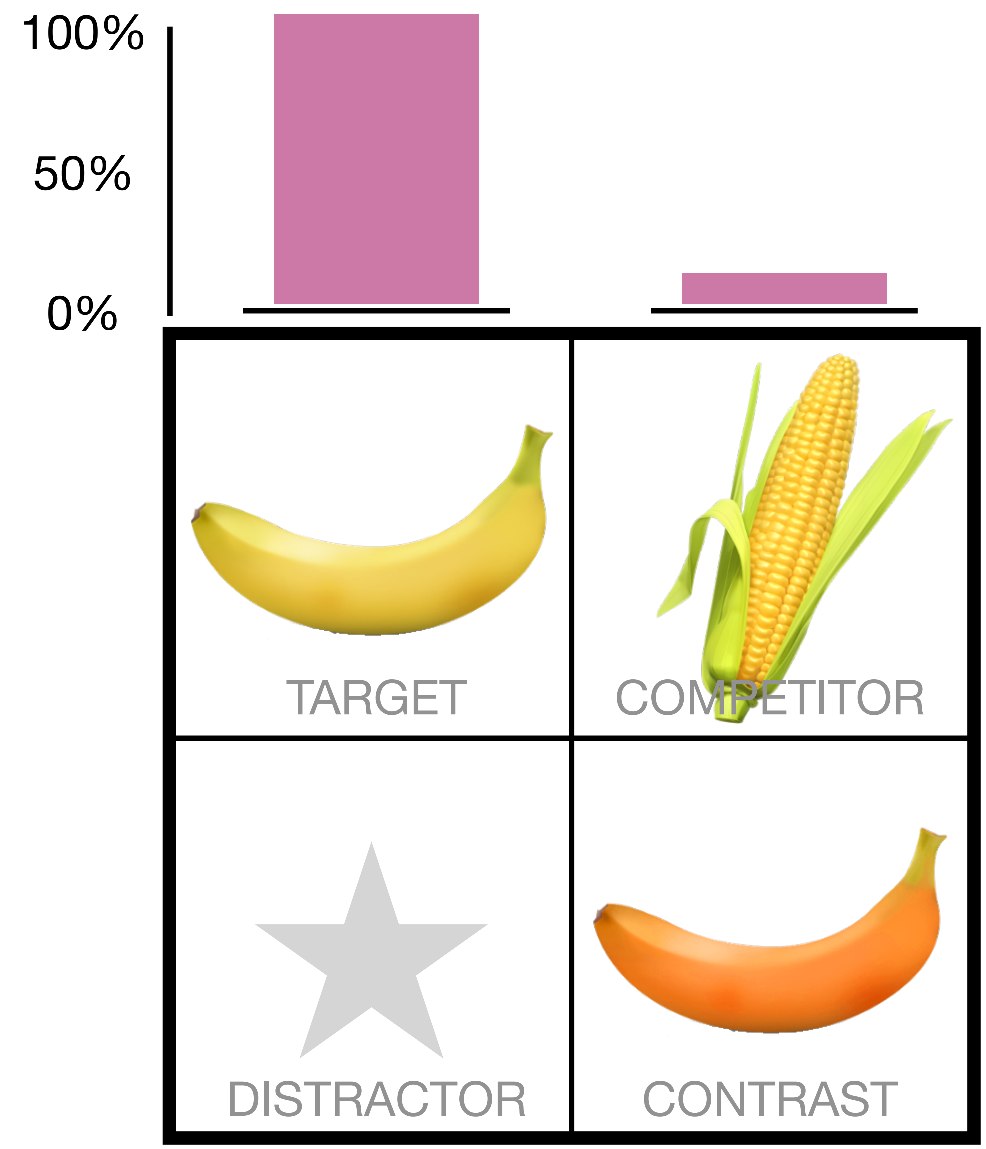


Rational Speech Act Model

$$P_{L_1}(r \mid u) \propto \boxed{P_{S_1}(u \mid r)} * P_{S_1}(r)$$

“Click on the yellow…”

