GENERATION OF EFFECTIVE INSTRUCTIONS IN SITUATED DIALOGUE

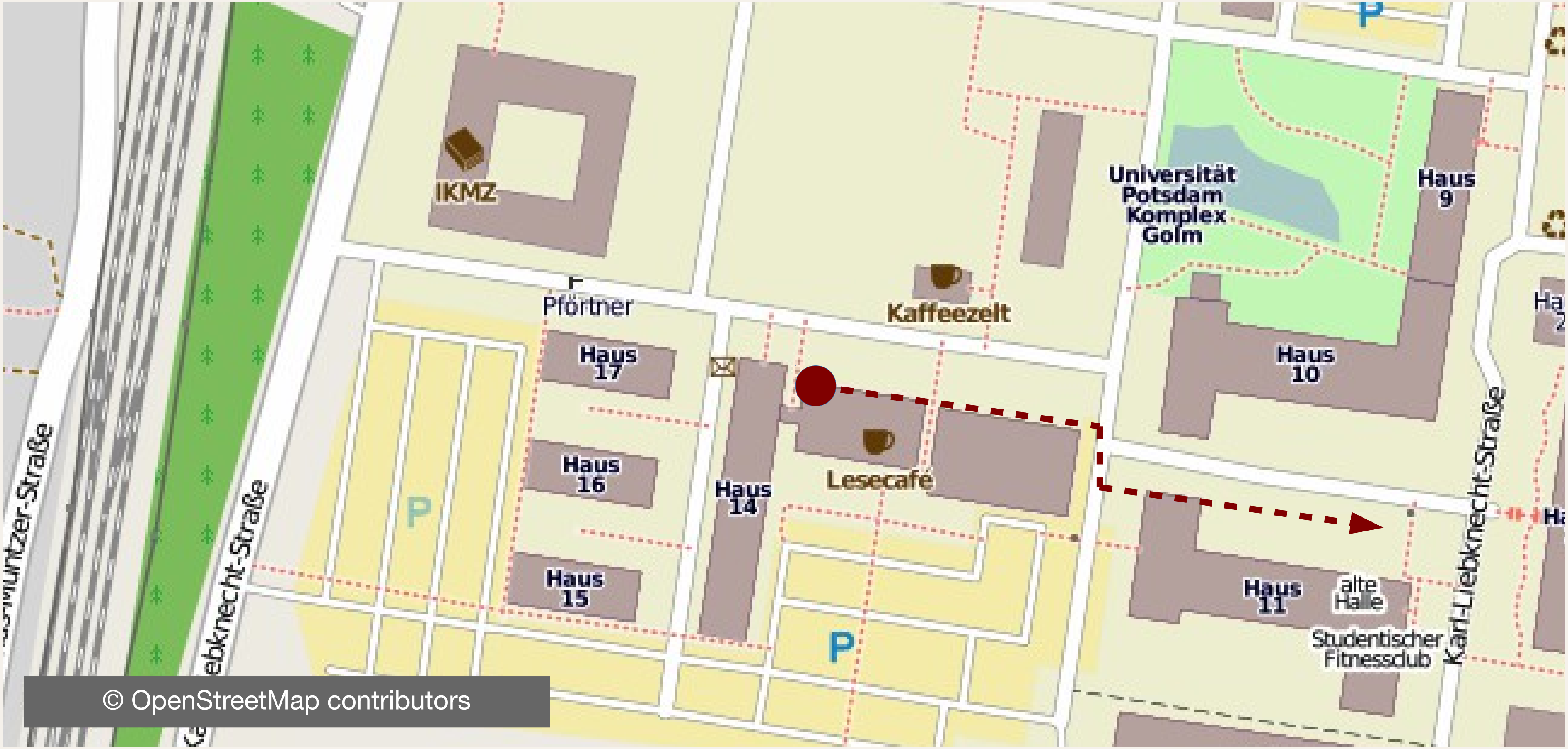
NIKOS ENGONOPOULOS
MARTIN VILLALBA
ALEXANDER KOLLER

SFB 632 - D6

What is **Pedestrian navigation** ?

(and why does it matter?)

Picture by Andrei Pop, via Flickr



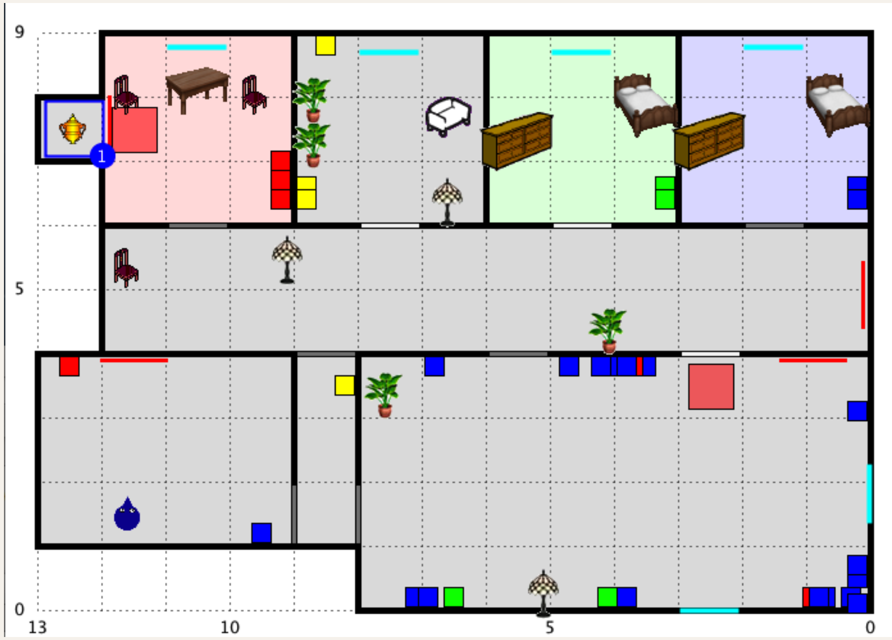
"See a big building to your right? Walk past it and then turn right... No, not that one, I meant the one that's like a *greenhouse*, it has some plants inside. Yeah that one. Now go left and then straight until I tell you."

REFERRING EXPRESSIONS

A NOUN PHRASE THAT IDENTIFIES UNIQUELY A CERTAIN OBJECT WITHIN A SCENE

METHODOLOGY: The GIVE Challenge

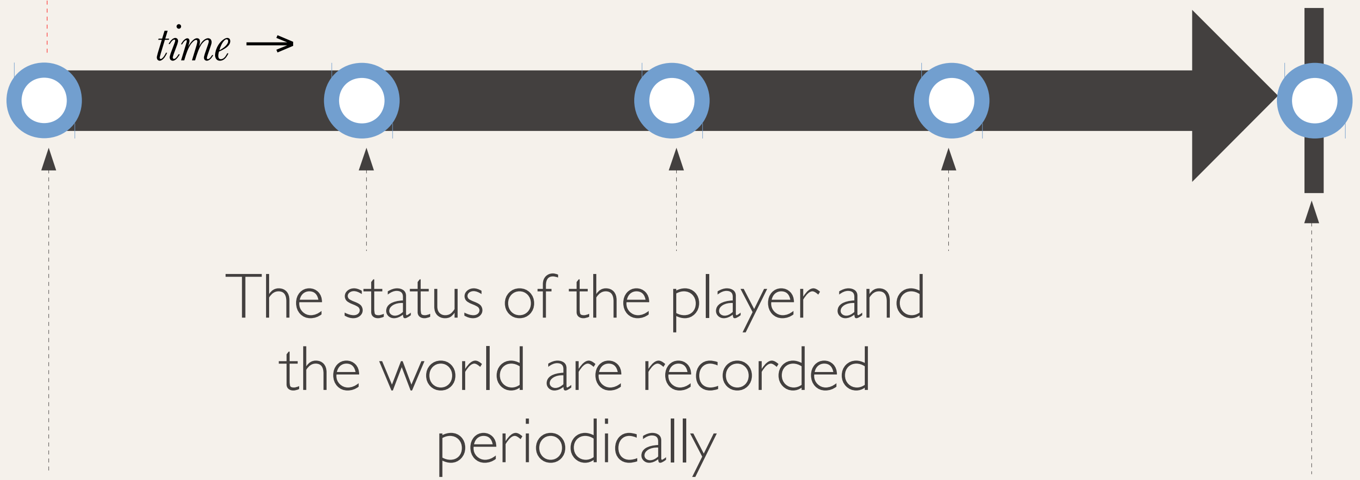
GENERATING INSTRUCTIONS IN VIRTUAL ENVIRONMENTS



Help a human player solve a puzzle through automatically generated, real-time instructions

EPISODES

One instance of recorded behavior is called an episode.



