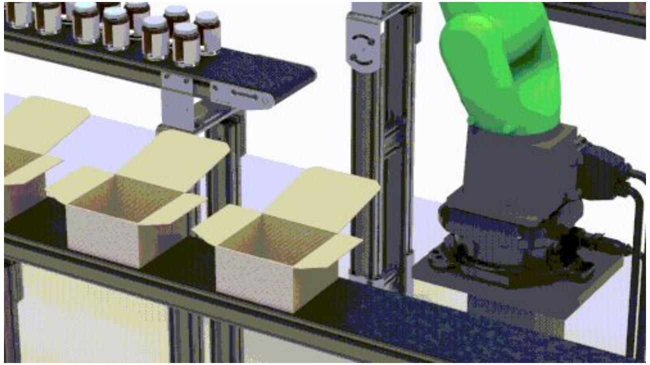
Inference of Human-derived Specifications of Object Placement via Demonstration

Alex Cuellar[1], Ho Chit Siu[2], Julie A Shah[1]

[1] MIT, [2] Lincoln Labs

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Interactive ROBOTICS



Source: Linus Projects

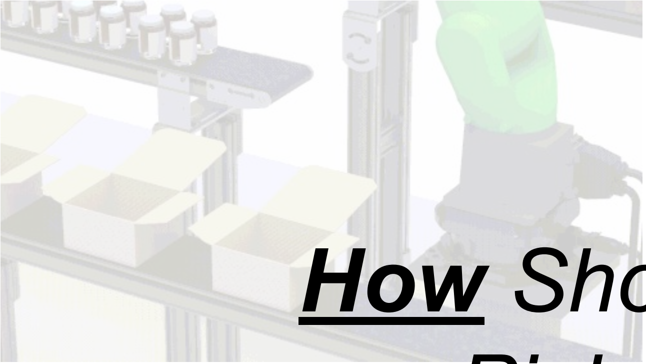
Dexterity

+

Perception

Source: Ambi Robotics

Source: Rutgers University



Source: Linus Projects

Source: Ambi Robotics

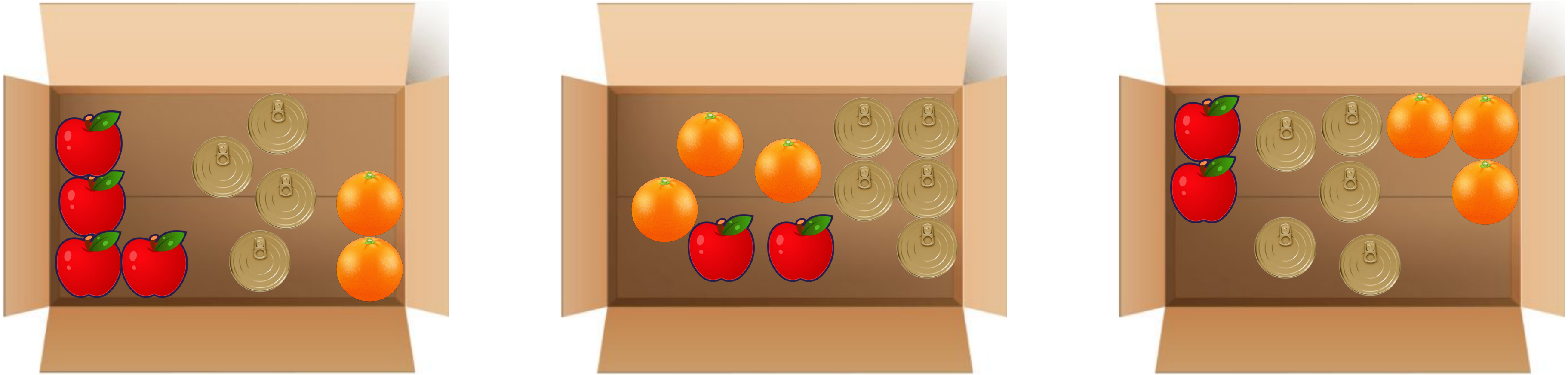
Question:

How *Should a Robot Complete Pick-And-Place Tasks?*

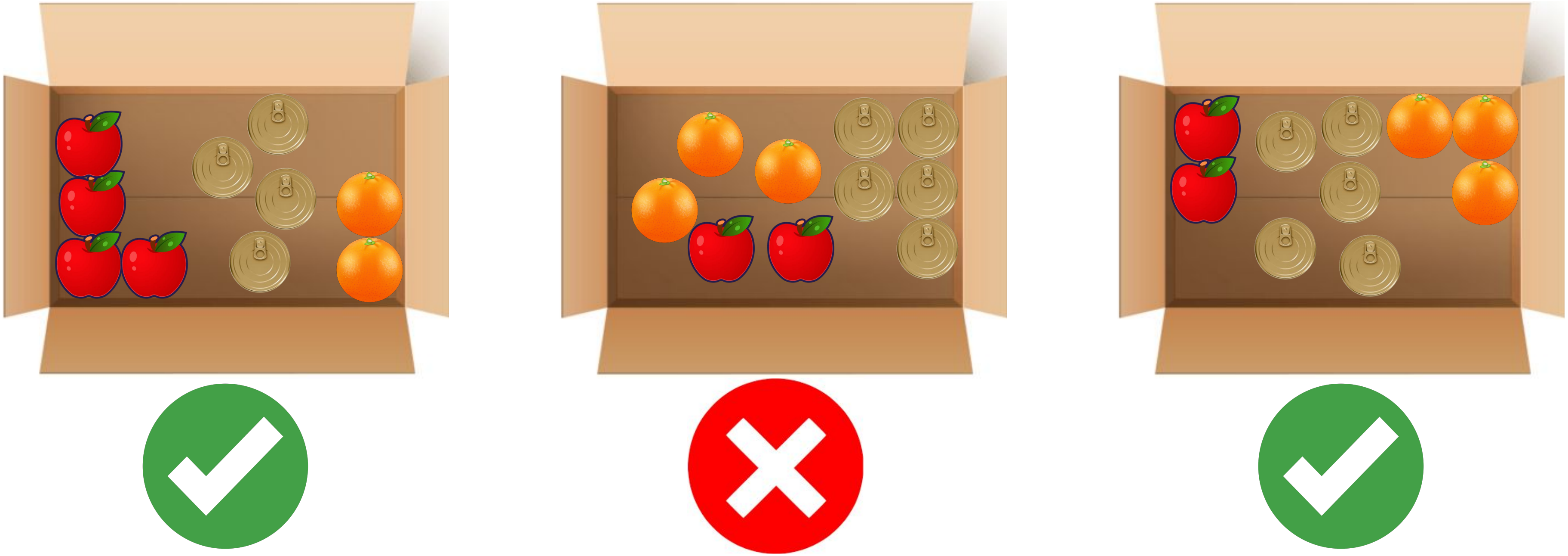
Dexterity
+
Perception

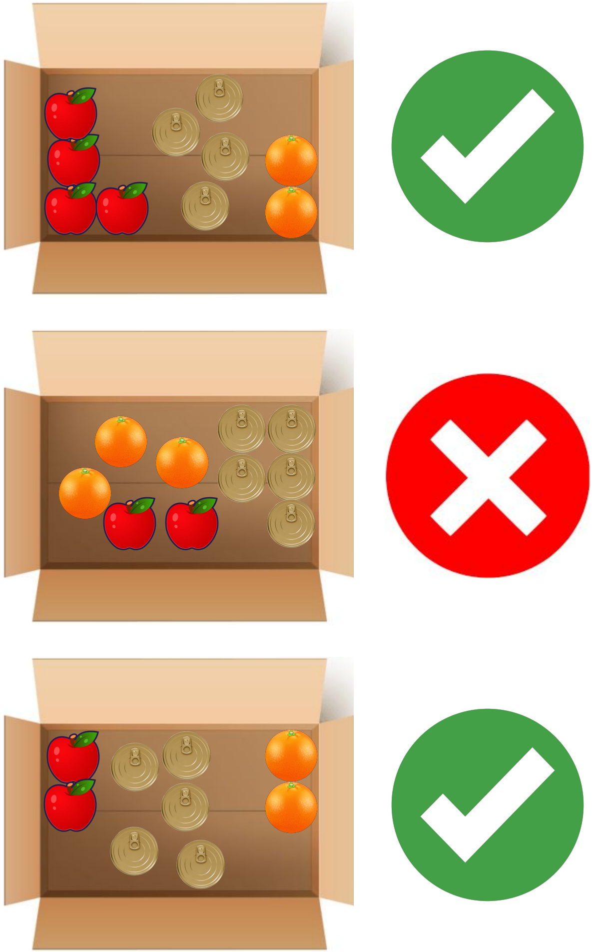
Source: Rutgers University

Which of These Doesn't Belong?



Which of These Doesn't Belong?





How do we know this?

Does not rely on:

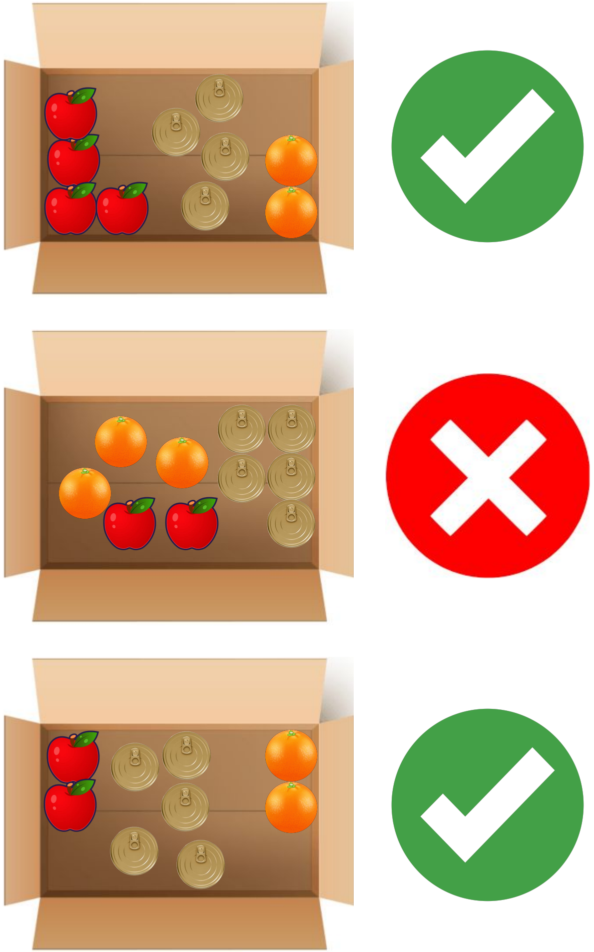
Exact Positions

Individual Objects

Does rely on:

Spatial Relationships

Object "Types"



How do we know this?

Does not rely on:	Does rely on:
Exact Positions	Spatial Relationships
Individual Objects	Object "Types"

Why it Matters?

User Preference/ Expectation

"Apples are to the left of cans"

Requirements

"Fruit should be secured against box walls"

PARCC: Positionally Augmented Region Connection Calculus

A specification language defining requirements on the spatial relationships between classes of objects

PARCC: Positionally Augmented Region Connection Calculus

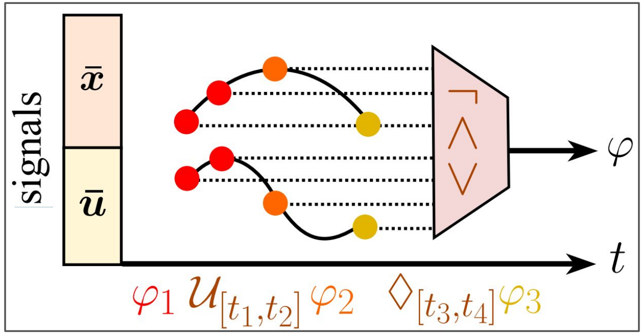
A specification language defining requirements on the spatial relationships between classes of objects

– AND –

An inference algorithm to learn specifications from demonstrations

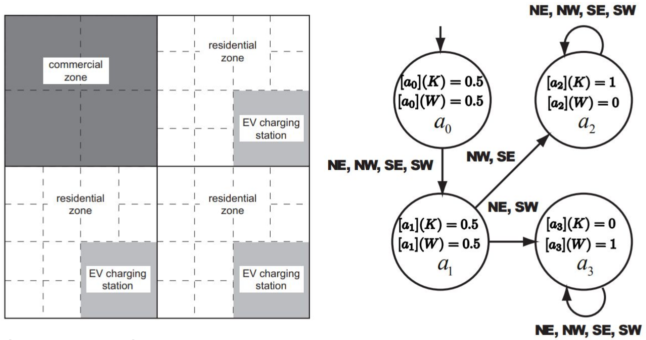
Existing Spatial Specification languages

Signal Temporal Logic [1]



Source: From [3]

Quad-Tree Representations [2]

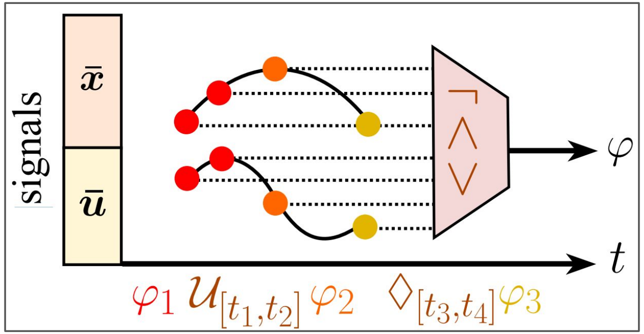


Source: From Citation [3]

[1] Linard, Alexis, and Jana Tumova. "Active learning of signal temporal logic specifications." 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). IEEE, 2020.
[3] Haghighi, Iman, et al. "SpaTeL: a novel spatial-temporal logic and its applications to networked systems." Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control. 2015.
[2] Gu, Zhaoyuan, et al. "Walking-by-logic: Signal temporal logic-guided model predictive control for bipedal locomotion resilient to external perturbations." EEE International Conference on Robotics and Automation (ICRA). 2024.