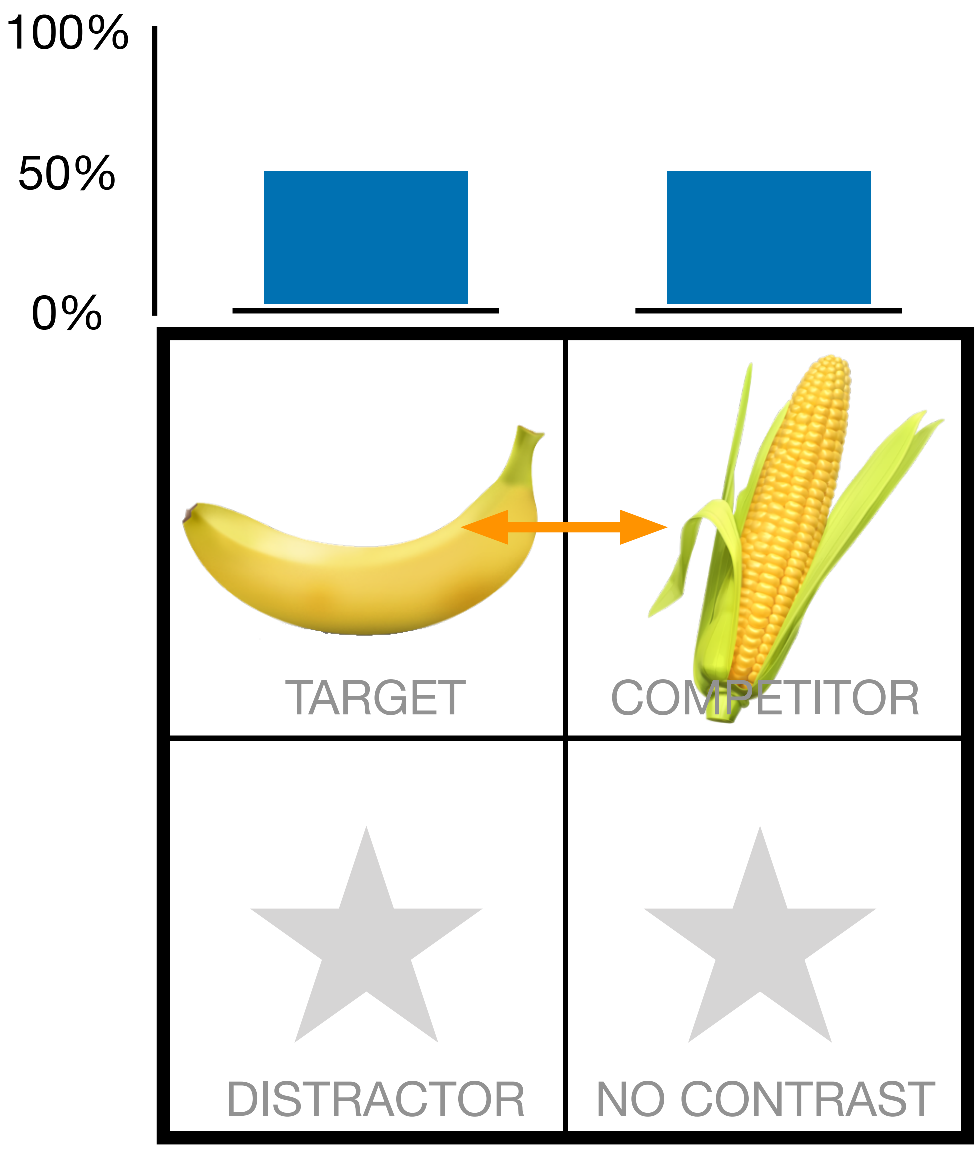




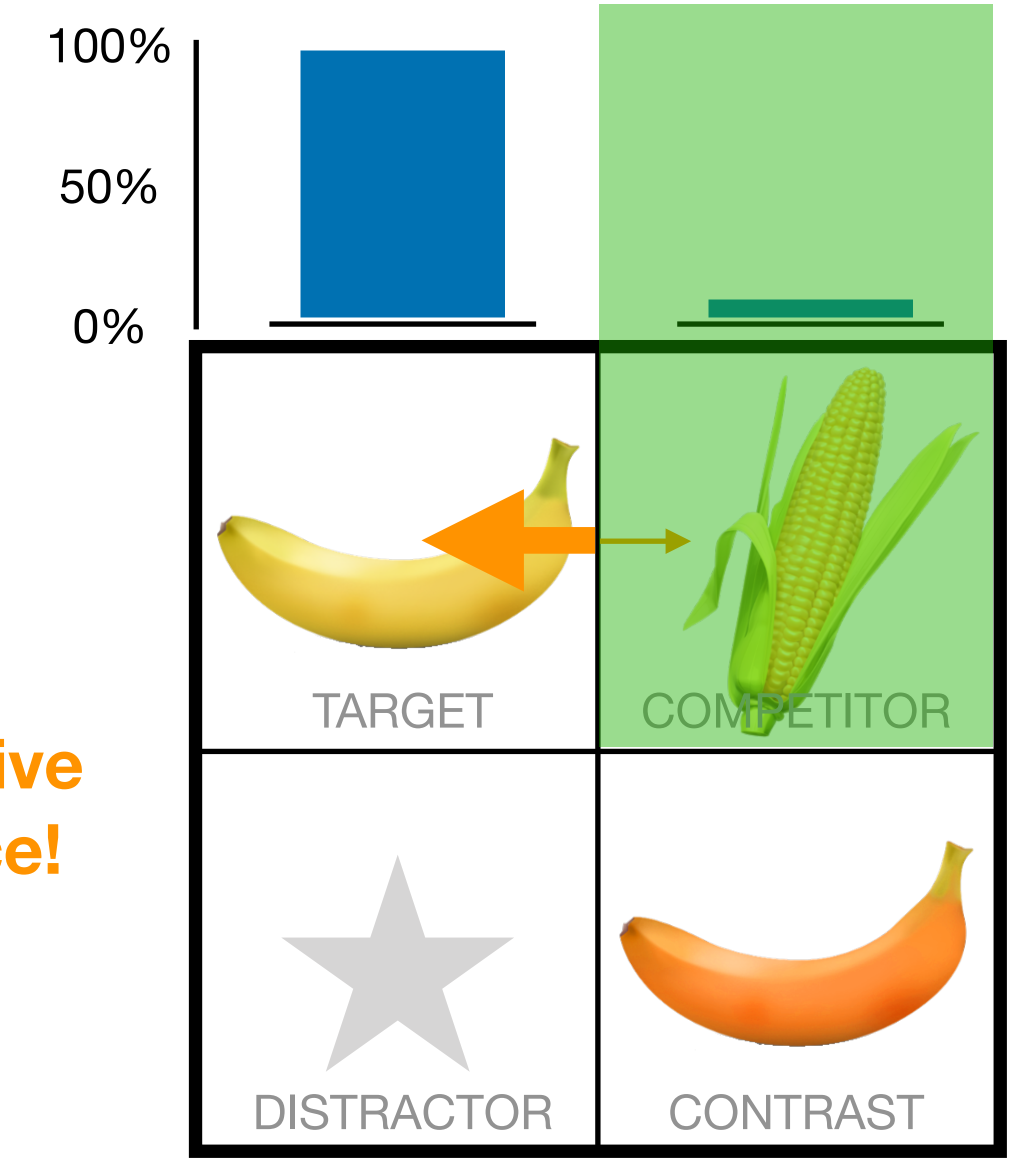
Rational Speech Act Model

$$\boxed{P_{L_1}(r\,|\,u)} \propto P_{S_1}(u\,|\,r) * P_{S_1}(r)$$

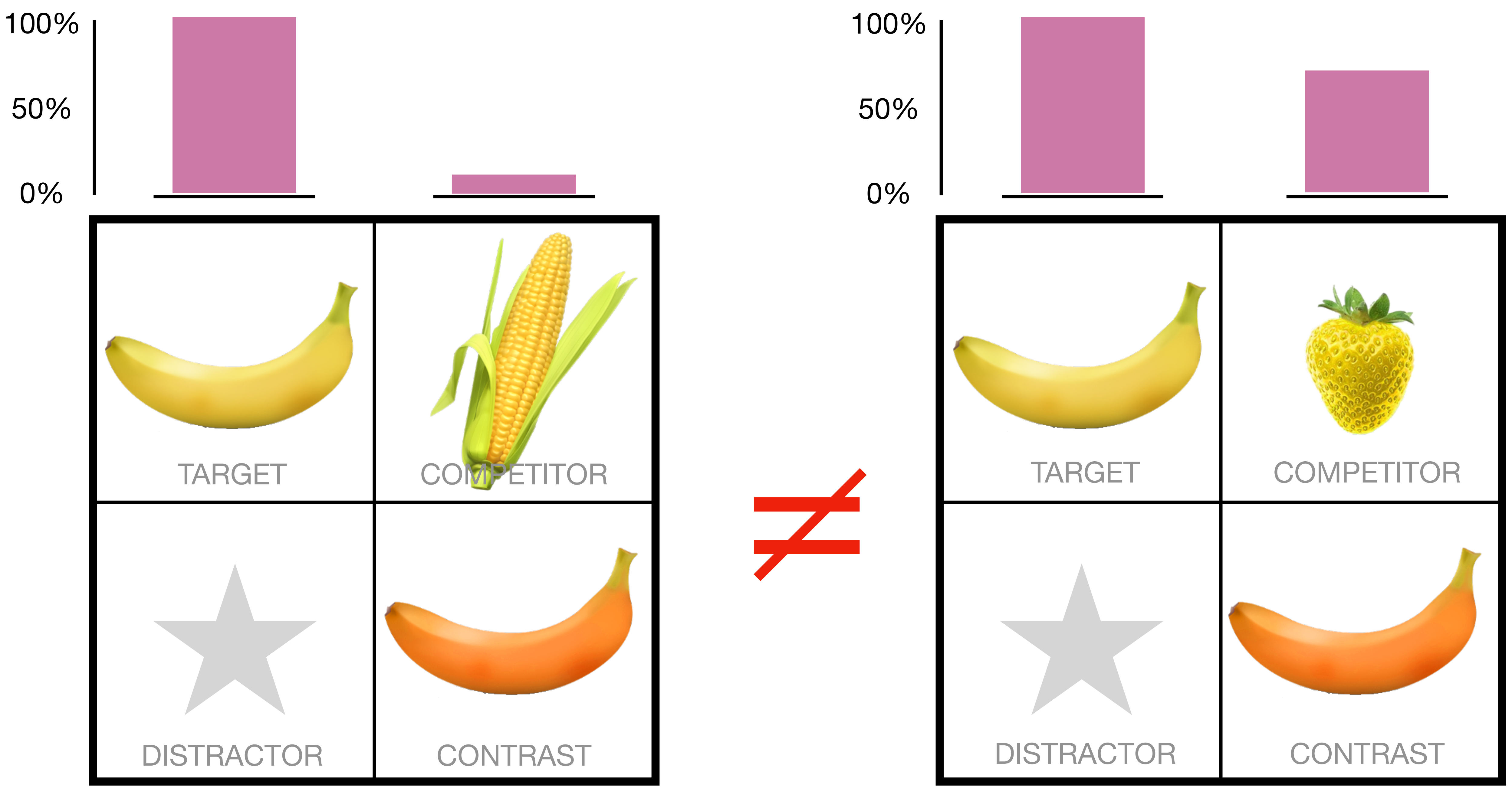
"Click on the yellow … !"



contrastive inference!