A RE is presented to the player

The player clicks a button

The status of the player and the world are recorded periodically

PROBABILISTIC FRAMEWORK

We want our instructions to have
a high degree of success.
For that, we need to maximize this probability

$$p(a \mid r, s, \sigma)$$

TARGET

BEHAVIOR

STATE OF THE WORLD

REFERRING EXPRESSION

PROBABILISTIC FRAMEWORK

We'll split this into two models:

$$p(a \mid r, s, \sigma) \propto \underbrace{p(a \mid r, s)}_{\text{SEMANTIC MODEL (Psem)}} \; \underbrace{p(a \mid \sigma)}_{\text{OBSERVATIONAL MODEL (Pobs)}}$$

The Psem model tells us which RE has a higher chance of success

The Pobs model tells us when we need to give you a new RE

LOG-LINEAR MODELS

Both models are log-linear,
because they are written in this form:

$$p(a \mid r, s) \propto \exp(w_1 f_1(a, r, s) + \ldots + w_n f_n(a, r, s))$$

f_i are called FEATURE FUNCTIONS

w_i are the associated WEIGHTS

We select the features, but the weights
are learned from the training data

SEMANTIC MODEL

EXAMPLE FEATURES FOR Psem

SEMANTIC FEATURES

Is the color of the item mentioned in the RE?
Is the relative position of an item mentioned in the RE?

CONFUSION FEATURES

Is the color of another item mentioned in the instruction?

SALIENCE FEATURES

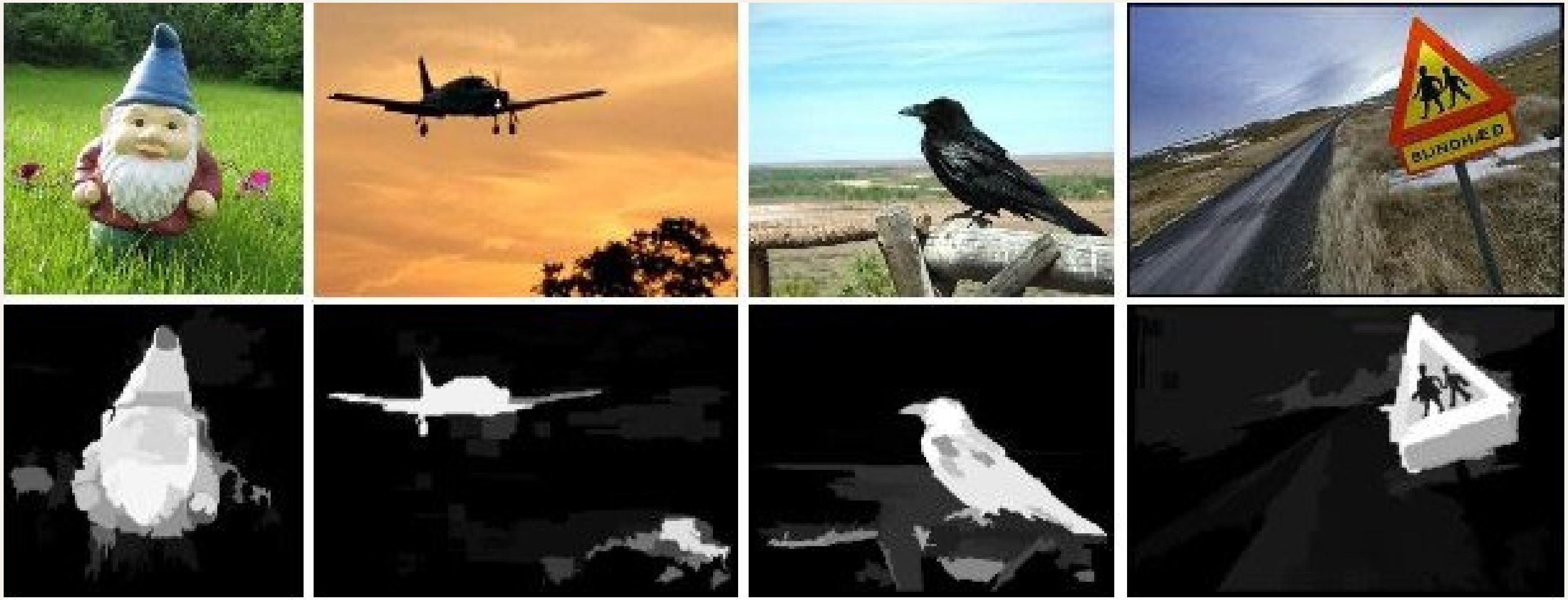
Is an item visible? Is it in the room?
How visually salient is it?

SEMANTIC MODEL

VISUAL SALIENCY

VISUAL SALIENCE

A weighted measure of centrality and size
for a target in a visual field



Ming-Ming Chen et al, IEEE TPAMI 2014

OBSERVATIONAL MODEL

EXAMPLE FEATURES FOR Pobs

Has the user remained still in the last seconds?

(might indicate confusion)

How much has the angle to an object changed?

(might indicate (dis)interest)

How much closer has the player moved towards an object? Has he entered the same room?

How has the visual salience of an object evolved?

(might indicate a loss of interest)

RESULTS

Training and testing were performed using recorded interactions between players and systems

Training data was obtained from the GIVE-2.5 Challenge

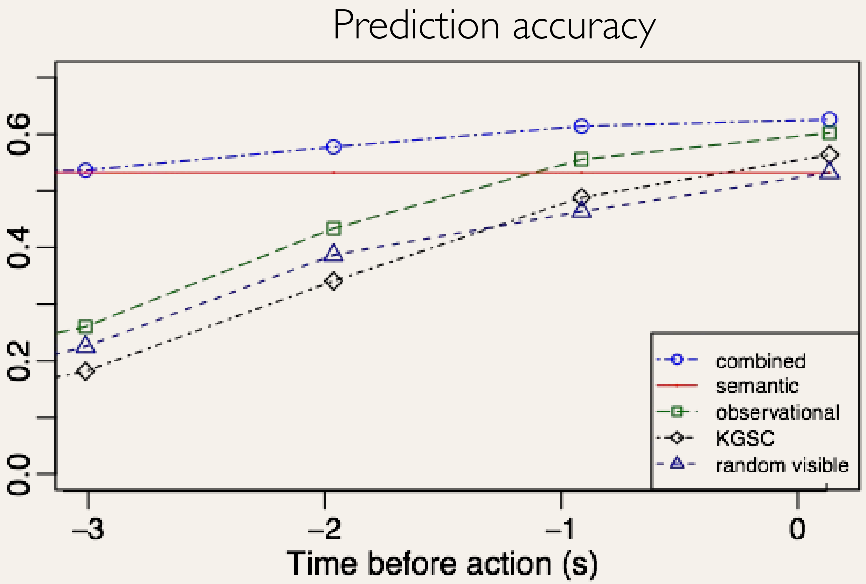
Test data was obtained from the GIVE-2 Challenge

