

# Grasping Beyond Reach: Selecting Out-of-Reach Objects in XR with Point&Grasp

Xuejing Luo<sup>1,2</sup>, Hee-Seung Moon<sup>3</sup>, Christian Holz<sup>4</sup>, Antti Oulasvirta<sup>1,2</sup>

<sup>1</sup> Aalto University, Finland <sup>2</sup> ELLIS Institute, Finland <sup>3</sup> Chung-Ang University, Republic of Korea <sup>4</sup> ETH Zürich, Switzerland



## Problem

*Ambiguity Makes Out-of-reach Selection Hard.*

In XR, out-of-reach objects are distant and similar. A single cue is not enough to infer the intended target.

Imagine wearing a VR headset and viewing this virtual kitchen. You need to select distant objects without physically moving.

### Spatial Ambiguity

Pointing direction meets multiple candidates.

### Semantic Ambiguity

Grasping gestures are compatible with multiple objects.

?

How can we flexibly integrate cues to select the right target?

## Method

### Probabilistic Cue Integration

Both cues are extracted from the same hand-tracking inputs.

#### Directional Cue

$$p(c_D | \sigma) \propto \exp\left(-\frac{\|p_e - p_o\|^2}{2\sigma_d^2}\right)$$

- $p_e$  : Pointing endpoint
- $p_o$  : Object center
- $\sigma_d$  : Directional uncertainty (controls spread)

#### Grasping Gestural Cue

$P(\text{grasp} | \text{object})$

**Bayesian Integrator**

$$p(o | c_D, c_G) \propto p(c_D | o) \cdot p(c_G | o) \cdot p(o)$$

↓  
**Posterior**  
 $p(o | c_D, c_G)$

→ **Inferred Target**  
apple

Compute the posterior for all candidate objects; the one with the highest posterior is selected as the inferred object.

Combine spatial and semantic cues via probabilistic inference.

## Results

### User Study under Controlled Ambiguity

#### Overall Performance

Selection Time [seconds]

#### Overall Performance

Trial Completion Rate

Point&Grasp significantly improves both efficiency and success rate over single-cue methods.

#### Performance under 4 Ambiguity Conditions

Selection Time [seconds]

Point fails under spatial ambiguity, while Grasp fails under semantic ambiguity. Point&Grasp combines both cues to remain robust across conditions.

#### Overall Patterns

#### High Spatial + High Semantic

#### High Spatial + Low Semantic

#### Low Spatial + High Semantic

#### Low Spatial + Low Semantic

Success is usually driven by **one dominant cue**, not cue agreement. Gesture dominates under high spatial ambiguity, direction becomes more informative under semantic ambiguity, and **fusion can still recover some trials when neither cue alone is sufficient.**

Fusion yields faster, more accurate, and more robust selection.

## Takeaway:

Out-of-reach object selection is an inference problem under ambiguity, not a deterministic mapping from a single cue. Probabilistic cue integration makes selection faster, more accurate and more robust in XR.

**Project page**