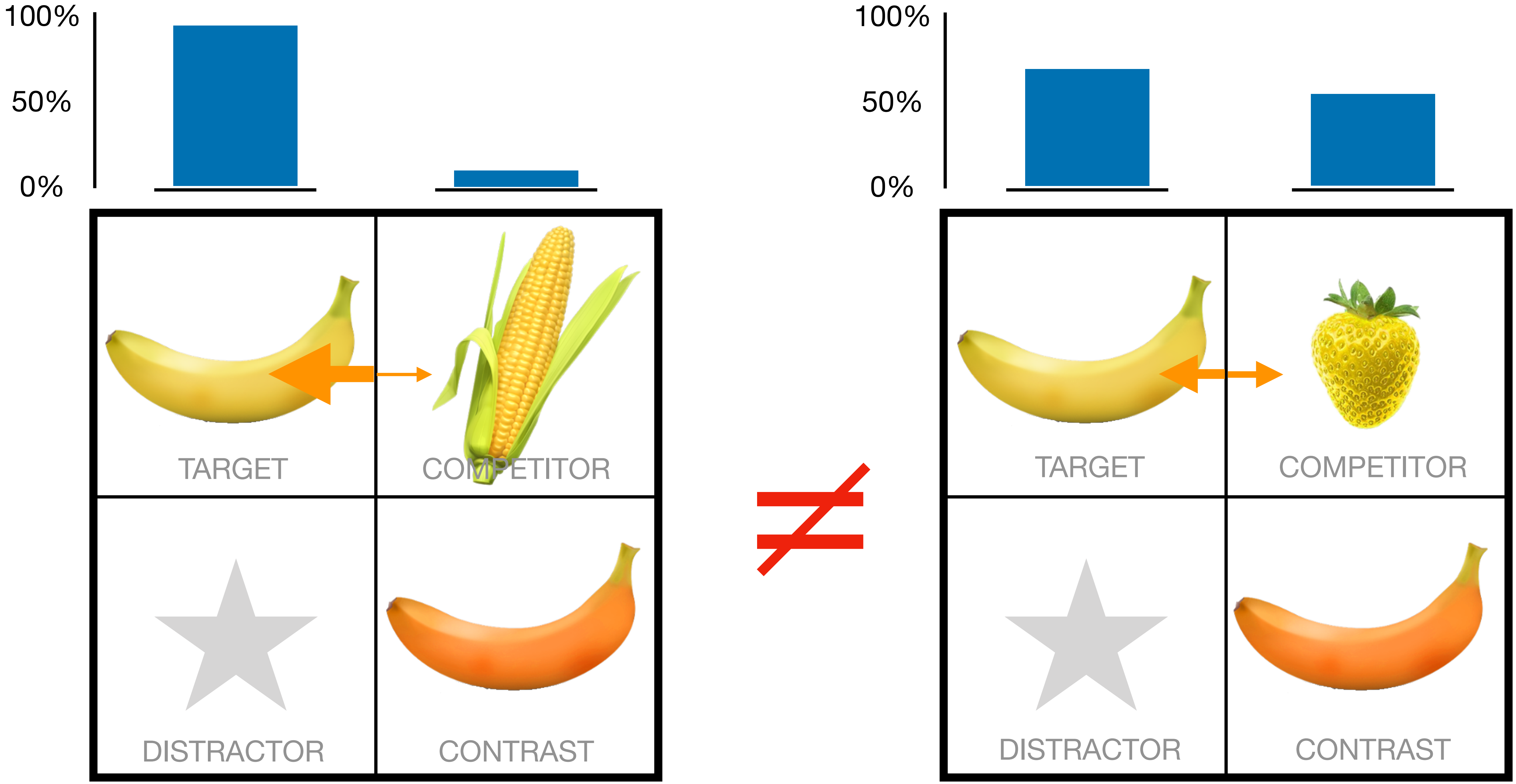


Prediction



100%
50%
0%
TARGET
COMPETITOR
DISTRACTOR
CONTRAST
100%
50%
0%
TARGET
COMPETITOR
DISTRACTOR
CONTRAST

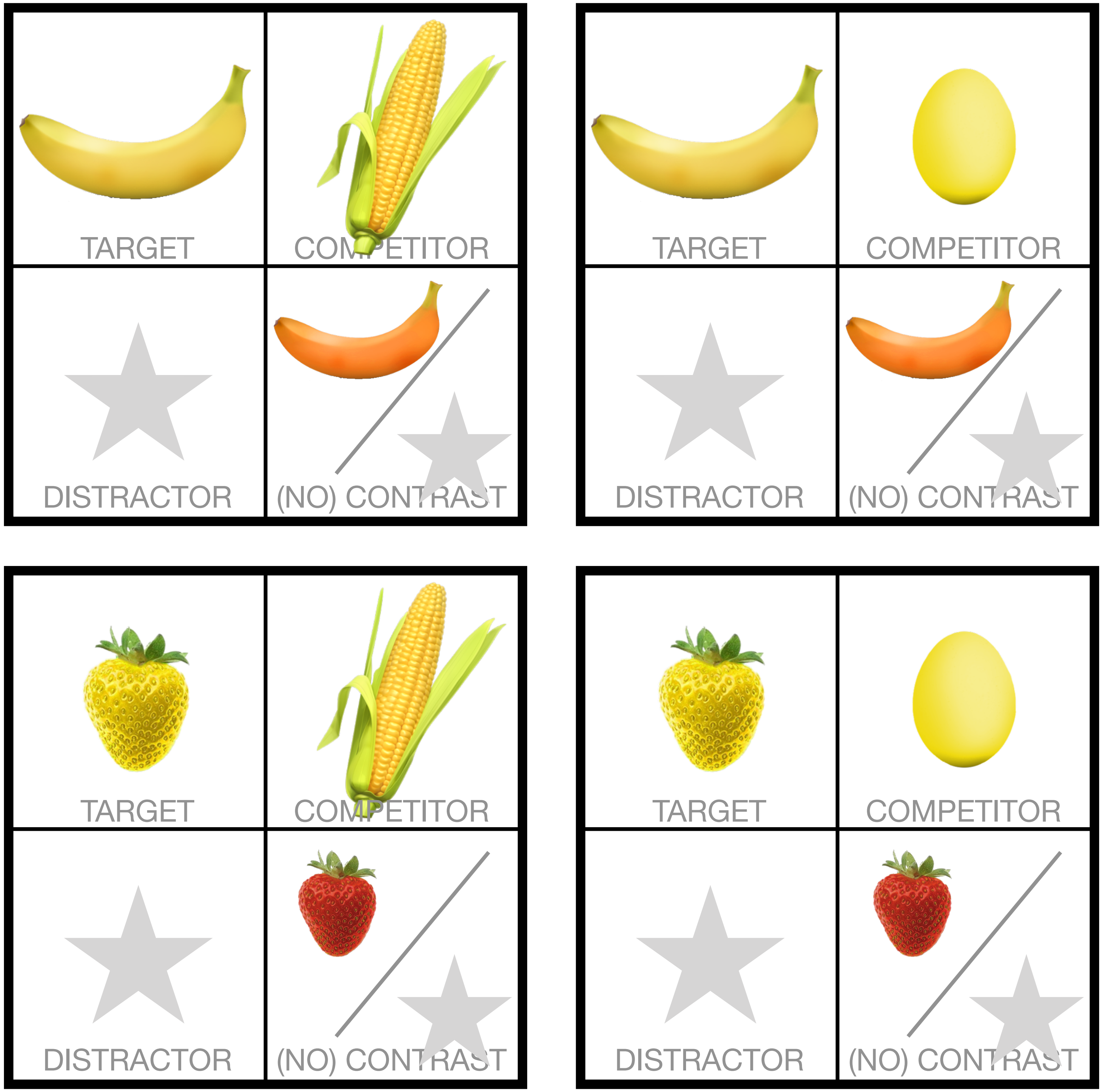
Prediction


100%
50%
0%
TARGET
COMPETITOR
DISTRACTOR
CONTRAST
100%
50%
0%
TARGET
COMPETITOR
DISTRACTOR
CONTRAST

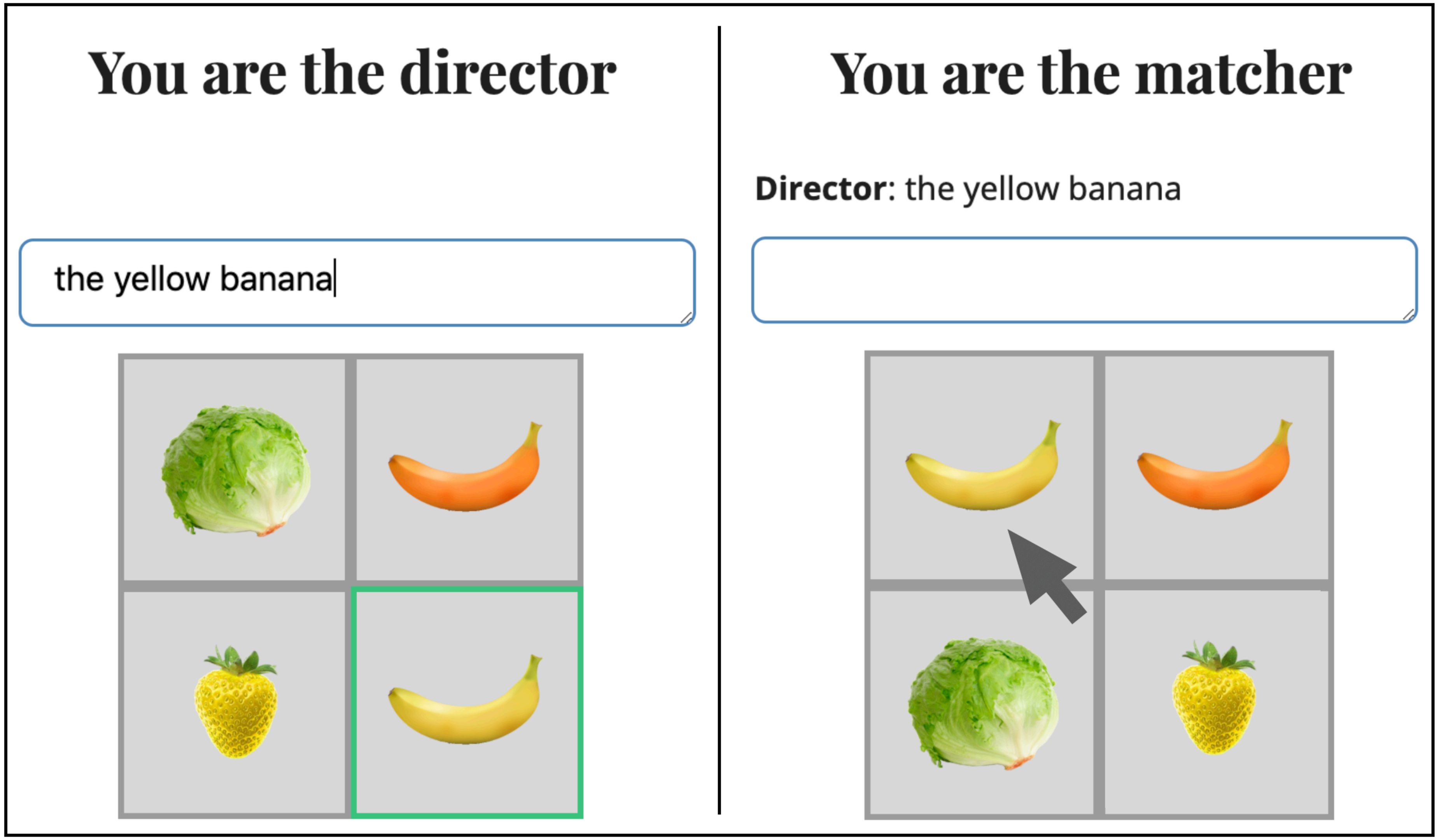
Test RSA

1. **Production experiment** to obtain model predictions
2. **Model predictions** using production data
3. **Comprehension experiment** to assess model predictions

2 x 2 x 2 design:
contrast presence, target typicality, competitor typicality



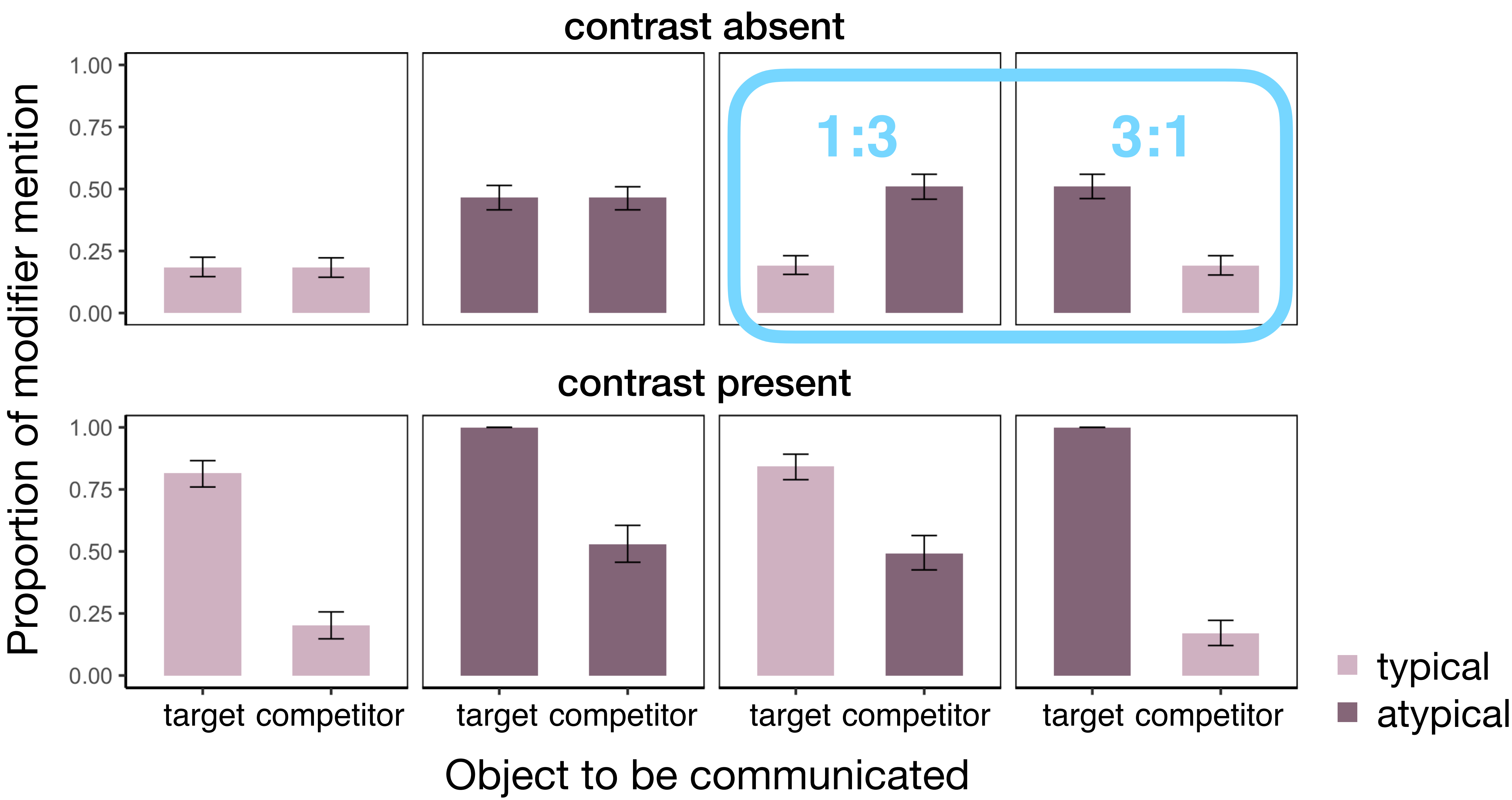
Production experiment



- Interactive reference game
- MTurk, n=141 participant pairs (112)
- 60 trials (32 critical; 28 fillers)

Clark & Wilkes-Gibbs (1986), Hawkins et al. (2015)

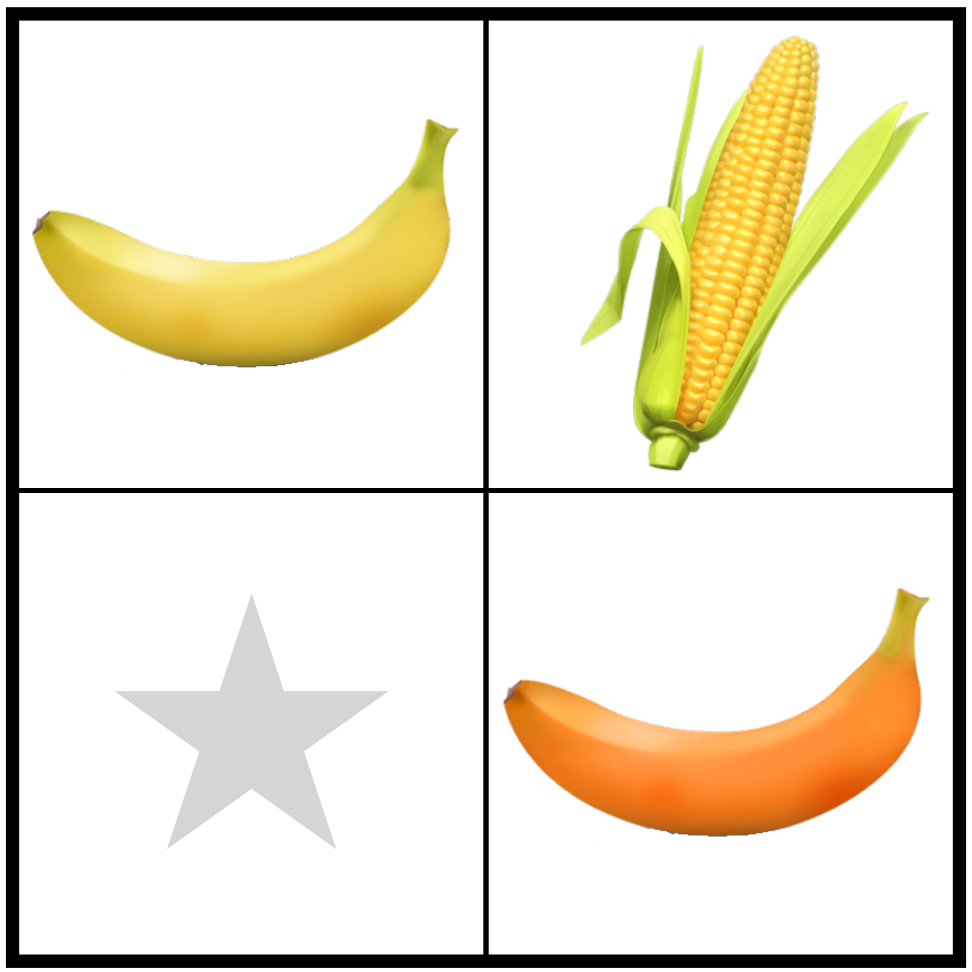
Production results



$$P_{L_1}(r|u) \propto \boxed{P_{S_1}(u|r)} \cdot \boxed{P(r) \quad \text{uniform}}$$

Comprehension experiment

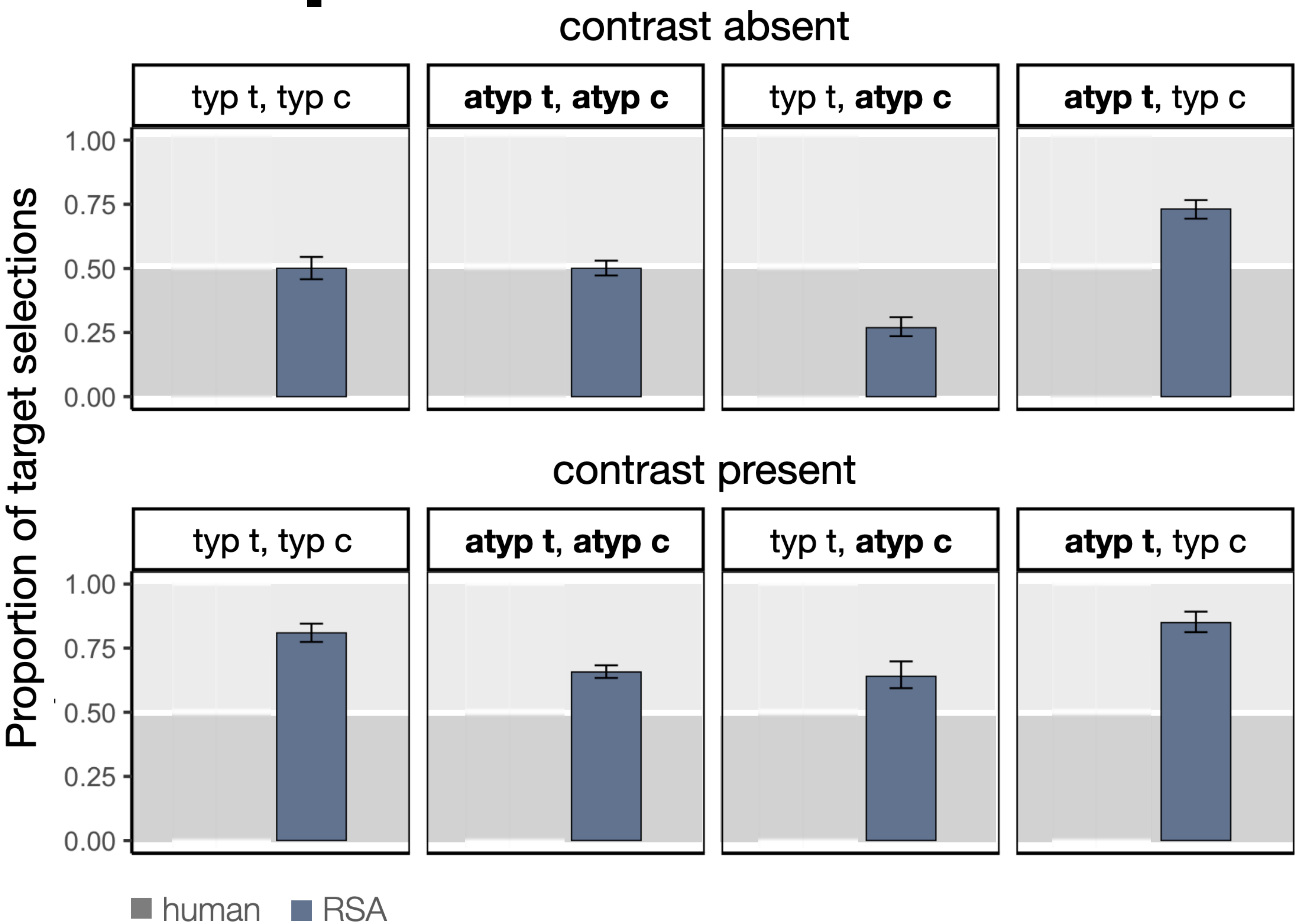
Click on the yellow banana!