RESULTS

The combined model outperforms both individual models

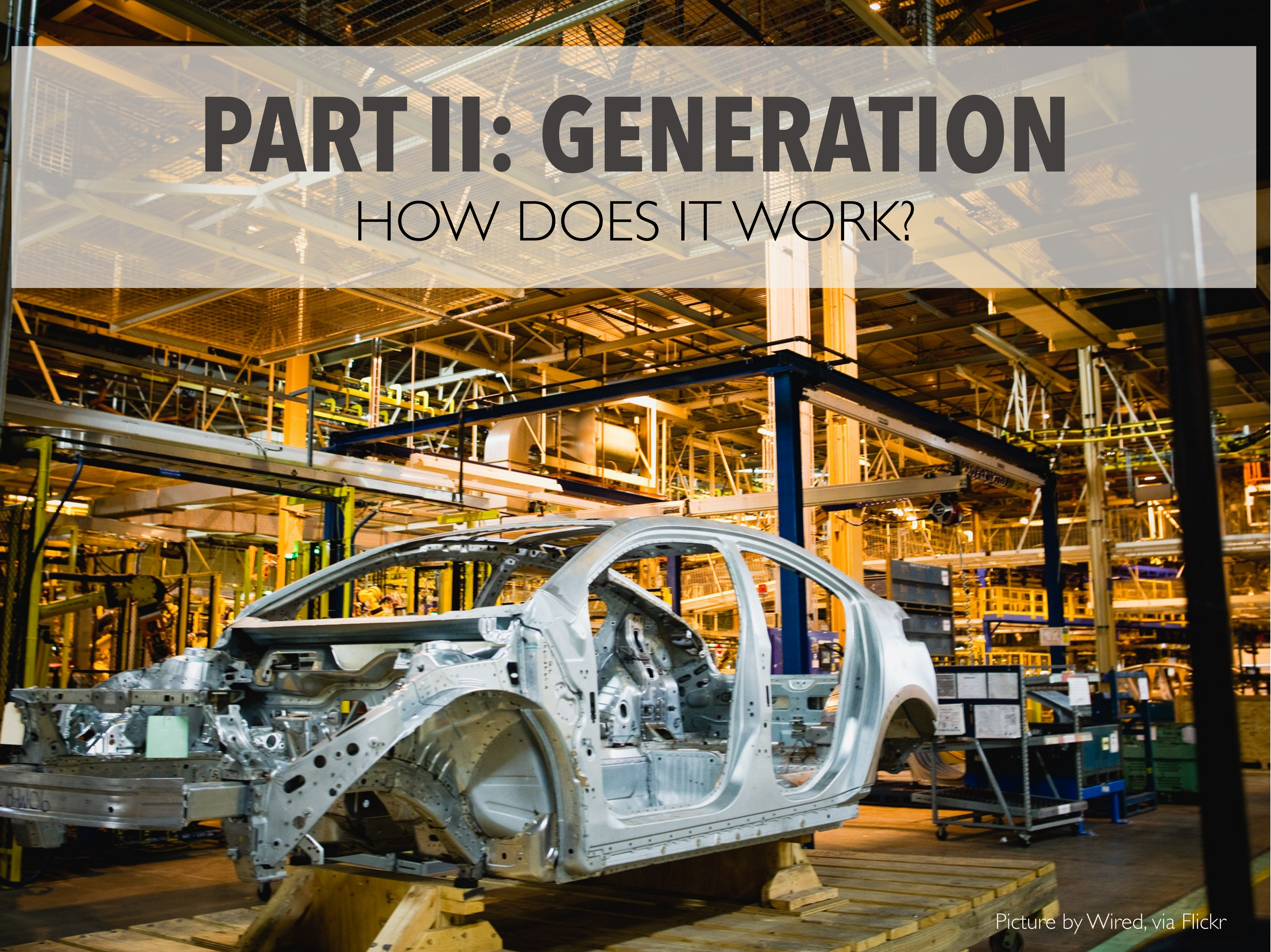
The Psem model outperforms Pobs and the baseline early on

The Pobs model improves late accuracy



PART II: GENERATION

HOW DOES IT WORK?



Picture by Wired, via Flickr

IRTG

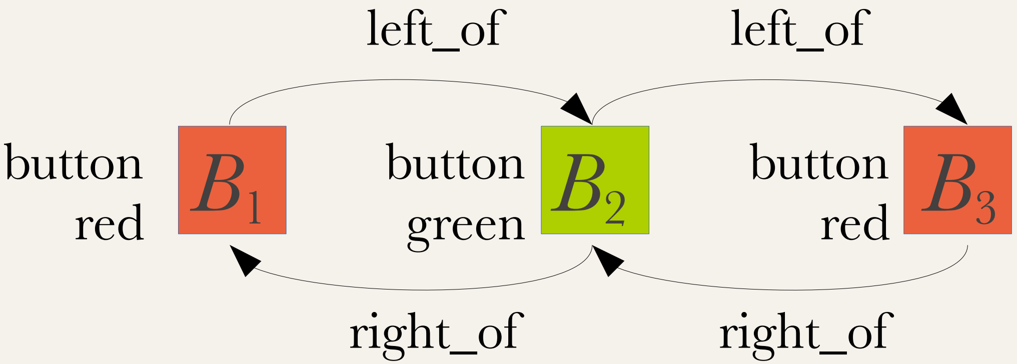
INTERPRETED REGULAR TREE GRAMMAR

GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · *w1*	$\text{uniq}(R_1) = \text{if } (R_1 \text{ is singleton}) \text{ then } R_1 \text{ else } \varnothing$
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{ a \in R_1 \mid \textit{exists } b \in R_2 \text{ s.t. } (a,b) \in \lvert\text{left\_of}\rvert \}$
N → green(N)	green · *w1*	$\lvert \textit{green} \rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert \textit{red} \rvert \cap R_1$
N → button	button	$\lvert\text{button}\rvert$

IRTG

INTERPRETED REGULAR TREE GRAMMAR

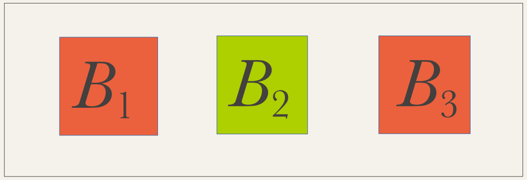
GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · $w1$	uniq(R_1) = if (R_1 is singleton) then R_1 else ∅
N → leftof(N, NP)	$w1$ · to the left of · $w2$	$\{\, a \in R_1 \mid \textit{exists } b \in R_2 \text{ s.t. } (a,b) \in \lvert\text{left\_of}\rvert \,\}$
N → green(N)	green · $w1$	$\lvert \textit{green} \rvert \cap R_1$
N → red(N)	red · $w1$	$\lvert \textit{red} \rvert \cap R_1$
N → button	button	$\lvert\text{button}\rvert$