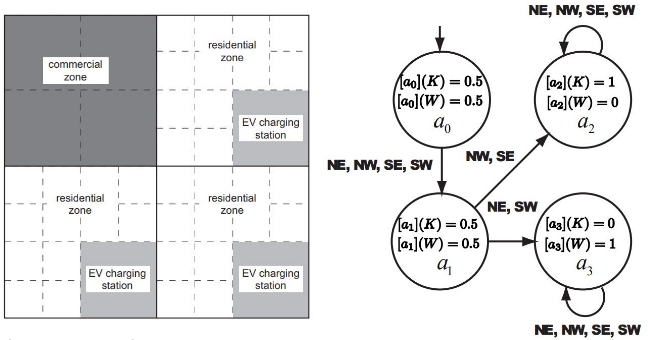
Existing Spatial Specification languages

Signal Temporal Logic [1]



Source: From [3]

Quad-Tree Representations [2]



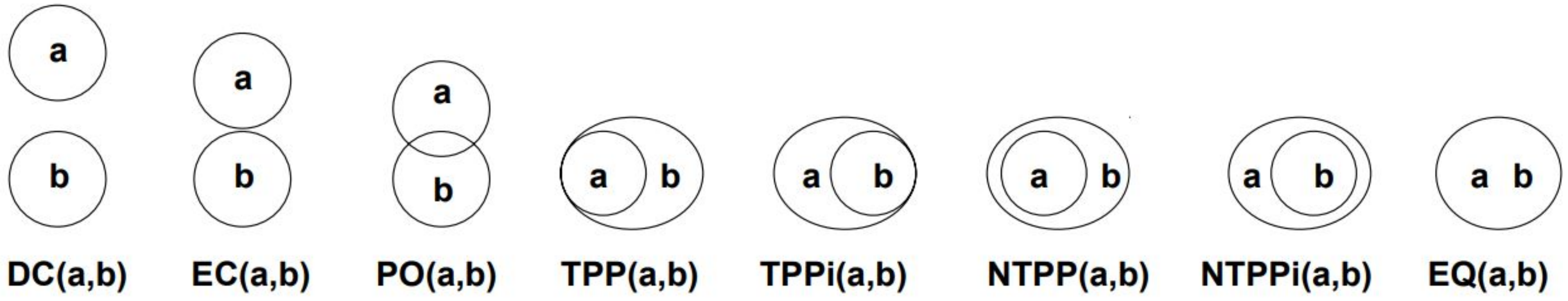
Source: From Citation [3]

Represent specifications over ***specific regions*** or ***precise distance*** in space

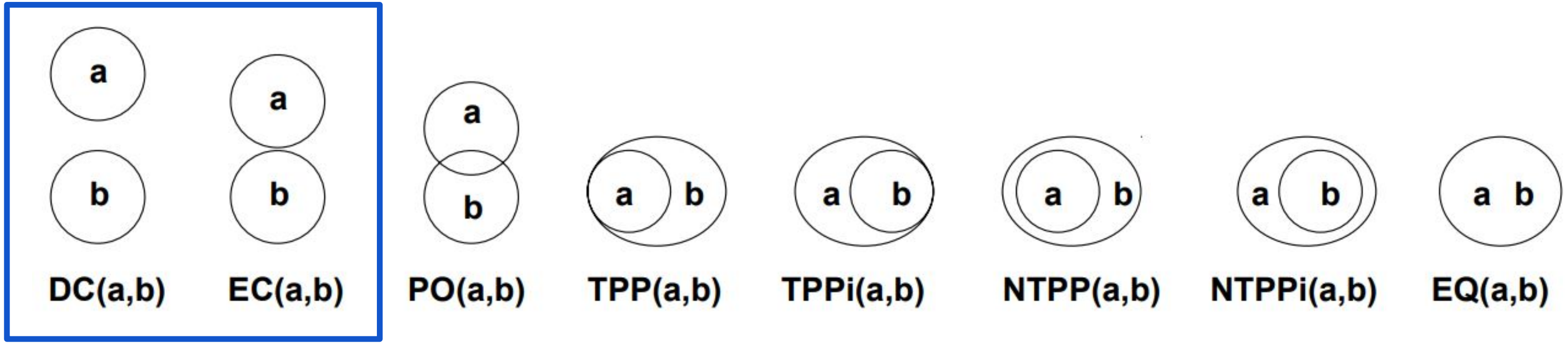
Don't easily represent ***relationships between objects***

[1] Linard, Alexis, and Jana Tumova. "Active learning of signal temporal logic specifications." 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). IEEE, 2020.
[3] Haghighi, Iman, et al. "SpaTeL: a novel spatial-temporal logic and its applications to networked systems." Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control. 2015.
[2] Gu, Zhaoyuan, et al. "Walking-by-logic: Signal temporal logic-guided model predictive control for bipedal locomotion resilient to external perturbations." EEE International Conference on Robotics and Automation (ICRA). 2024.

RCC: Reasoning Over ***Relationships*** of Spatial Regions

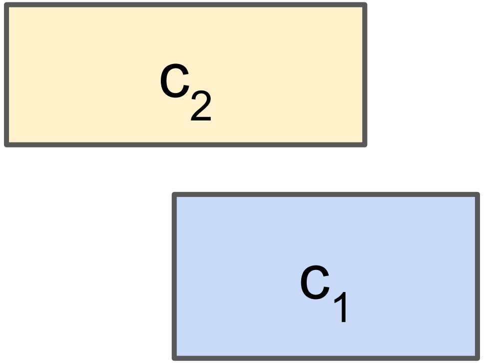


RCC: Reasoning Over *Relationships* of Spatial Regions

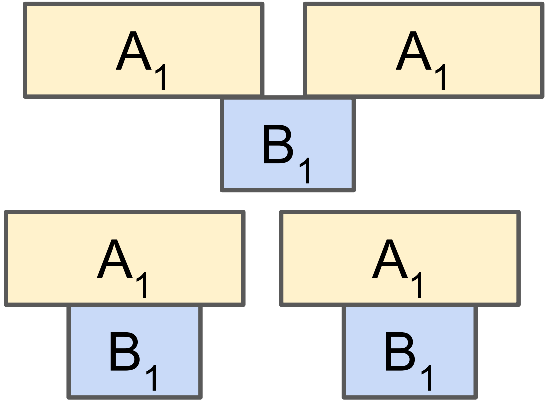


We use the fragment that assumes objects that do not share regions in space

Defining PARCC: *Object Relations*



$$DR_N(c_2, c_1) \rightarrow DR(c_2, c_1) \wedge y_{c_2} \geq y_{c_1}$$
$$\forall (x_{c_2}, y_{c_2}) \in c_2, (x_{c_1}, y_{c_1}) \in c_1$$