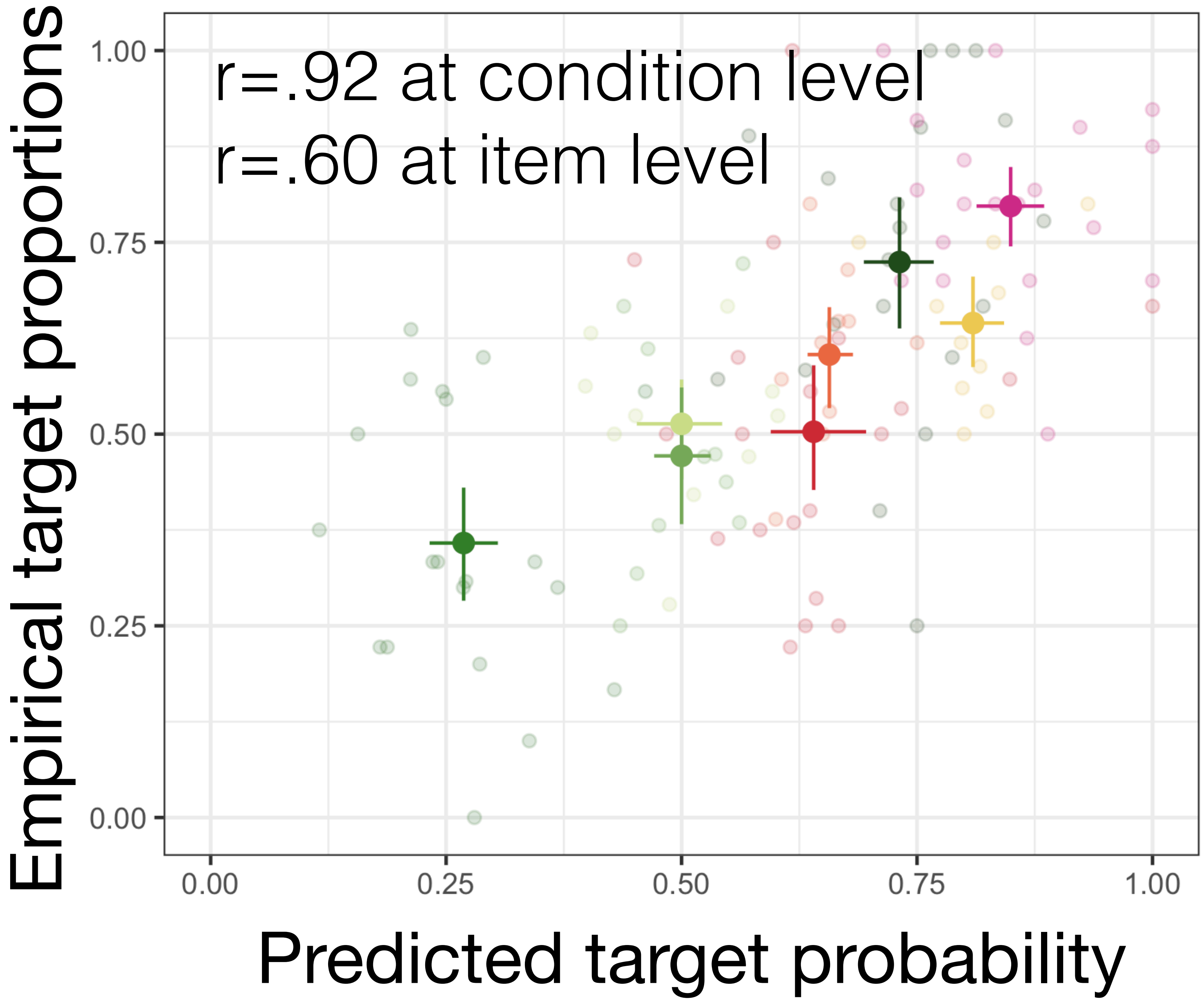
- incremental decision task rather than eye-tracking study
- competitor typicality manipulated between-subjects
- MTurk, n=79 (73)
  40 per competitor typicality condition
- 55 trials (20 critical; 35 fillers)

Qing et al. 2018, Degen et al under review

Comprehension results



Comparison of predicted and empirical target selections



Simple pragmatic RSA listener captures variability in contrastive inference strength by reasoning about (variability in) modifier production probabilities.

Interim summary I

A simple pragmatic RSA listener captures variability in contrastive inference strength by reasoning about (variability in) modifier production probabilities.

Some consequences:

effects of adjective function or semantics only operative via their effect on listeners' production expectations

no need for "default descriptions"

Caroline Graf

Robert Hawkins

Leyla Kursat

Brandon Waldon

Noah Goodman

Elisa Kreiss

PART II

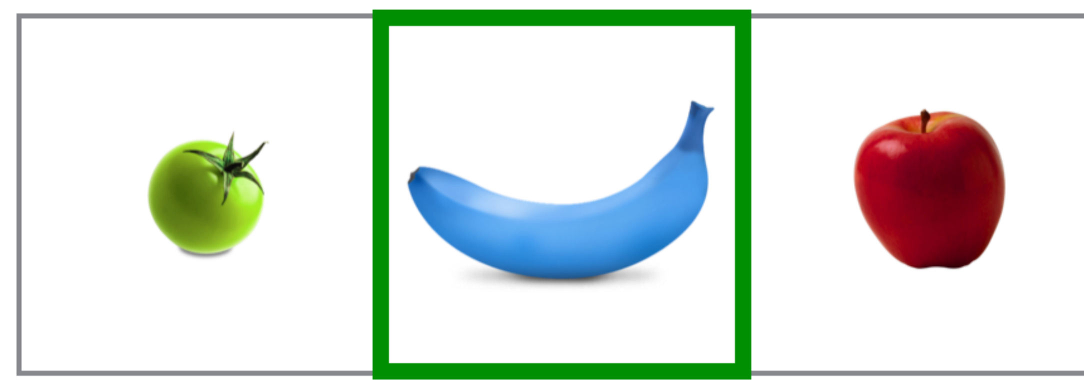
—

Production of referring expressions: redundant modification

Degen et al 2020; Waldon & Degen 2021; Kursat & Degen to appear

1. **Redundant color and size modification**
   Degen et al 2020

2. **Model extension: color typicality**
   Degen et al 2020



3. **Model extension: cross-linguistic variability**
   Waldon & Degen 2021

the small blue pin
la tachuela pequeña

Production: redundant referring expressions

size sufficient

the small pin

color sufficient

the blue pin

75-80% ***the small blue pin*** **8-10%**

Speakers produce seemingly overinformative referring expressions.

Deutsch 1976; Pechmann 1989; Sedivy 2003; Gatt et al. 2011; Koolen et al 2013; Rubio-Fernández 2016; Westerbeek et al 2015; Davies & Katsos 2013; van Gompel et al 2019

The RSA framework

Frank & Goodman 2012

$O = \{$  $\}$

$U = \{\text{big}, \text{small}, \text{green}, \text{black}\}$
$\text{big green}, \text{small green}, \text{small black}\}$

obvious problem:
no complex utterances

Literal listener

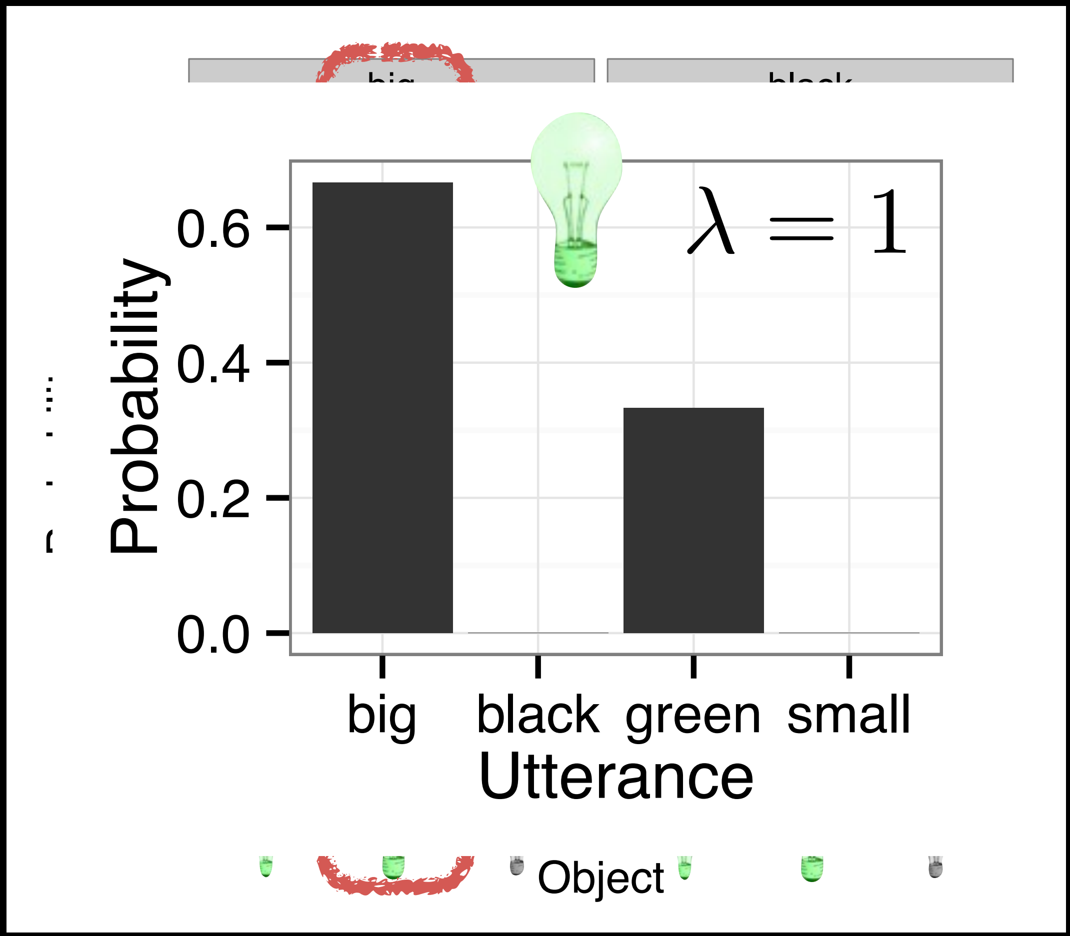
$$P_{L_0}(o|u) = \mathcal{U}(o|\{u \text{ is true of } o\})$$

$$[[u]] : O \to \{\text{true}, \text{false}\}$$

Pragmatic speaker

$$P_{S_1}(u|o) \propto e^{\lambda \cdot (\ln P_{L_0}(o|u) - C(u))}$$

Quantity Manner



Utterance semantics & cost

Intersective semantics

$$[[u]] = [[u_1]] \wedge [[u_2]]$$

$$[[\text{big green}]] = [[\text{big}]] \wedge [[\text{green}]]$$

Cost

$$C(u) = C(u_1) + C(u_2)$$

RSA does not produce overinformative REs…

Gatt et al 2013; Westerbeek et al 2015

…with deterministic Boolean semantics

Motivation for relaxed semantics?