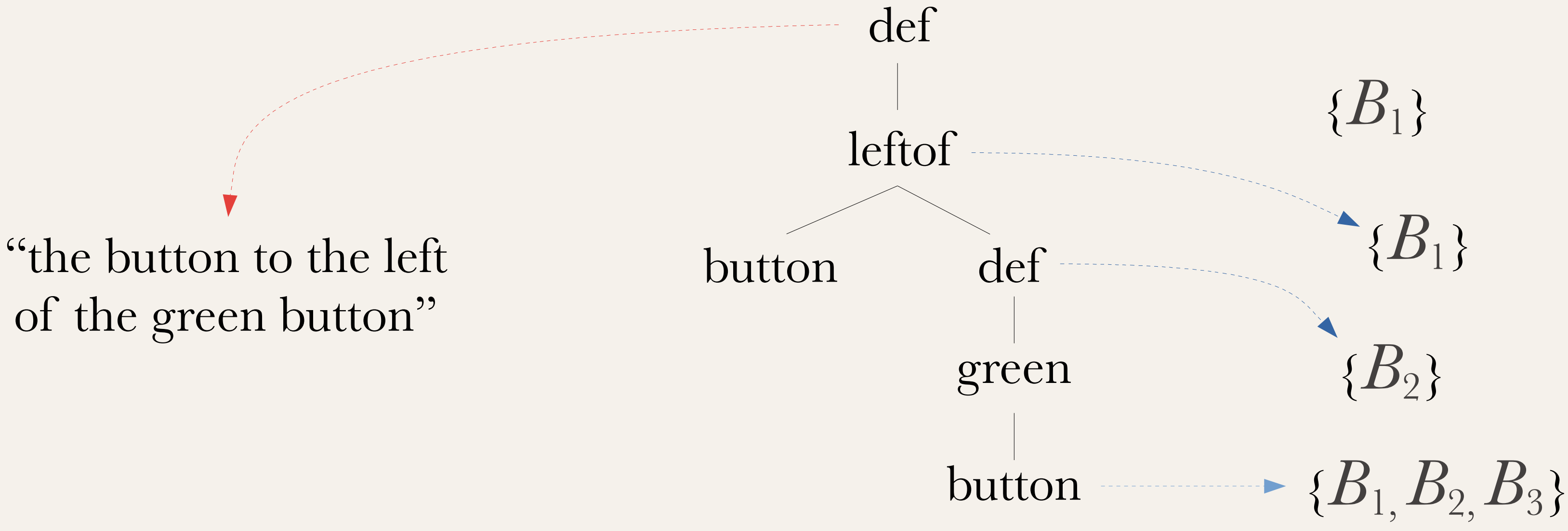


IRTG

INTERPRETED REGULAR TREE GRAMMAR

GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · $w1$	uniq(R_1) = if (R_1 is singleton) then R_1 else ∅
N → leftof(N, NP)	$w1$ · to the left of · $w2$	$\{\ a \in R_1 \ \vert\ \textit{exists}\ b \in R_2$ s.t. (a,b) ∈ $\vert$left_of$\vert$ }
N → green(N)	green · $w1$	$\vert \textit{green} \vert \cap R_1$
N → red(N)	red · $w1$	$\vert \textit{red} \vert \cap R_1$
N → button	button	$\vert$button$\vert$

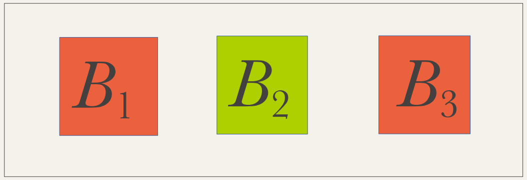




IRTG

CHART-BASED GENERATION

GRAMMAR RULE	STRING	DENOTATION
NP → def(N)	the · *w1*	uniq(R_1) = if (R_1 is singleton) then R_1 else $\varnothing$
N → leftof(N, NP)	*w1* · to the left of · *w2*	$\{ a \in R_1 \mid \textit{exists } b \in R_2 \text{ s.t. } (a,b) \in \lvert \text{left\_of} \rvert \}$
N → green(N)	green · *w1*	$\lvert \textit{green} \rvert \cap R_1$
N → red(N)	red · *w1*	$\lvert \textit{red} \rvert \cap R_1$
N → button	button	$\lvert \text{button} \rvert$



$NP/\{b_1\} \rightarrow \text{def}(N/\{b_1\})$

$NP/\{b_2\} \rightarrow \text{def}(N/\{b_2\})$

$N/\{b_1\} \rightarrow \text{leftof}(N/\{b_1, b_2, b_3\}, NP/\{b_2\})$

$N/\{b_1, b_3\} \rightarrow \text{red}(N/\{b_1, b_2, b_3\})$

$N/\{b_2\} \rightarrow \text{green}(N/\{b_1, b_2, b_3\})$

$N/\{b_1, b_2, b_3\} \rightarrow \text{button}$

…

IRTG

CHART-BASED GENERATION

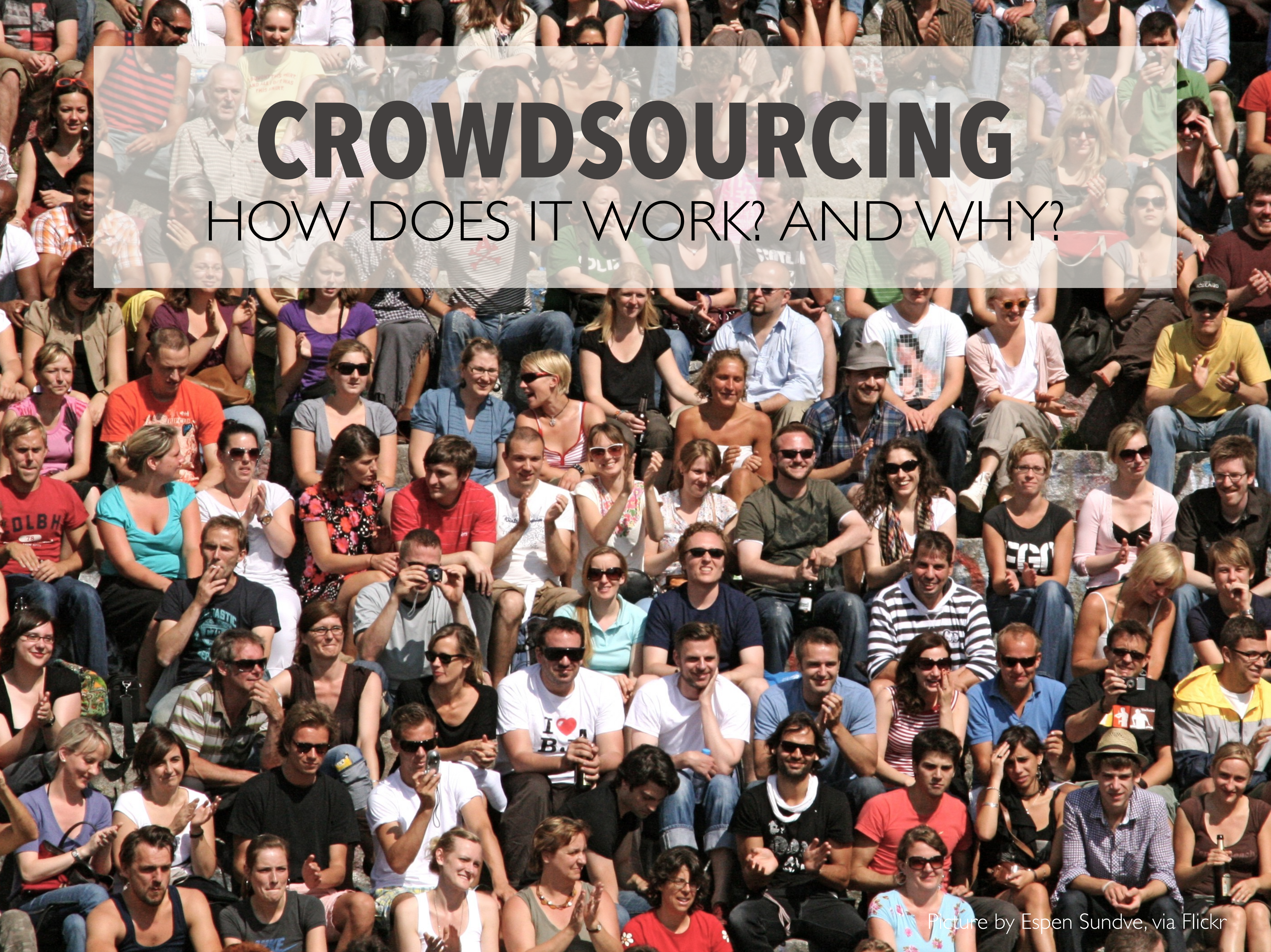
A chart can tell us how to generate all possible REs

Picking the best one is tricky

See (Engonopoulos & Koller 2014) for more details

CROWDSOURCING

HOW DOES IT WORK? AND WHY?