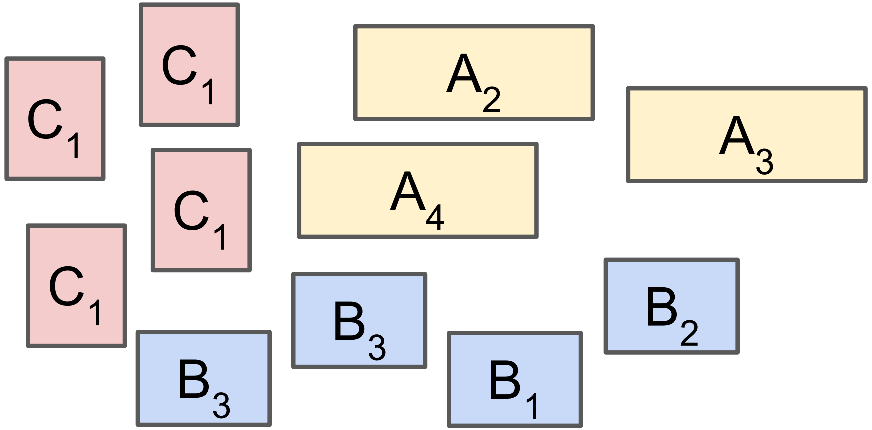


$$EC_N(A, B) \leftrightarrow EC_N(a, b)$$
$$\forall a \in \mathcal{A} \quad \exists b \in \mathcal{B}$$

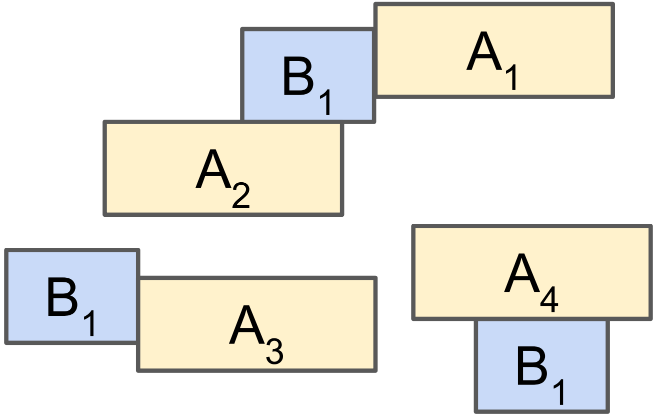
PARCC "Object" Relationships
specify a direction (i.e. N,E,S,W) between the objects it describes

PARCC "Class" Relationships:
Specify a relationship all objects of one class have with another class

Defining PARCC: *Logical Specifications*



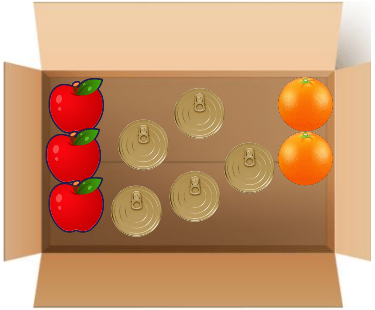
$$DR_N(A, B) \wedge DR_W(A, C)$$

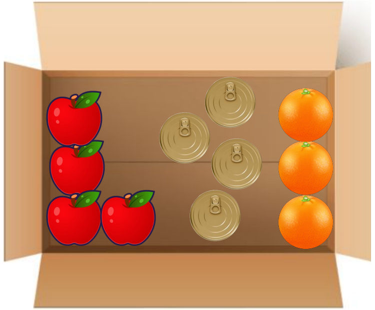


$$EC_E(A, B) \vee EC_S(A, B)$$

PARCC Specifications utilize boolean logic to define requirements involving multiple class relations

Inferring PARCC Specifications



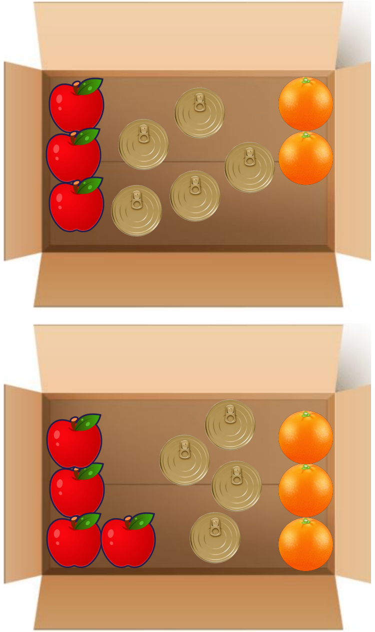


$$DR_E(O,C) \qquad DR_W(A,C)$$

$$\bigvee_{i\in\{N,S,E,W\}} EC_i(O,O)$$

$$\bigvee_{i\in\{N,S,W\}} EC_i(A,W)$$

Inferring PARCC Specifications



Inference

Disjunctive "constraints"

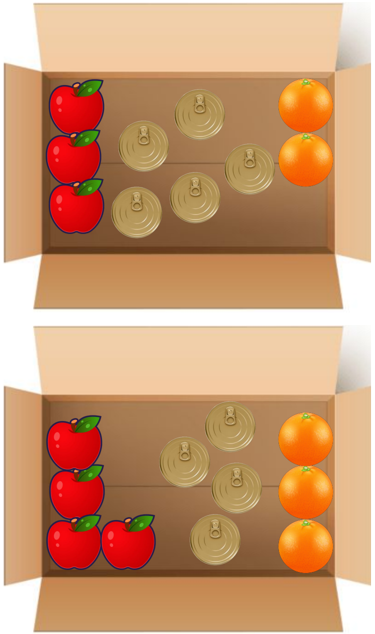
$$\mathcal{C} = \{\phi_1, \phi_2 ...\}$$

$$DR_E(O, C) \qquad DR_W(A, C)$$

$$\bigvee_{i \in \{N,S,E,W\}} EC_i(O, O)$$

$$\bigvee_{i \in \{N,S,W\}} EC_i(A, W)$$

Inferring PARCC Specifications



Disjunctive "constraints"

$$\mathcal{C} = \{\phi_1, \phi_2 \ldots\}$$

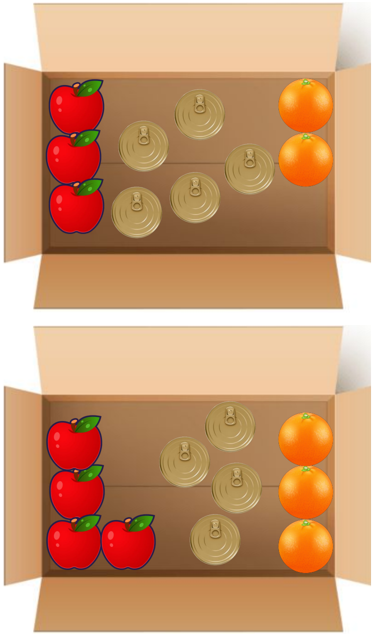
$$DR_E(O, C) \qquad DR_W(A, C)$$

$$\bigvee_{i \in \{N,S,E,W\}} EC_i(O, O)$$

$$\bigvee_{i \in \{N,S,W\}} EC_i(A, W)$$

Final Specification: $\Phi = \wedge_{\phi \in \mathcal{C}} \phi$

Inferring PARCC Specifications



Disjunctive "constraints"