Modifiers differ:

size adjectives are more vague and context-dependent than color adjectives

color is more salient than size
Arts et al 2011; Gatt et al 2013

size adjectives are judged to be more subjective than color adjectives
Scontras, Degen, & Goodman 2017

Non-deterministic semantics

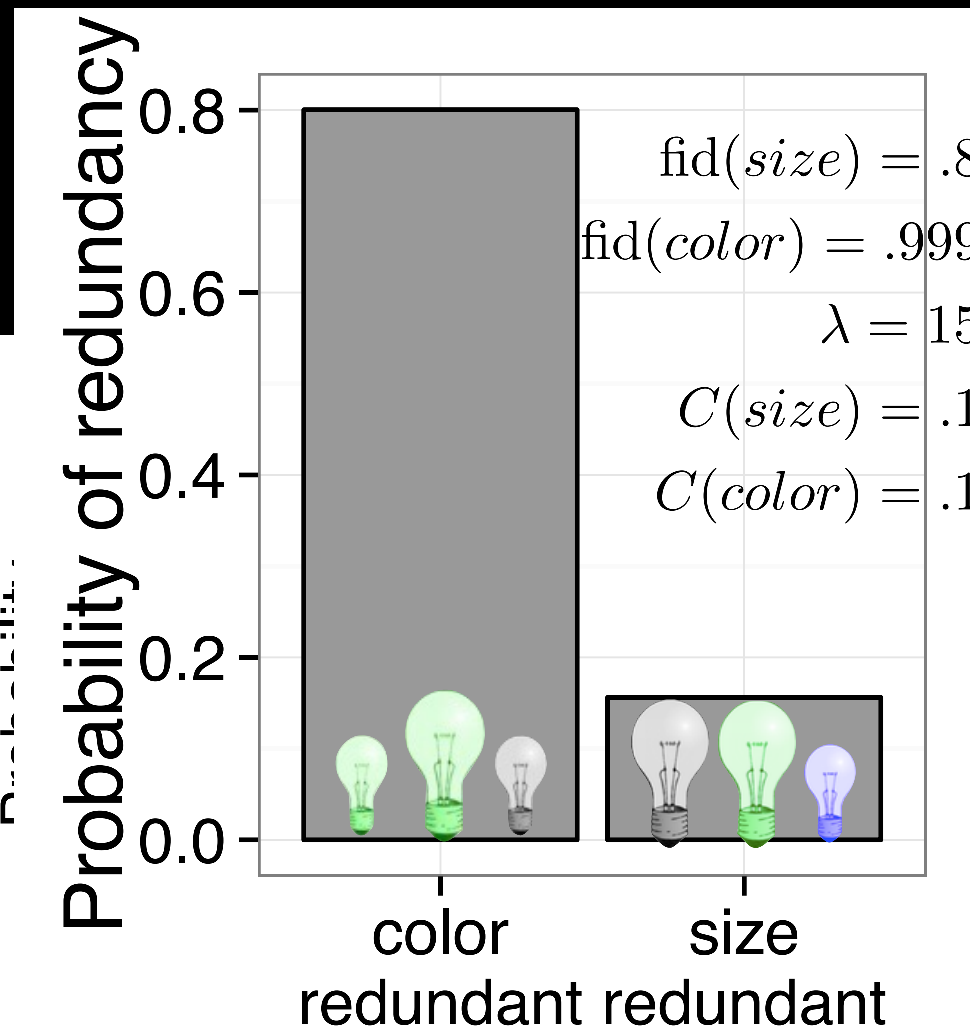
Literal listener

fidelity

$$P_{L_0}(o|u) \propto \begin{cases} \boxed{1-\epsilon} & [[u]](o) = \text{true} \\ \epsilon & \text{otherwise} \end{cases}$$

Pragmatic speaker

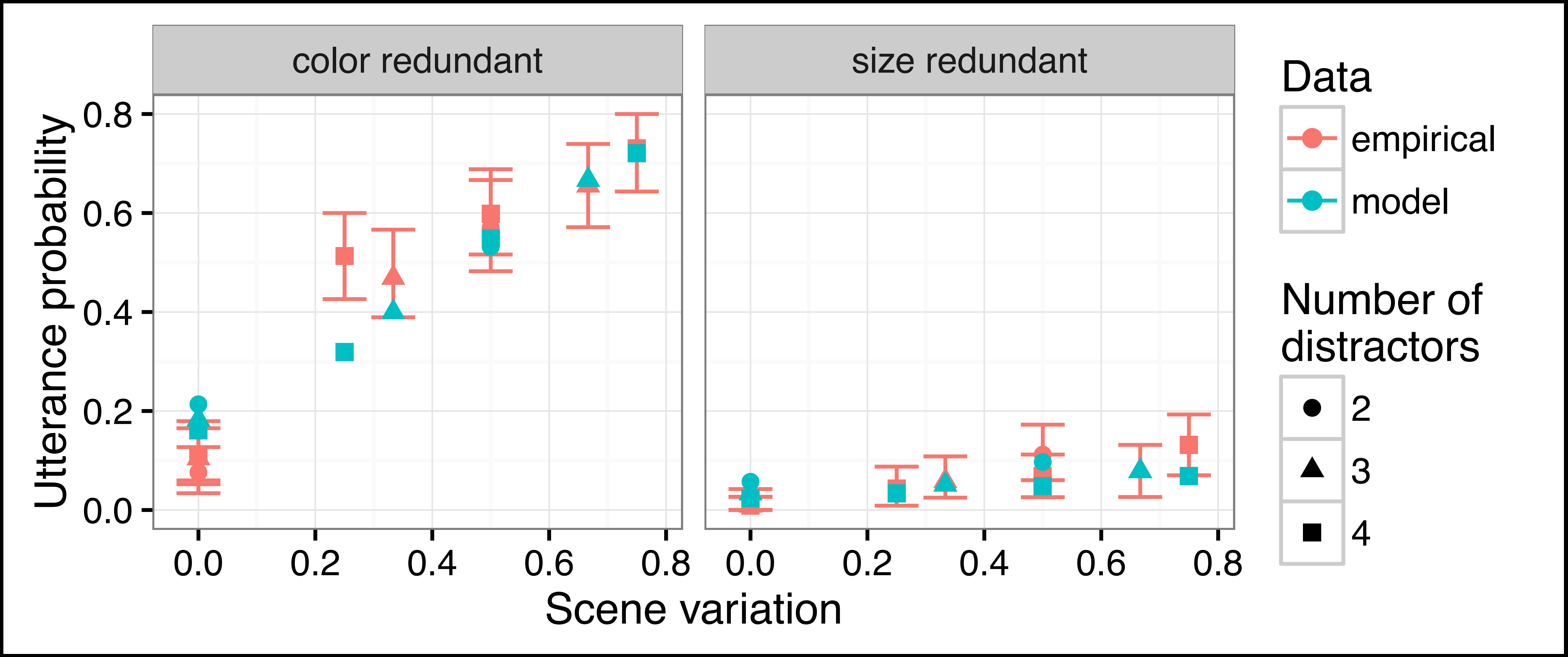
$$P_{S_1}(u|o) \propto e^{\lambda\cdot(\ln P_{L_0}(o|u) - C(u))}$$



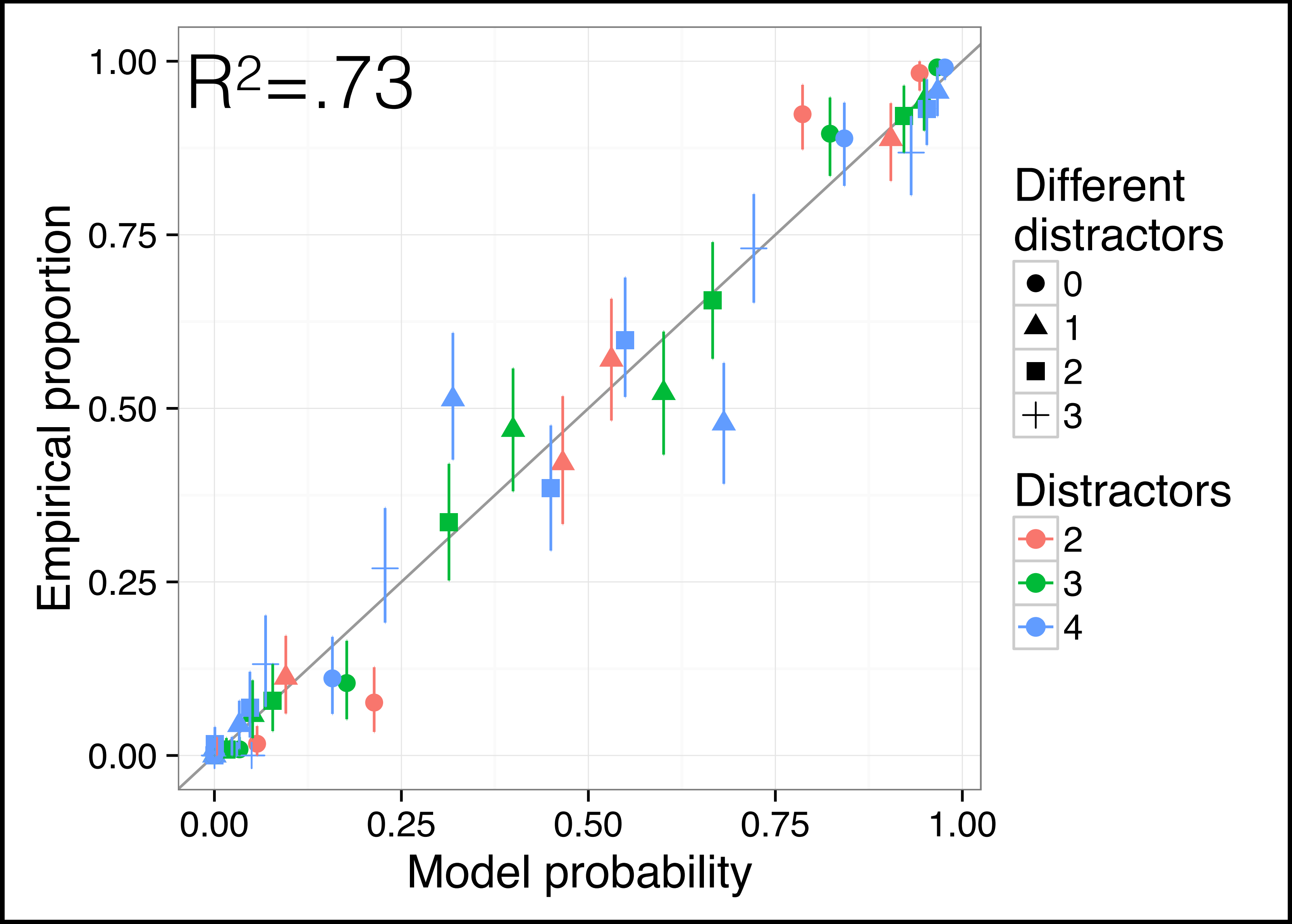
color-size asymmetry

If modifiers don't "work perfectly",
adding modifiers adds information

Interactive reference game results



Model fit



R²=.73
Empirical proportion
Model probability
0.00
0.25
0.50
0.75
1.00
Different distractors
0
1
2
3
Distractors
2
3
4