Picture by Espen Sundve, via Flickr

CROWDSOURCING

SOME STATISTICS

Accessible cost

“[Crowdsourcing] is, in short, extremely inexpensive relative to nearly every alternative other than uncompensated students” (Berinsky et al., 2012)

Estimated expected pay: $1/10min

CROWDSOURCING

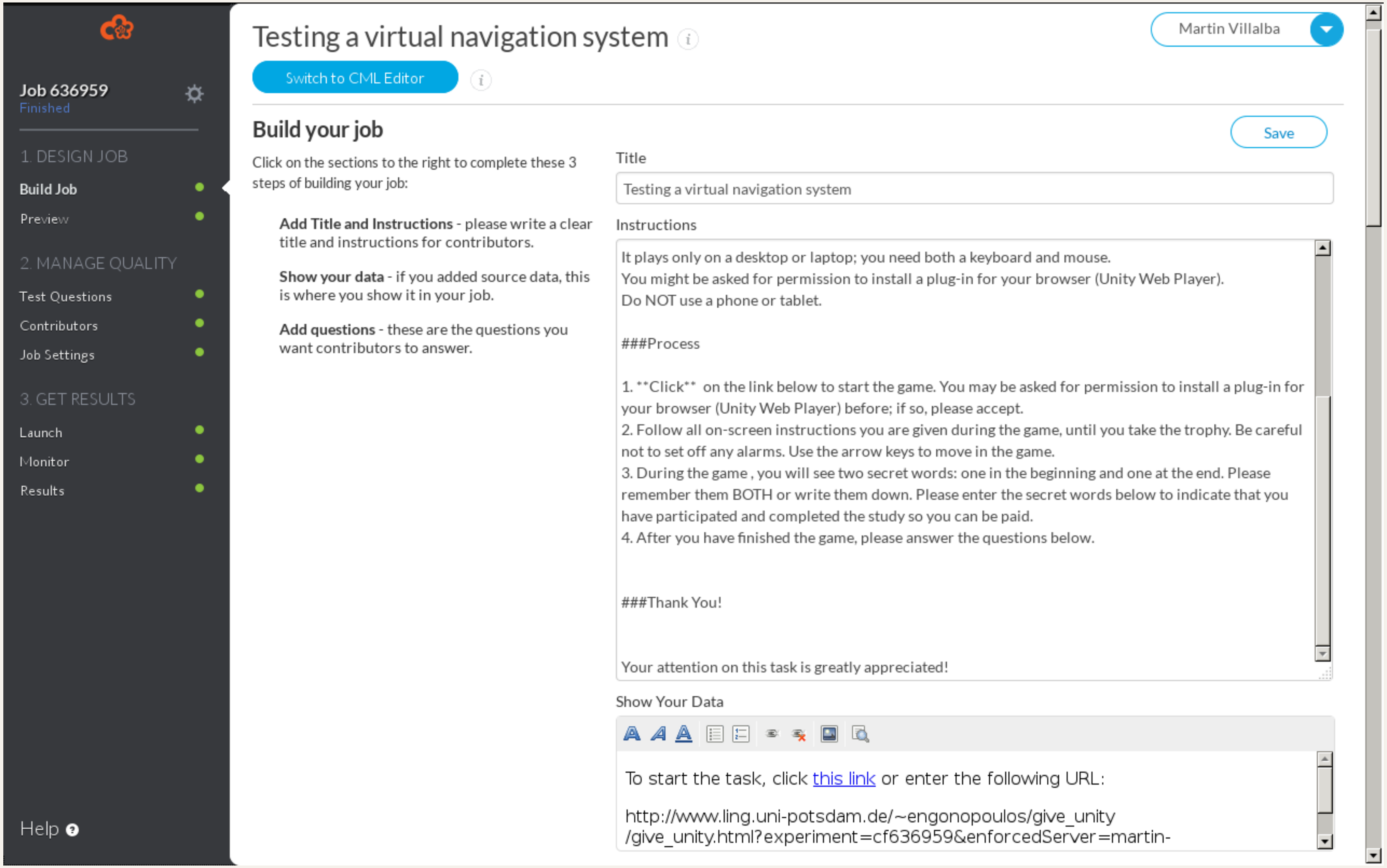
OUR EXPERIENCE

CrowdFlower

Available in Europe

Waived fee for educational purposes

CROWDSOURCING
SETTING UP OUR EXPERIMENT



Build your job

Click on the sections to the right to complete these 3 steps of building your job:

Add Title and Instructions - please write a clear title and instructions for contributors.

Show your data - if you added source data, this is where you show it in your job.

Add questions - these are the questions you want contributors to answer.

Title: Testing a virtual navigation system

Instructions:

It plays only on a desktop or laptop; you need both a keyboard and mouse.
You might be asked for permission to install a plug-in for your browser (Unity Web Player).
Do NOT use a phone or tablet.

###Process

1. **Click** on the link below to start the game. You may be asked for permission to install a plug-in for your browser (Unity Web Player) before; if so, please accept.
2. Follow all on-screen instructions you are given during the game, until you take the trophy. Be careful not to set off any alarms. Use the arrow keys to move in the game.
3. During the game , you will see two secret words: one in the beginning and one at the end. Please remember them BOTH or write them down. Please enter the secret words below to indicate that you have participated and completed the study so you can be paid.
4. After you have finished the game, please answer the questions below.

###Thank You!

Your attention on this task is greatly appreciated!