$\mathcal{C} = \{\phi_1, \phi_2 ...\}$

$DR_E(O, C) \qquad DR_W(A, C)$

$$\bigvee_{i \in \{N,S,E,W\}} EC_i(O, O)$$

$$\bigvee_{i \in \{N,S,W\}} EC_i(A, W)$$

Final Specification: $\Phi = \wedge_{\phi \in \mathcal{C}} \phi$

GOAL: Infer human's intended $\mathcal{C}$

Inferring PARCC Specifications

All Possible Disjunctive Formulas:

N=1:

$DR_W(A,C)$

$EC_E(O,A)$

$DR_N(A,O)$

$\dots$

N=2:

$DR_W(A,C) \vee DR_S(A,C)$

$EC_W(A,C) \vee DR_W(A,A)$

$EC_W(O,O) \vee DR_N(O,O)$

$\dots$

$\dots$

Satisfying Formulas:

N=1:

~~$DR_W(A,C)$~~

$EC_E(O,A)$

~~$DR_N(A,O)$~~

$\dots$

N=2:

$DR_W(A,C) \vee DR_S(A,C)$

~~$EC_W(A,C) \vee DR_W(A,A)$~~

$EC_W(O,O) \vee DR_N(O,O)$

$\dots$

$\dots$

For each satisfying formula ϕ:

Probability that ϕ is incidentally satisfied:

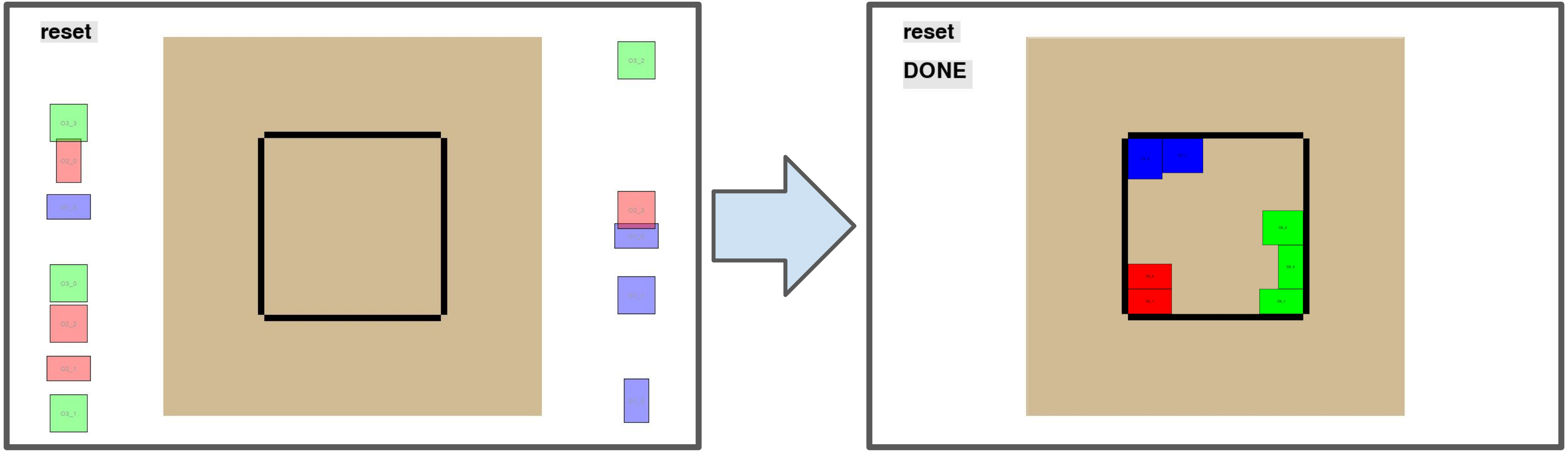
$$P(D \to \phi|\mathcal{R}) = \prod_{o \in \mathcal{O}_D^C} \max\left(\epsilon, \frac{\sum_{R \in \mathcal{R}} \sum_{o' \in \mathcal{O}_R^C} \mathbf{1}(o' \to \phi)}{\sum_{R \in \mathcal{R}} \sum_{o' \in \mathcal{O}_R^C} 1}\right)$$

- $P(D \to \phi|\mathcal{R}) < p_c$: Keep ϕ
- $P(D \to \phi|\mathcal{R}) > p_c$: Discard ϕ

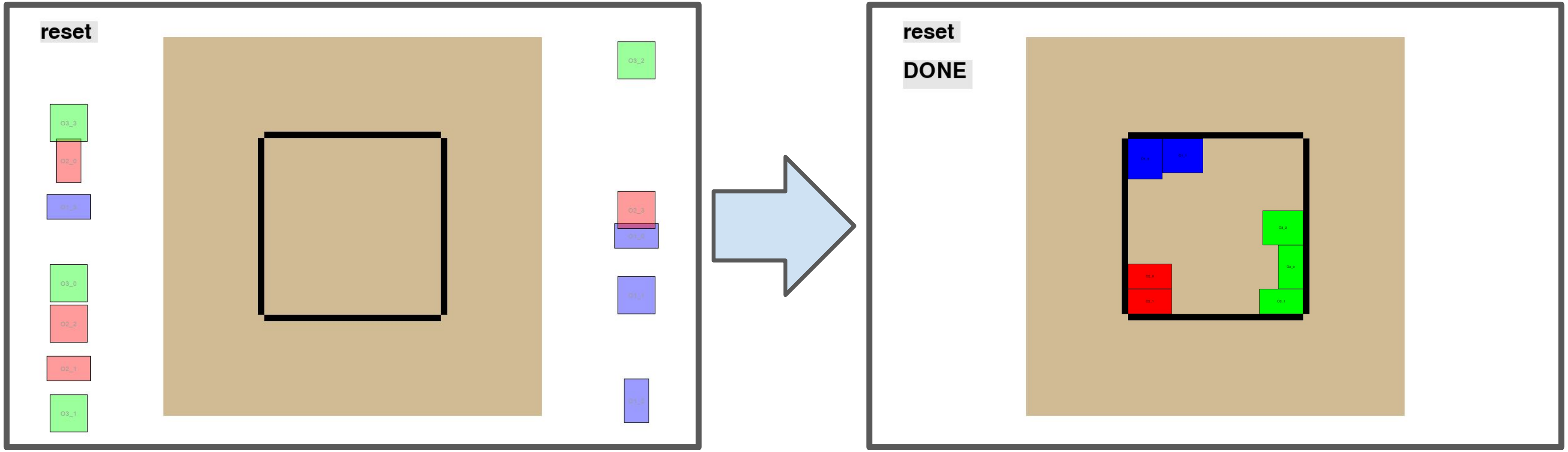
Inferring PARCC Specifications

Does this really capture human's ground truth object placement preferences?