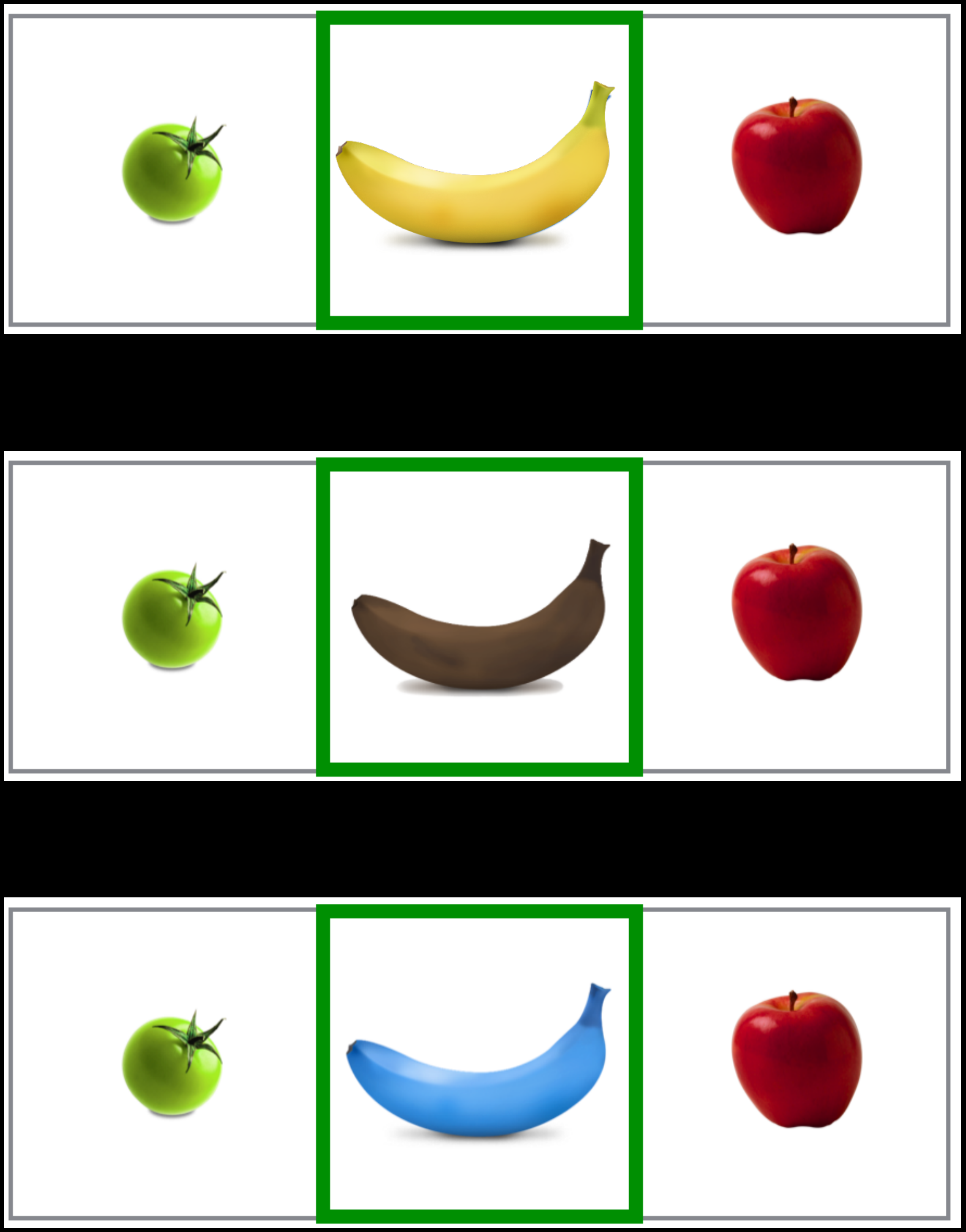
Interim summary II

A continuous-semantics RSA speaker captures variability in overmodification by reasoning about a continuous semantics literal listener: overmodification is more likely, the less inferable the intended referent is from the simply modified description.

A consequence:

“overinformative” —> “usefully redundant”

Extending the model to within-color variability



Extending the continuous semantics

Literal listener

$$P_{L_0}(o|u) \propto [[u]](o)$$

$$[[u]](o) = \boxed{\text{typicality}(u, o)}$$

Pragmatic speaker

$$P_{S_1}(u|o) \propto e^{\lambda \ln P_{L_0}(o|u) - \text{cost}(u)}$$

How typical is *o* for *u*?

x

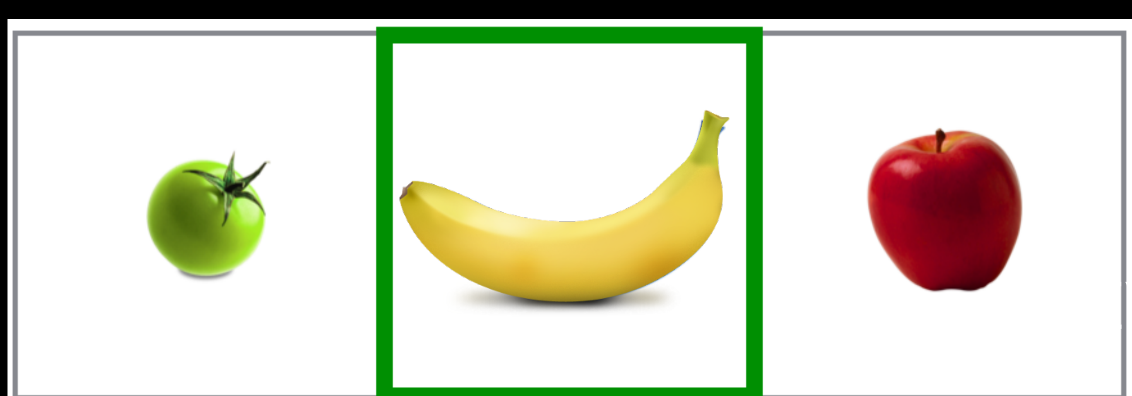
"banana"
"yellow banana"
"yellow"
"brown banana"
"brown"
…