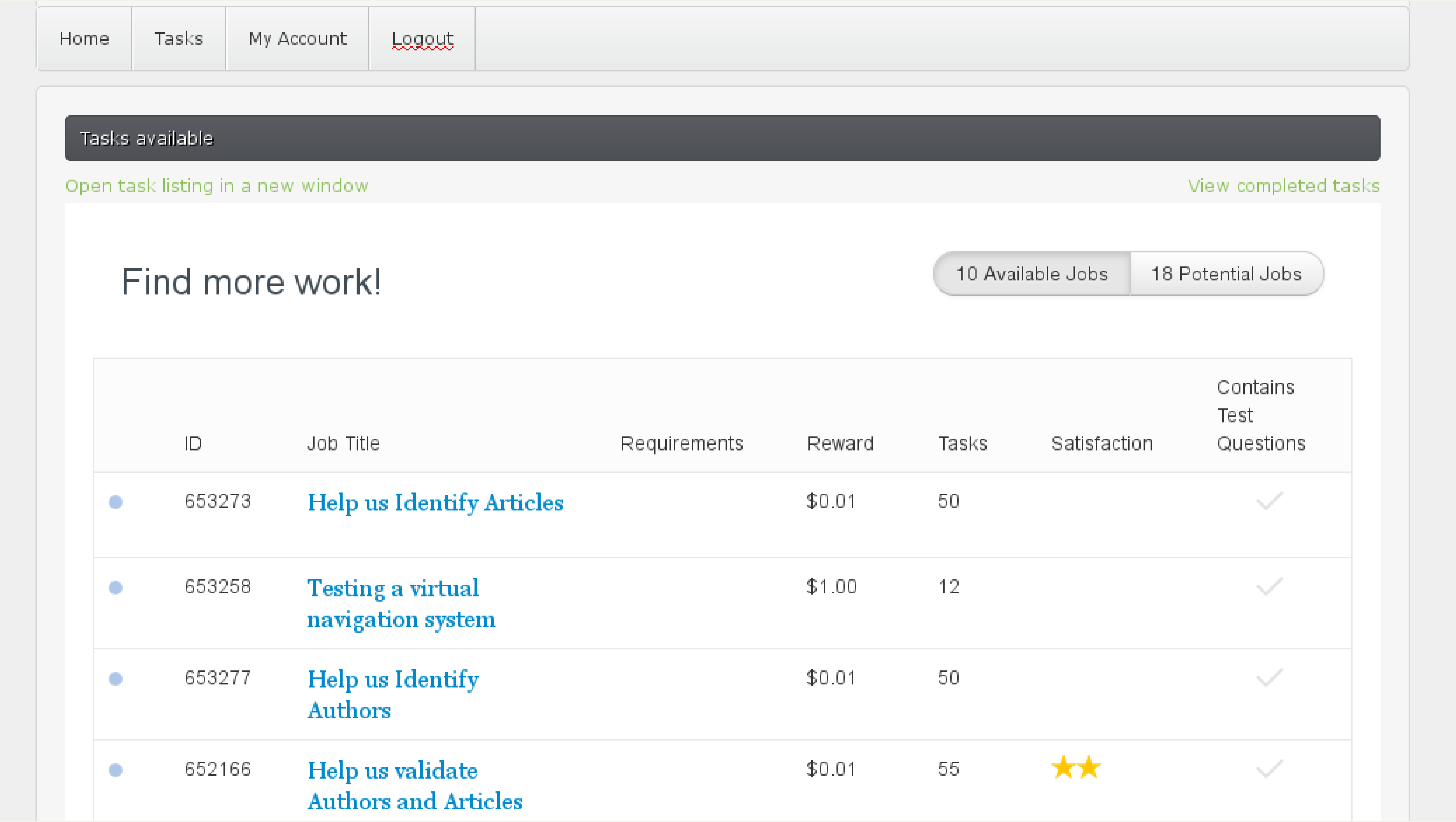
Show Your Data:

To start the task, click this link or enter the following URL:

http://www.ling.uni-potsdam.de/~engonopoulos/give_unity/give_unity.html?experiment=cf636959&enforcedServer=martin-

CROWDSOURCING

SETTING UP OUR EXPERIMENT

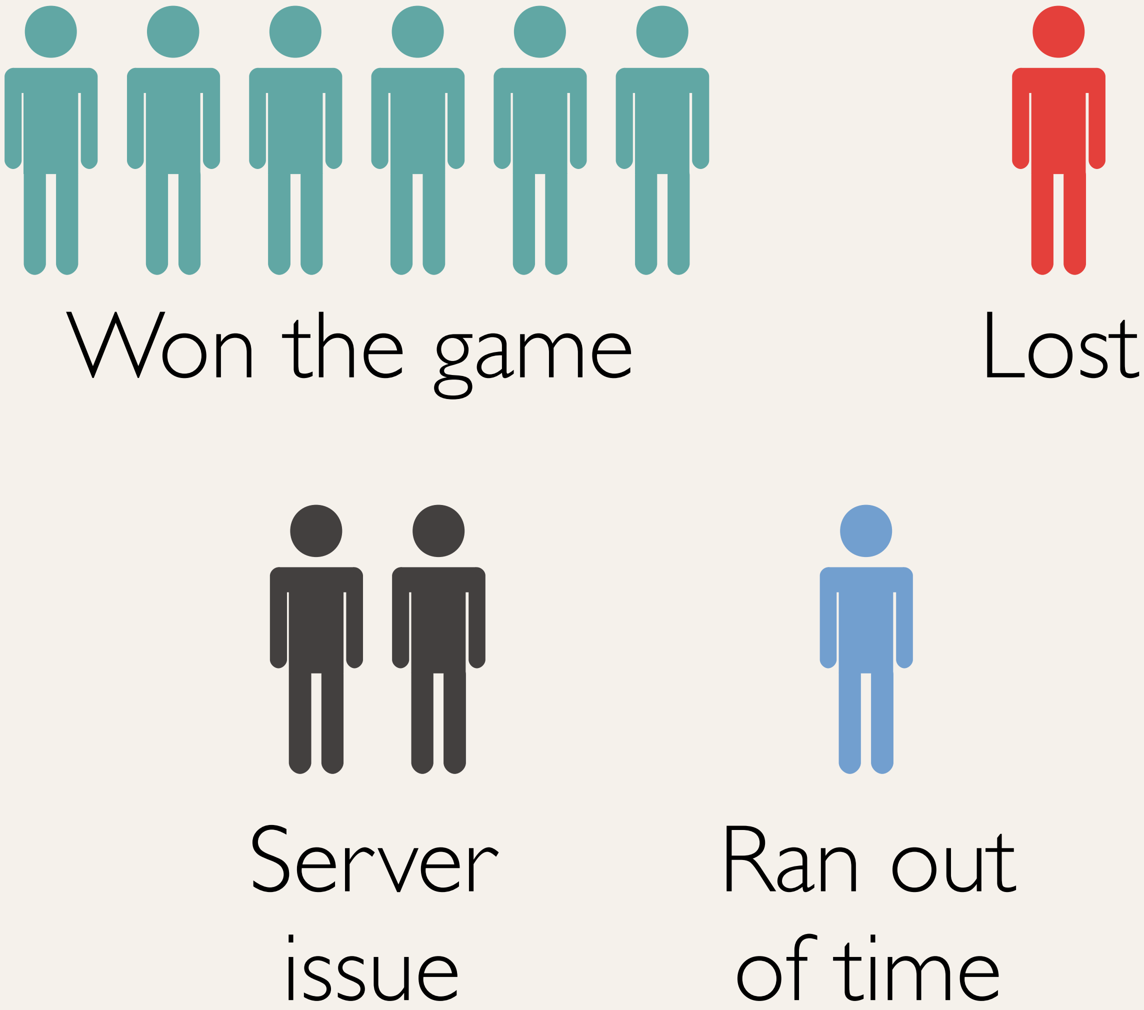


Find more work!

	ID	Job Title	Requirements	Reward	Tasks	Satisfaction	Contains Test Questions
●	653273	Help us Identify Articles		$0.01	50		✓
●	653258	Testing a virtual navigation system		$1.00	12		✓
●	653277	Help us Identify Authors		$0.01	50		✓
●	652166	Help us validate Authors and Articles		$0.01	55	★★	✓

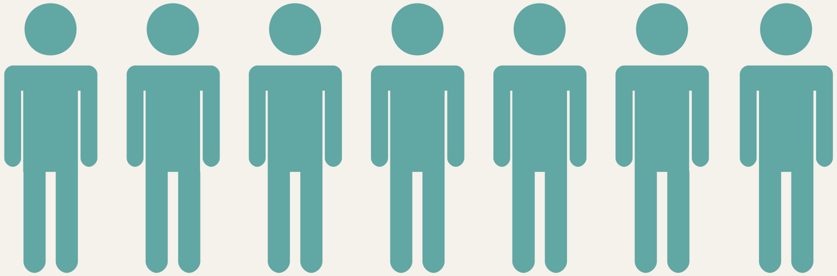
CROWDSOURCING

SOME RESULTS – 1:15H TASK

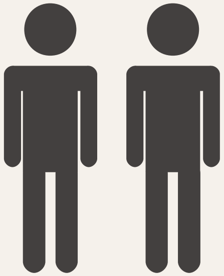


CROWDSOURCING

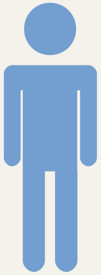
SOME RESULTS – 1:15H TASK



Won the game



Server issue



Ran out of time

CROWDSOURCING

OTHER ISSUES

We have to keep cheaters in mind

Incentives are effective,
but tricky to get right

FUTURE WORK

WHERE DO WE GO FROM HERE?