Box Packing Domain



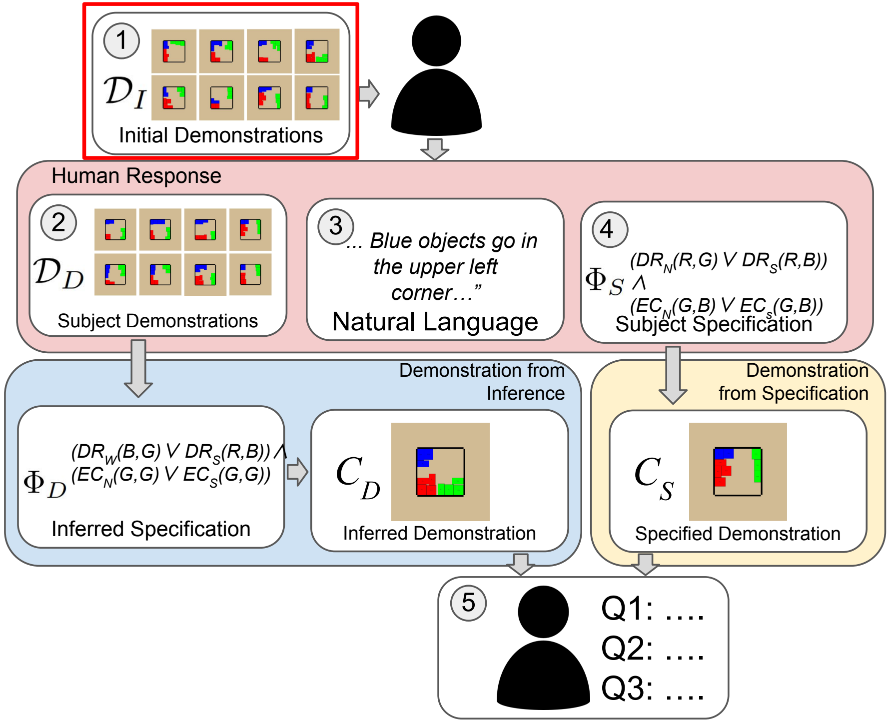
Box Packing Domain

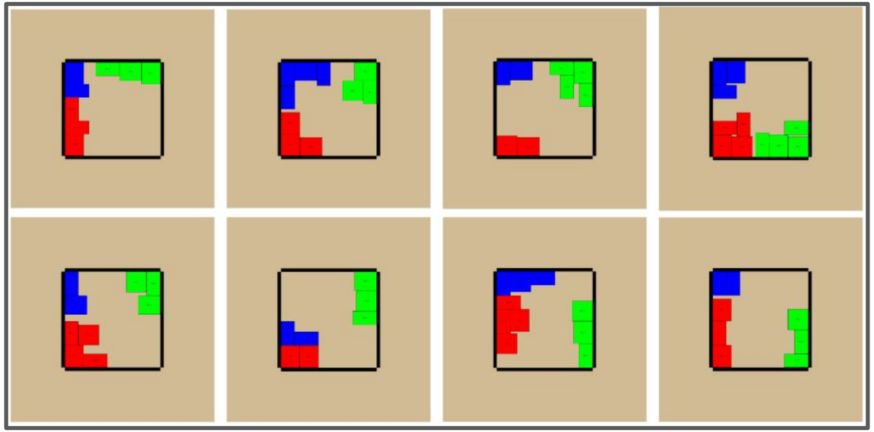


Questions:

- Can the PARCC inference algorithm capture intuitive requirements on object placement?
- Can the inference algorithm capture requirements better than direct specification?

Human Study Design

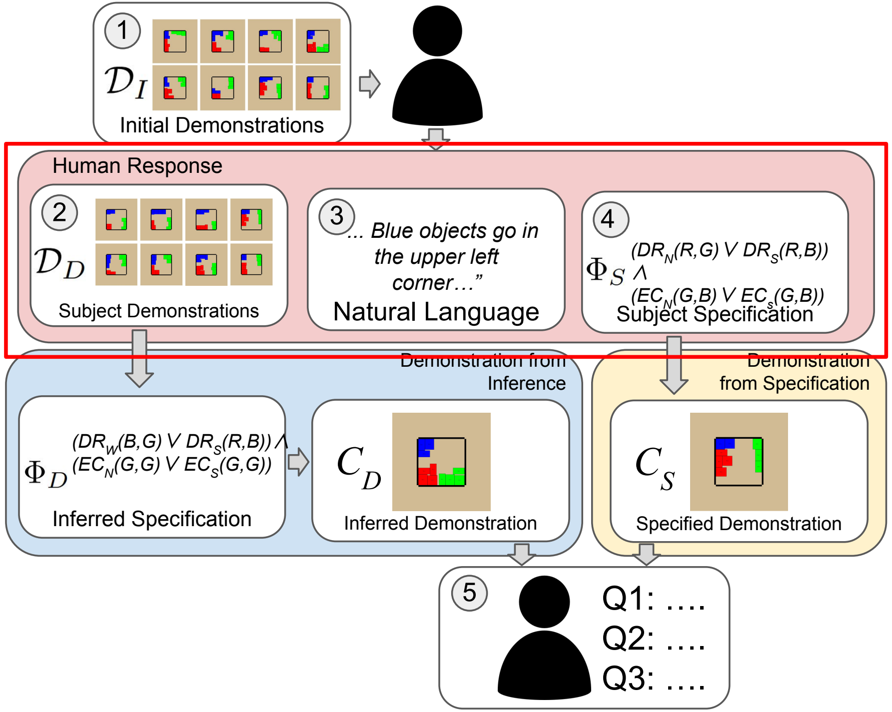




Show the participant 8 *initial demonstrations* that intentionally use a specification of 12 formulas:

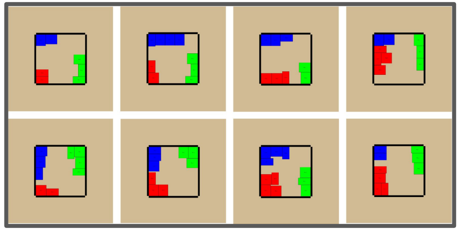
$$\phi_1 \ldots \phi_{12}$$

Human Study Design



The subject provides:

1



Demonstrations matching the initial examples

2

"... Blue objects go in the upper left corner…"

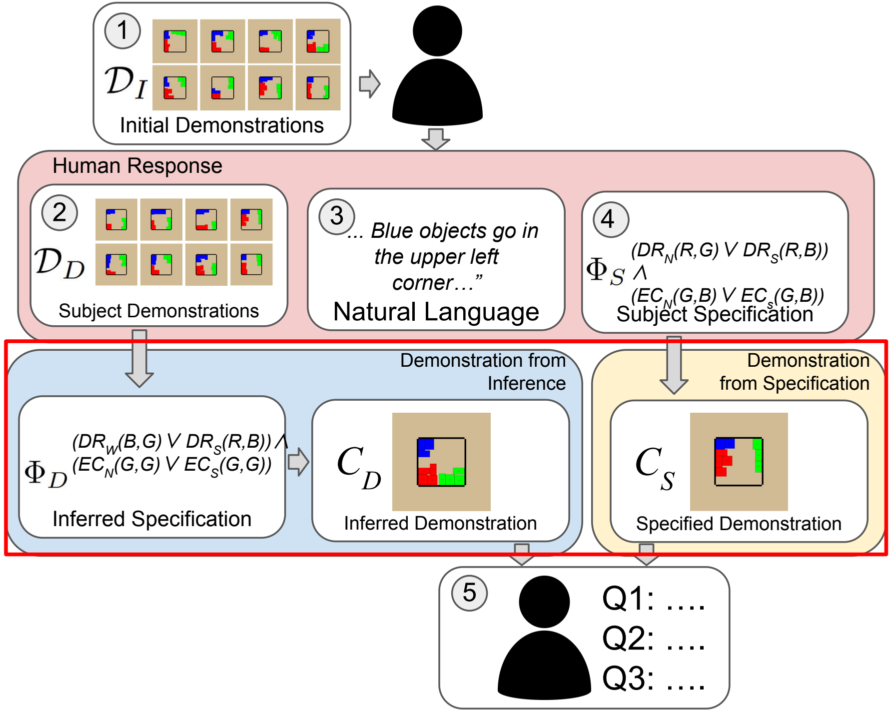
Natural language description of requirements