RSA predictions with continuous semantics

typicality("banana", ) = .4

typicality("blue banana", ) = .98

typicality("banana", ) = .01

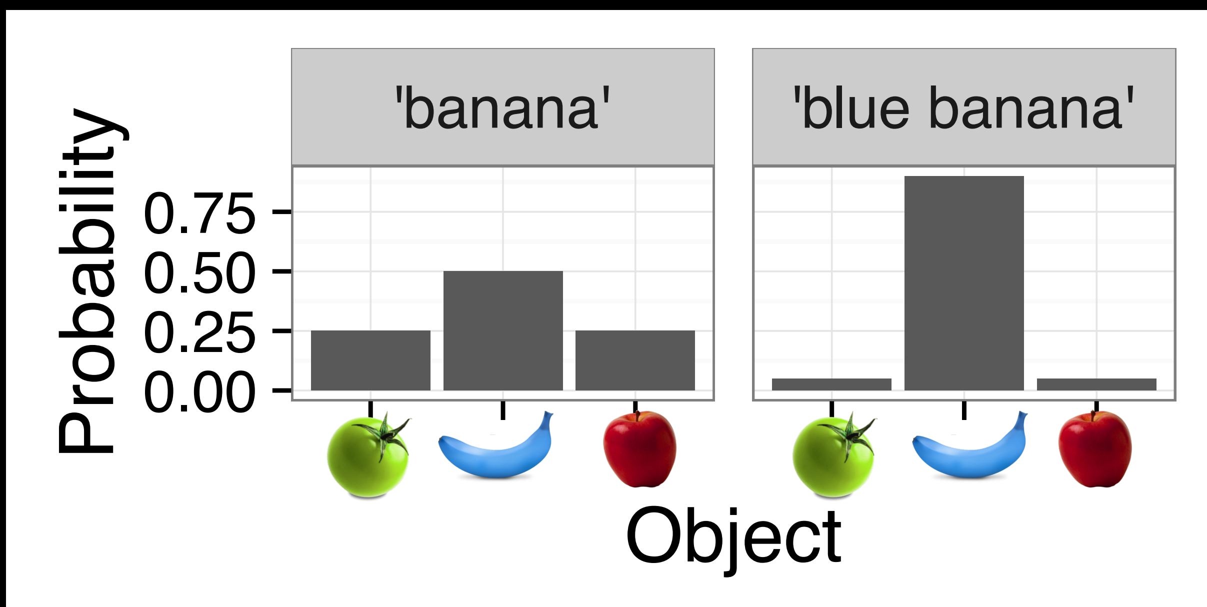


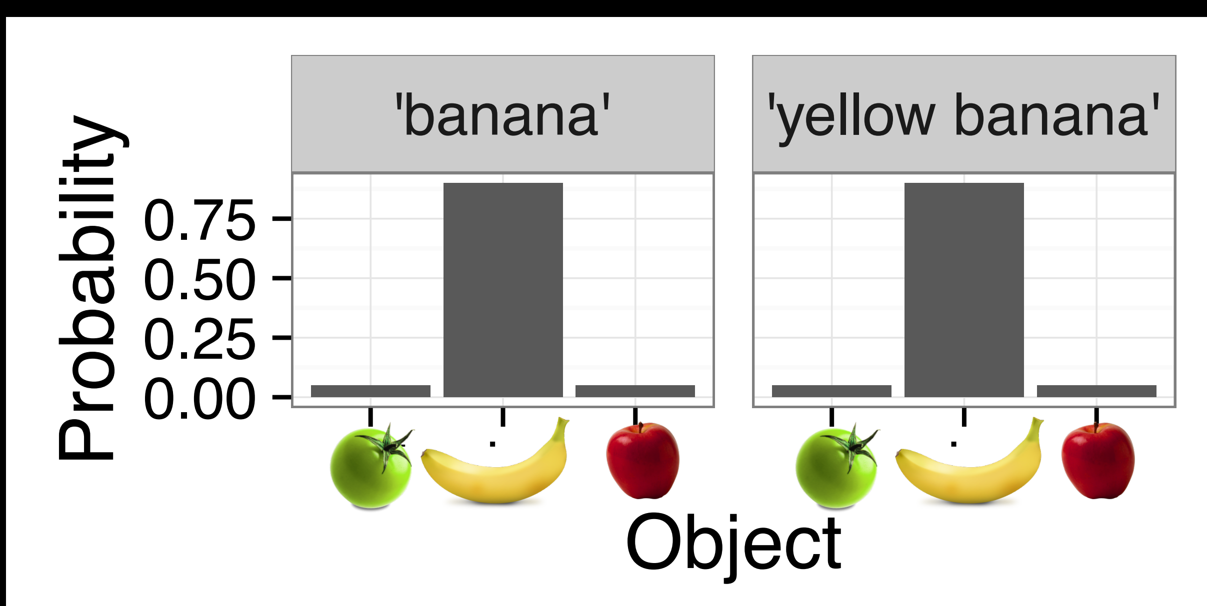
typicality("banana", ) = .98

ypicality("yellow banana", ) = .98

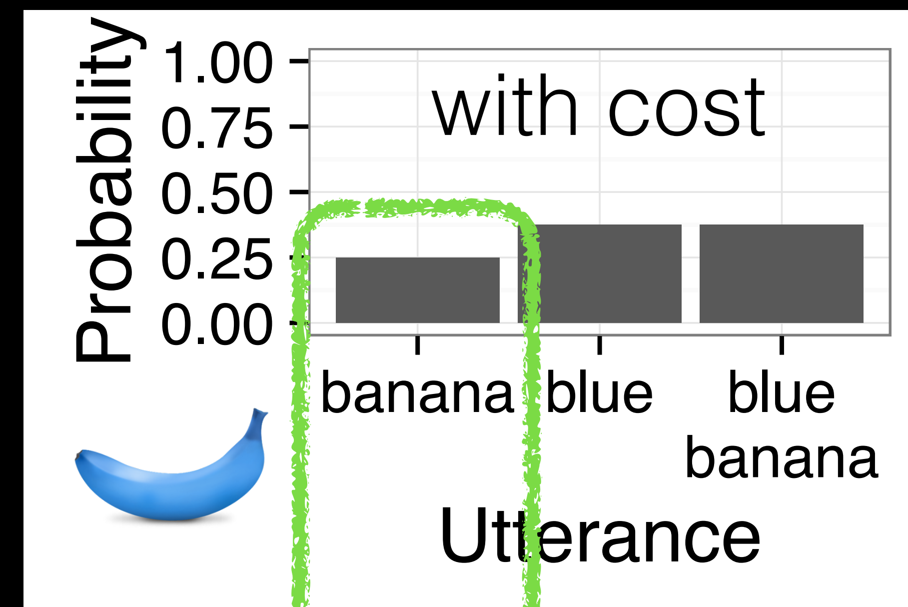
Predictions

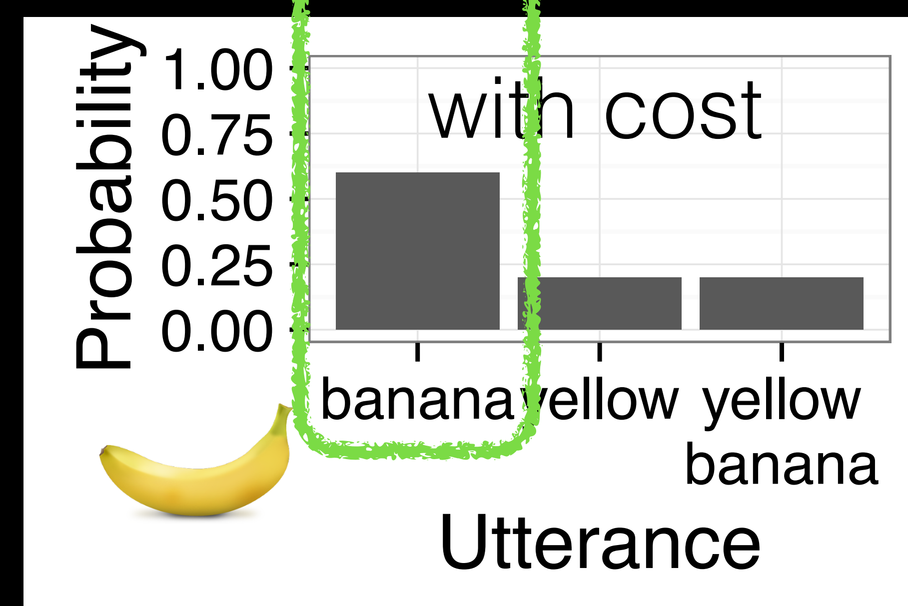
Literal listener





Pragmatic speaker





Redundancy more likely when probability of confusion is high

Independent empirical evidence for RSA with continuous semantics?

Literal listener

$$P_{L_0}(o|u) \propto [[u]](o)$$

$$[[u]](o) = \boxed{\text{typicality}(u, o)}$$

Pragmatic speaker

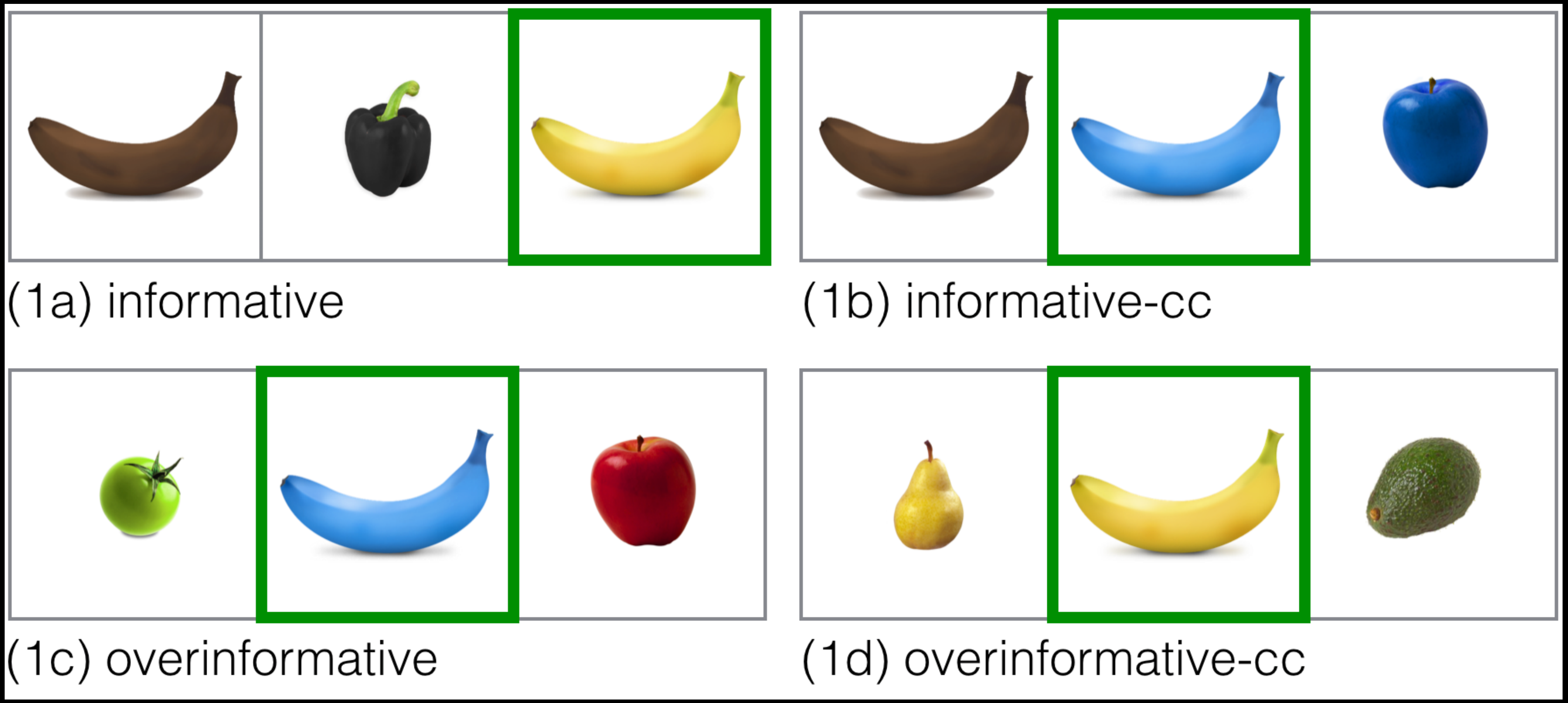
$$\boxed{P_{S_1}(u|o)} \propto e^{\lambda \ln P_{L_0}(o|u) - \text{cost}(u)}$$

1. Typicality norming
2. Production study
3. Model evaluation

Production study: interactive reference game experiment

Production study

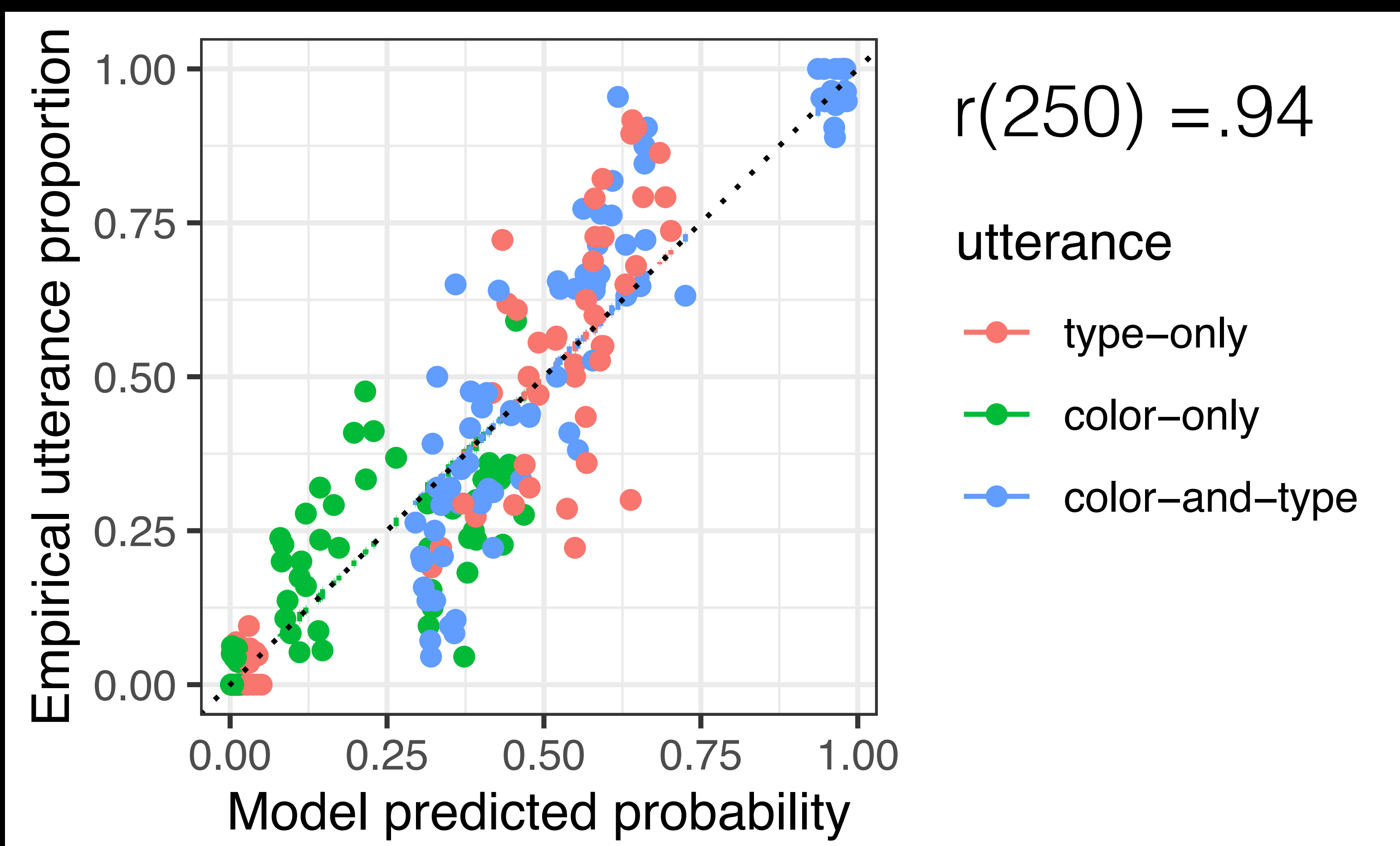
- 60 pairs of participants on Mechanical Turk
- random assignment to speaker/listener role
- 42 trials
- varied contextual informativeness of utterances:



presence of same type x presence of color competitor

Model evaluation

		Semantics	
		empirical	*fixed plus empirical*
	empirical	-1474.6 (4)	-1354.4 (7)
Cost	*fixed*	-1434.8 (4)	-1321.9 (7)
	none	-1372.9 (2)	**-1209.8 (5)**



A continuous-semantics RSA speaker captures variability in color modification by reasoning about a continuous semantics literal listener: the less inferable the intended referent is from the unmodified noun, the more likely modification is.

Final extension: cross-linguistic variability

Color overmodification less likely in Spanish (with post-nominal adjectives) than in English. Rubio-Fernández 2016

Strong argument for the role of incrementality.