Picture by Jen Scheer, via Flickr

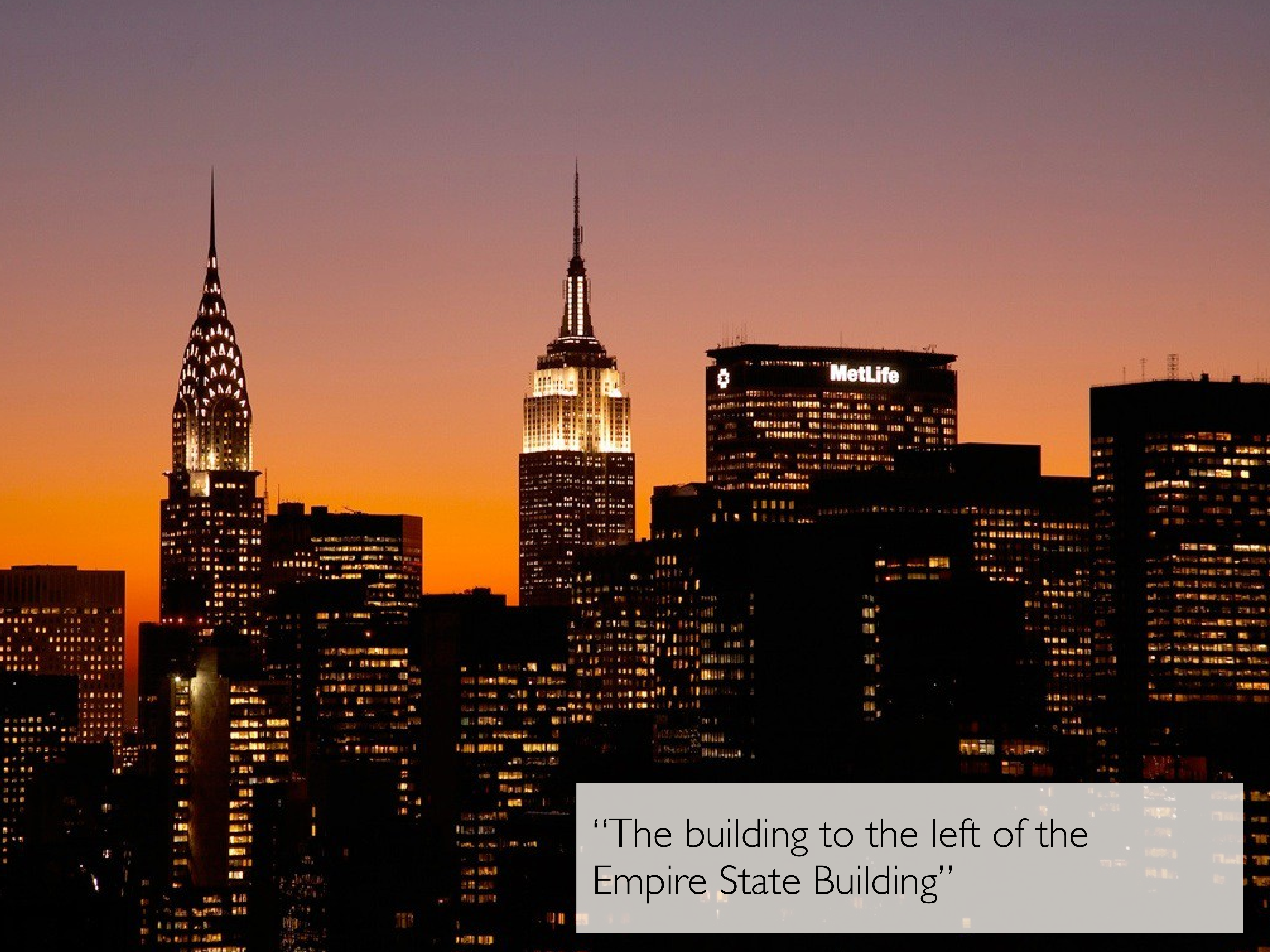
FUTURE WORK

RESTRICTED CONTEXT SET

We defined a referring expression as

A NOUN PHRASE THAT IDENTIFIES UNIQUELY A CERTAIN OBJECT WITHIN A SCENE

We rarely make those



“The building to the left of the Empire State Building”

FUTURE WORK

RESTRICTED CONTEXT SET

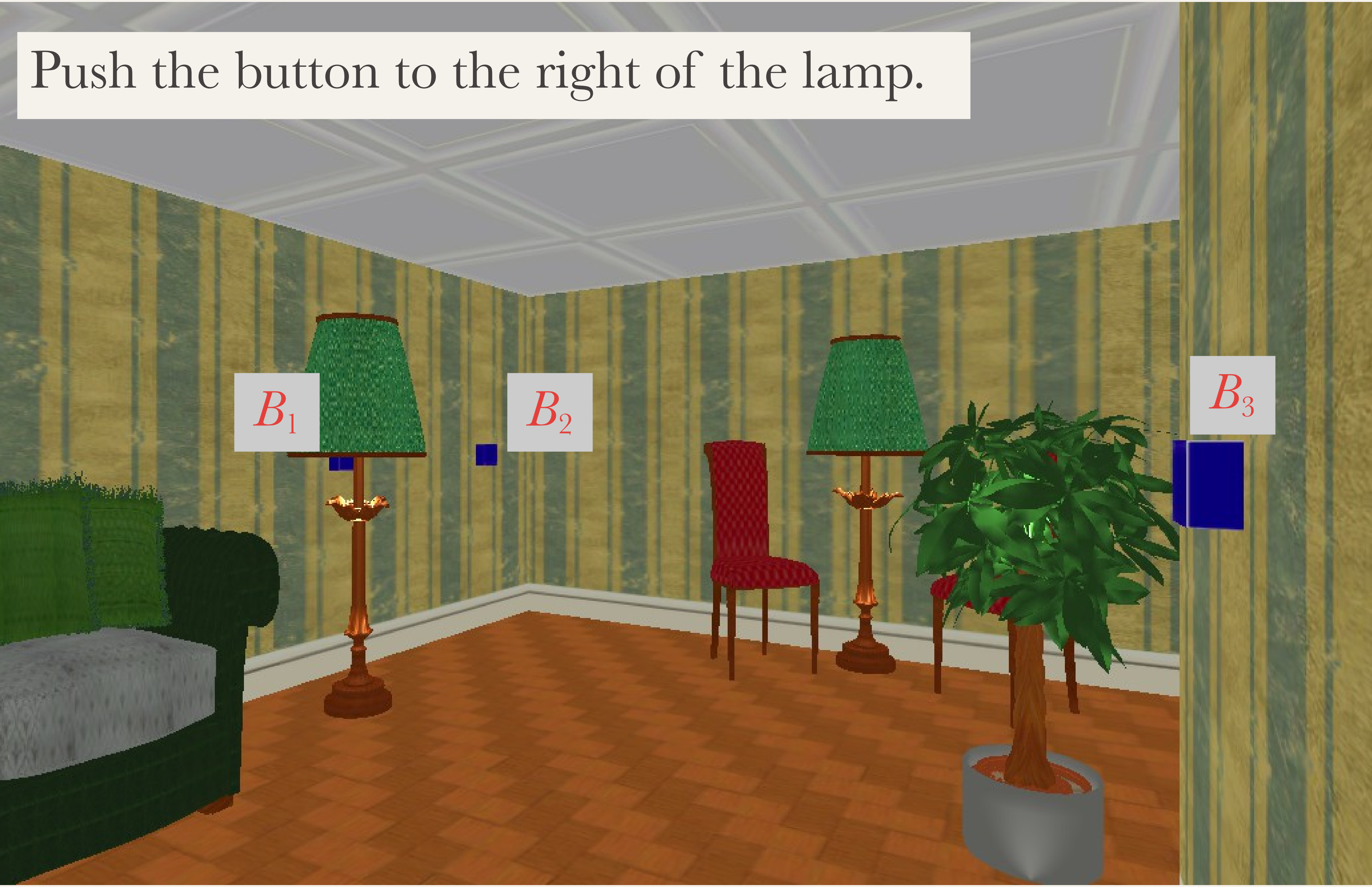
We say the viable candidates for an interpretation process are part of the context set

But how do we know which targets should be part of it?