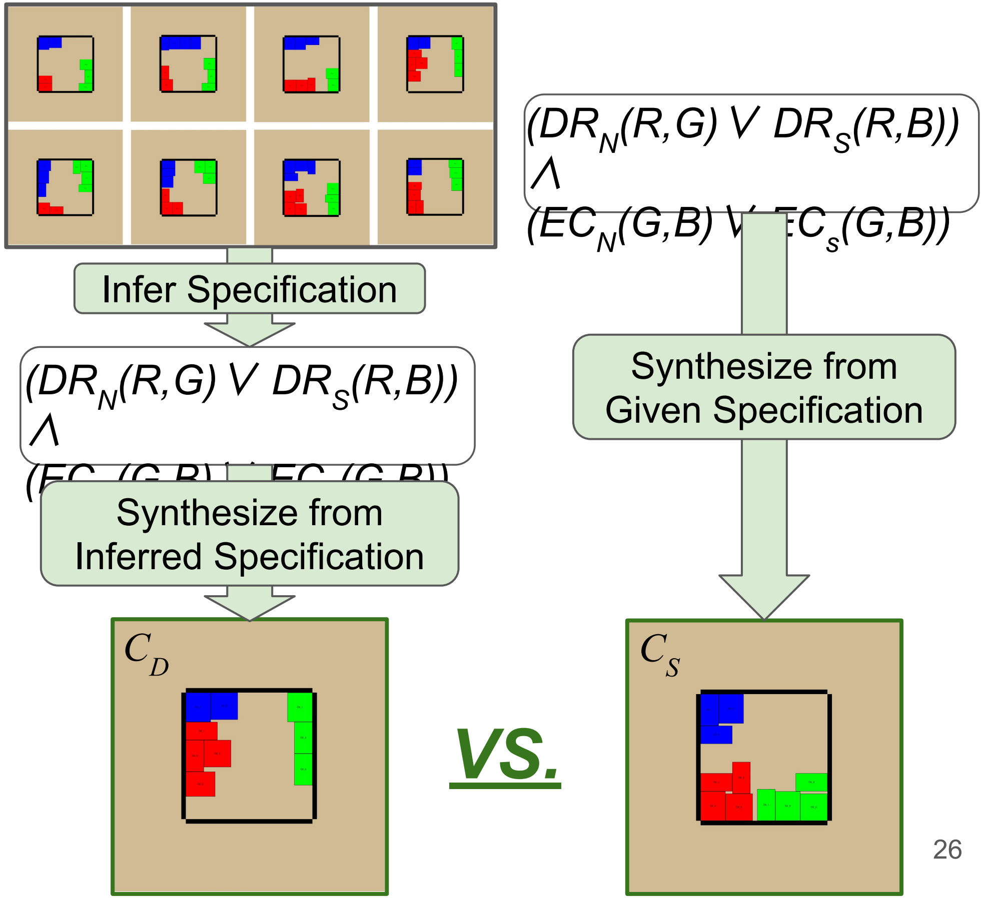
3

$(DR_N(R,G) \vee DR_S(R,B)) \wedge (EC_N(G,B) \vee EC_s(G,B))$

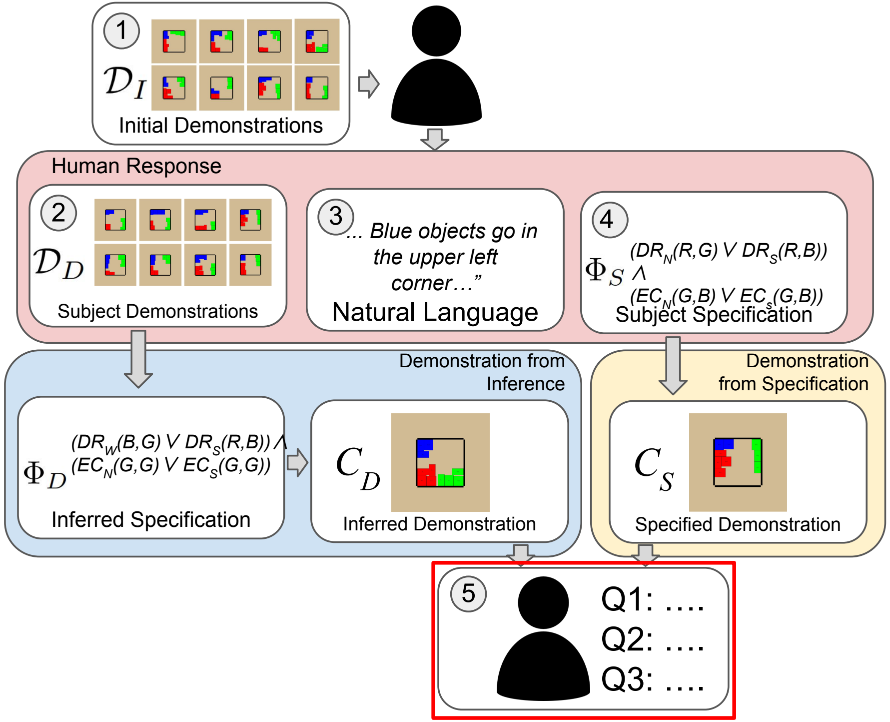
PARCC specification

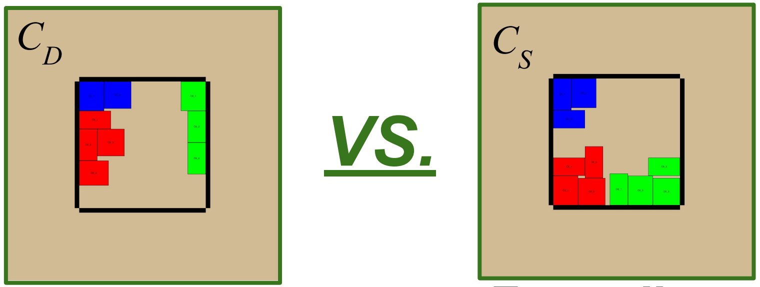
Human Study Design





Human Study Design



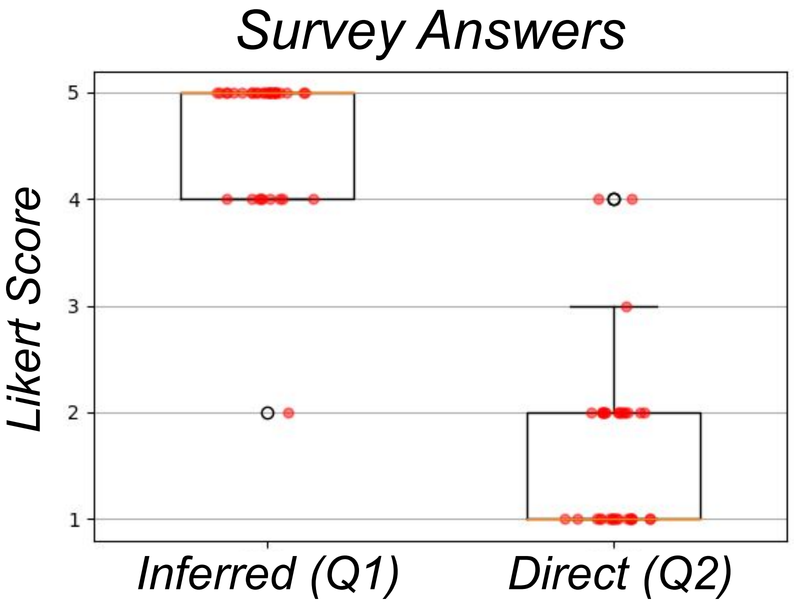


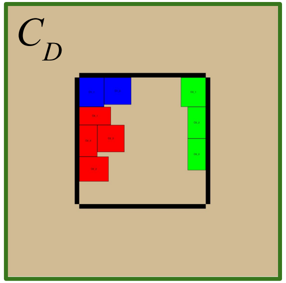
From Inferred specification ***From direct specification***

1) The *[inferred example]* matches patterns in my demonstrations.

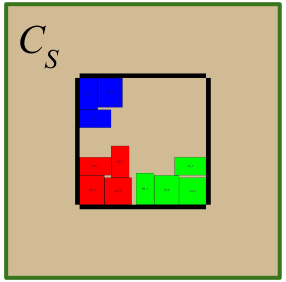
2) The *[direct example]* matches patterns in my demonstrations.

Human Study Results: Inferred vs Direct Specifications





VS.



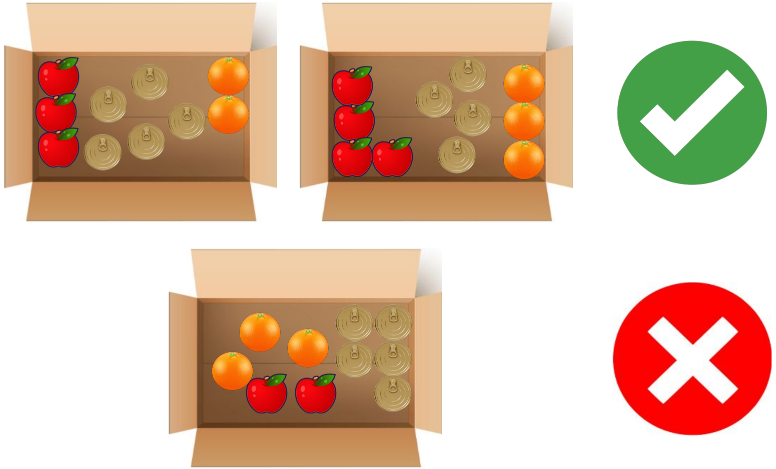
From Inferred specification

From direct specification

1) The *[inferred example]* matches patterns in my demonstrations.
2) The *[direct example]* matches patterns in my demonstrations.

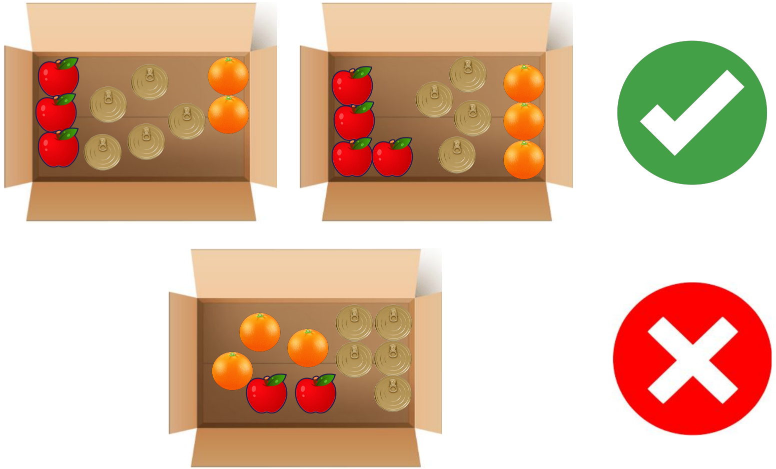
Limitations

- PARCC Representations:
  - Constrained to a 2D space
  - Only uses rectangular objects
- Inference Algorithm:
  - More efficient sampling of candidate disjunctions



Limitations

- PARCC Representations:
  - Constrained to a 2D space
  - Only uses rectangular objects



- Inference Algorithm:
  - More efficient sampling of candidate disjunctions

Questions?