Utterances	**Size-sufficient (SS) scene** o_{big_blue} o_{big_red} o_{small_blue} 	**Color-sufficient (CS) scene** o_{small_red} o_{big_red} o_{small_blue} 
English	*blue pin, red pin, big pin, small pin, big blue pin, big red pin,* ***small blue pin***	*blue pin, red pin, big pin, small pin, small red pin, big red pin,* ***small blue pin***
Spanish -postnom.	*pin blue, pin red, pin big, pin small, pin blue big, pin red big,* ***pin blue small***	*pin blue, pin red, pin big, pin small, pin red small, pin red big,* ***pin blue small***

Incremental RSA

Cohn-Gordon, Goodman, & Potts 2018

$$L_0^{INCR}(r|c,i) \propto \mathcal{X}^D(c,i,r) \cdot P(r)$$

$$\mathcal{X}^D(c,i,r) = \frac{|u{:}[[u]]^D(r)=1 \wedge u \text{ is a continuation of } c+i|}{|u{:}u \text{ is a continuation of } c+i|}$$

proportion of applicable continuations

$$S_1^{INCR}(i|c,r) \propto e^{\alpha(L_0^{INCR}(r|c,i) - C(i))}$$

$$S_1(u|r) = \prod_{j=1}^{n} S_1^{INCR}(i_j|c = [i_1 \ldots i_{j-1}], r)$$

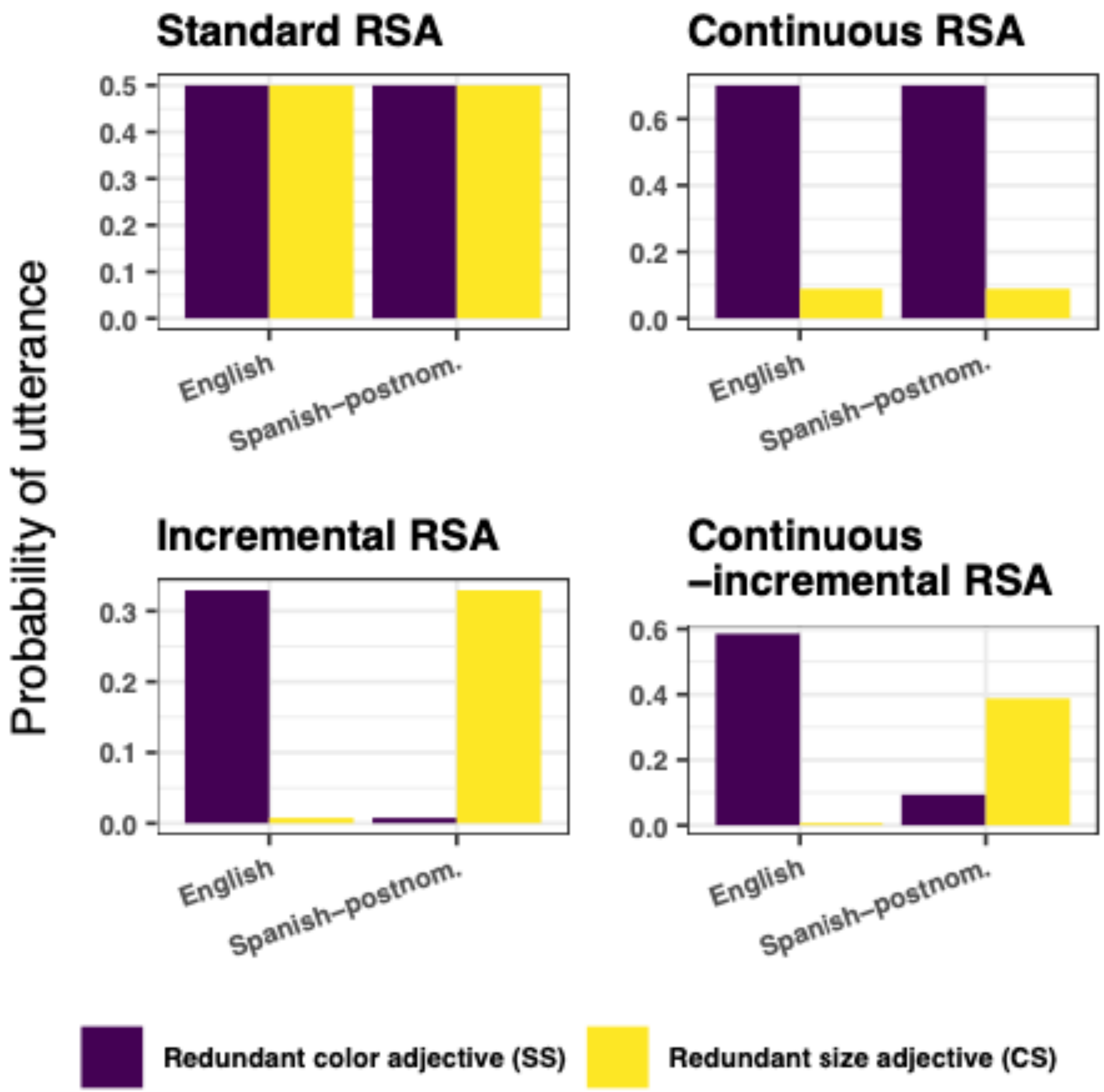
Continuous-Incremental RSA

Waldon & Degen 2021

$$\mathcal{X}^C(c,i,r) = \frac{\sum [[u]]^C(r){:}u \text{ is a continuation of } c+i}{|u{:}u \text{ is a continuation of } c+i|}$$

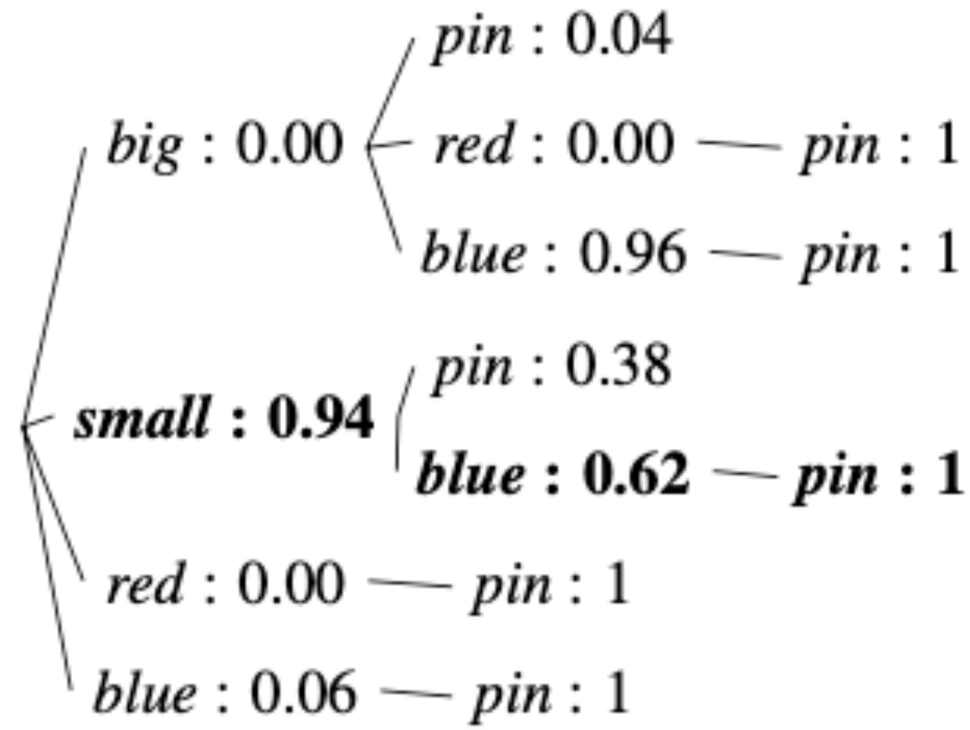
sum of semantic values over number of continuations

Model predictions



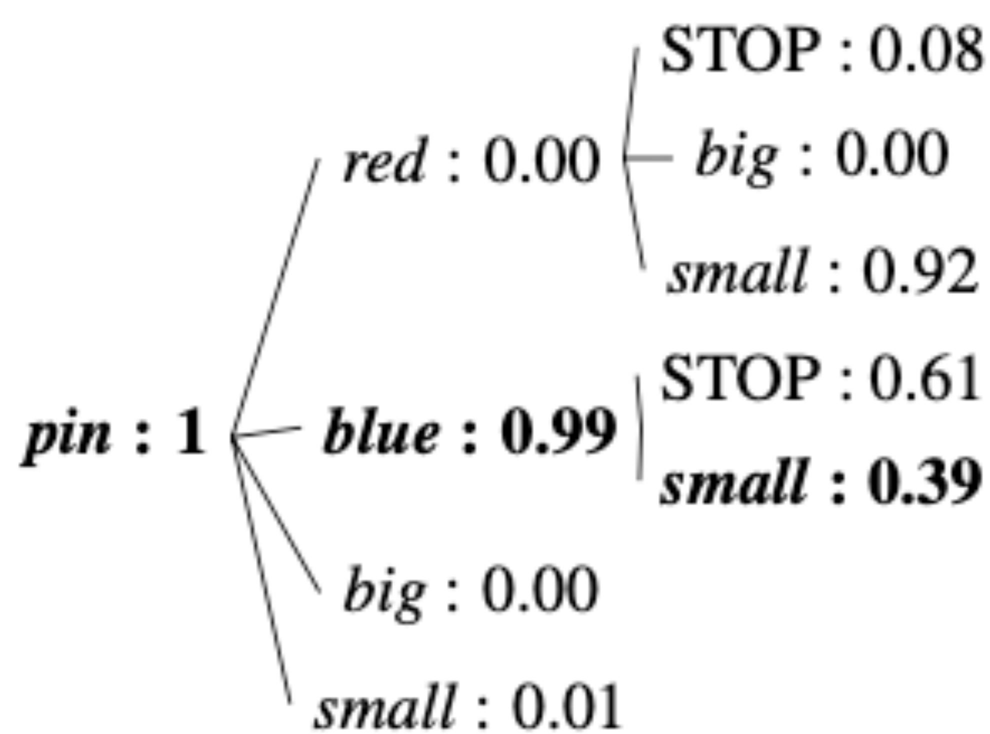
Continuous-Incremental RSA

English size-sufficient



Analogous contexts (symmetric in Incremental RSA)

Spanish color-sufficient



Continuous-Incremental RSA

Combining incremental and continuous RSA

- provides some support for Rubio-Fernández's claim that modification is generally less useful post-nominally

- makes interesting novel prediction for flipped color/size overmodification asymmetry in post-nominal adjective languages

Much more empirical work needed!

Reference comprehension and production

Comprehension:

Listeners draw inferences based on the expectation that speakers not be over-informative.

Production:

Speakers produce seemingly overinformative referring expressions.

Quantity-2

Quantity-2

HOW TO RESOLVE?

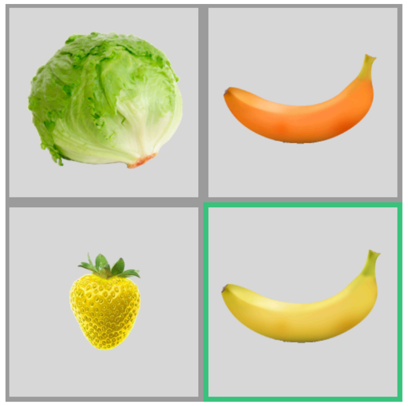
Reference comprehension and production

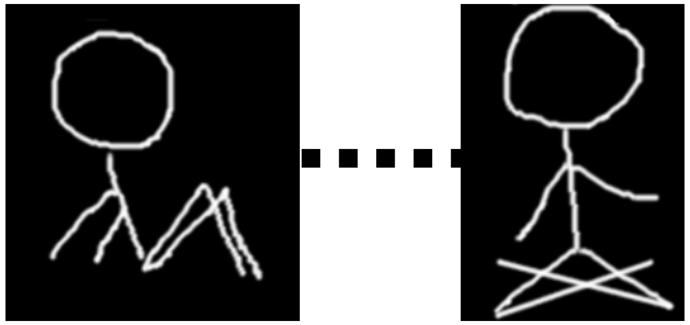
Comprehension:

Listeners draw inferences based on the expectation that speakers be **sufficiently contextually informative**.

Production:

Speakers produce **usefully redundant** referring expressions.





RSA: useful modeling framework to formalize the link between production and comprehension via probabilistic inference.

http://www.problang.org

Probabilistic language understanding

An introduction to the Rational Speech Act framework

By Gregory Scontras, Michael Henry Tessler, and Michael Franke

The present course serves as a practical introduction to the Rational Speech Act modeling framework. Little is presupposed beyond a willingness to explore recent progress in formal, implementable models of language understanding.

Main content

The literal listener rule can be written as follows:

```
// set of states (here: objects of reference)
// we represent objects as JavaScript objects to demarcate them from utterances
// internally we treat objects as strings nonetheless
var objects = [{color: "blue", shape: "square", string: "blue square"},
               {color: "blue", shape: "circle", string: "blue circle"},
               {color: "green", shape: "square", string: "green square"}]

// set of utterances
var utterances = ["blue", "green", "square", "circle"]

// prior over world states
var objectPrior = function() {
  var obj = uniformDraw(objects)
  return obj.string
}

// meaning function to interpret the utterances
var meaning = function(utterance, obj){
  _.includes(obj, utterance)
}

// literal listener
var literalListener = function(utterance){
  Infer({model: function(){
    var obj = objectPrior();
    var uttTruthVal = meaning(utterance, obj);
    condition(uttTruthVal == true)
    return obj
  }})
}

viz.table(literalListener("blue"))
```

run

(state)	probability
blue circle	0.5
blue square	0.5

Exercises:

1. In the model above, `objectPrior()` returns a sample from a `uniformDraw` over the possible objects of reference. What happens when the listener's beliefs are not uniform over the

Scontras, Tessler, & Franke

forestdb.org/models/overinf.html

Thank you!