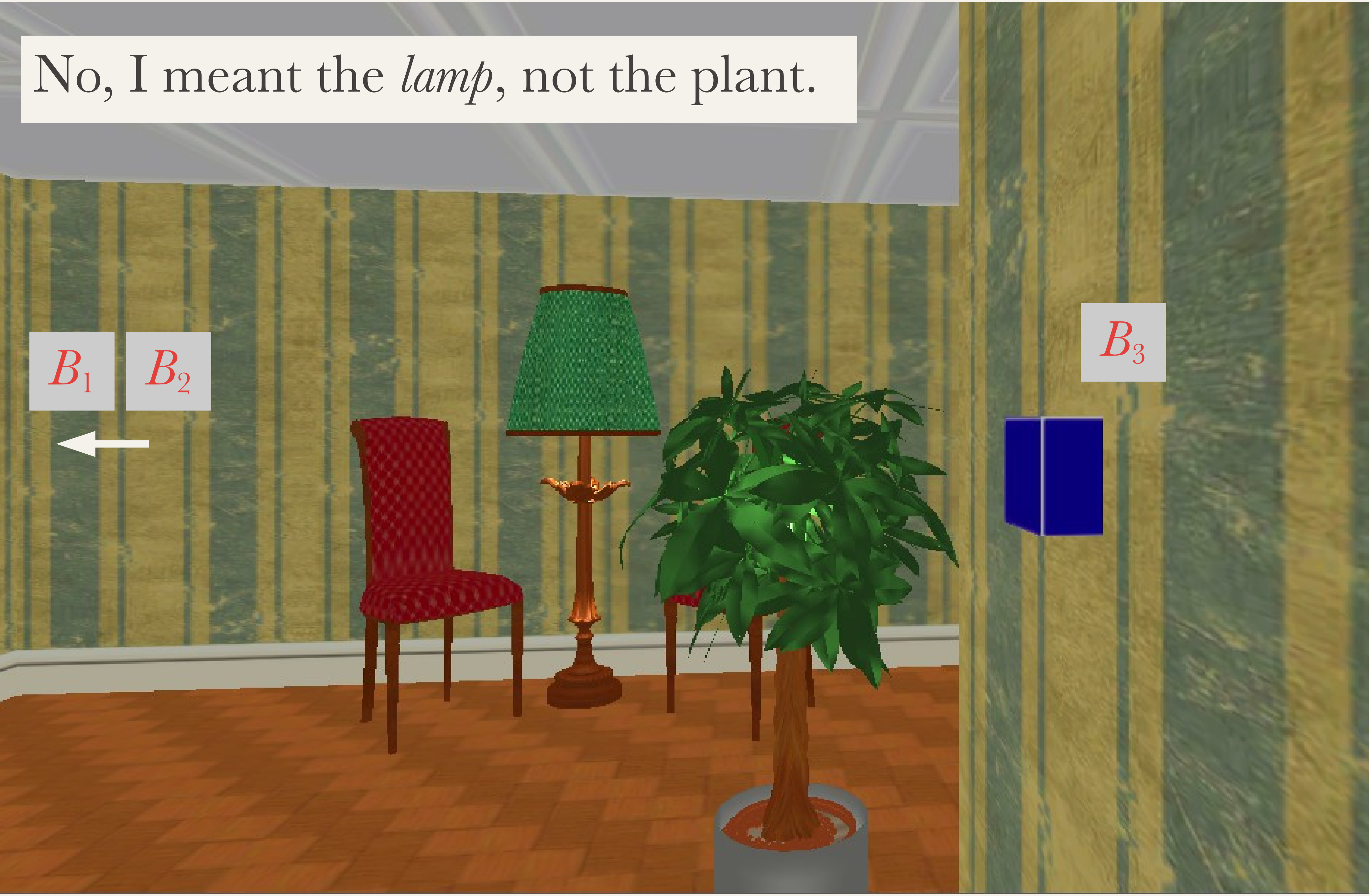
FUTURE WORK

CONTRASTIVE REs

Contrastive REs are vital to keep users from making (possibly costly) mistakes



Push the button to the right of the lamp.



No, I meant the *lamp*, not the plant.

FUTURE WORK

CONTRASTIVE REs

We have structured information,
but we don't have the right structures.

Which strategies should we look into?

QUESTIONS?

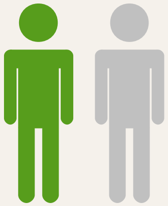
THANK YOU FOR YOUR ATTENTION

CROWDSOURCING

WORKERS' DEMOGRAPHICS

1 in 4 have a Bachelor's degree

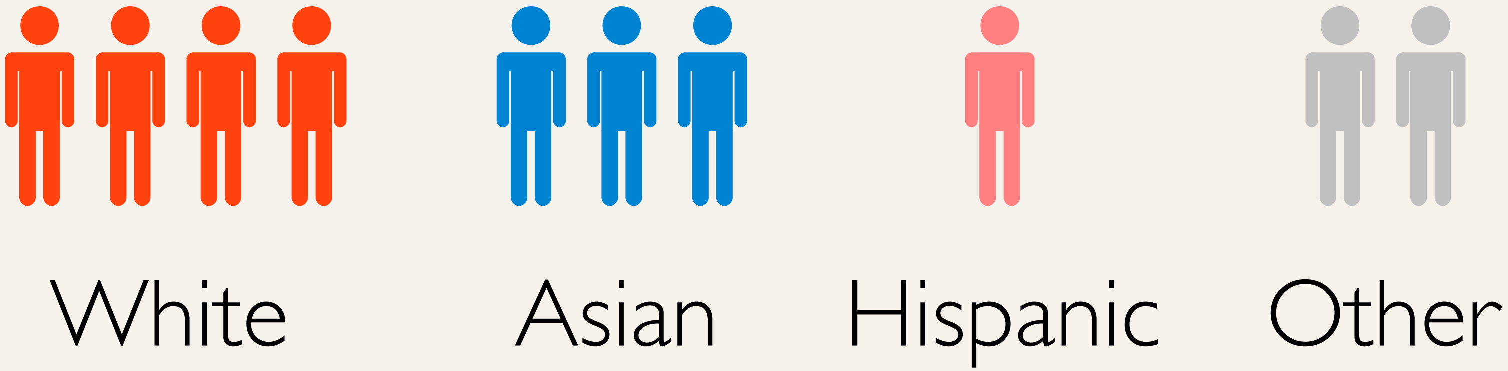
3 in 4 are men



Half of them are single,
and/or under 30

CROWDSOURCING

WORKERS' DEMOGRAPHICS



4 out of 5 own
a smartphone