# Grounding Verbs in Intuitive Physics: Few-Shot Categorization of Action Events

Jingyi Wu[1, 2] Joshua Hartshorne[3] Tao Gao[2]

[1]Department of Linguistics, Zhejiang University [2]Department of Commnuication, UCLA [3]MGH Institute of Health Professions

## Introduction

### (No) Double-object Dative

Lafleur **slid** the puck **to** the goalie.

Lafleur **slid** the goalie the puck.

Lafleur **lifts** the crate **to** him.

Lafleur **lifts** him the crate.

Verb concepts grounded in a physical world model → Verb argument structures

### Physically-grounded verb argument structure

**Double-object construction:** Instantaneous application of force to send the object on a trajectory to a recipient

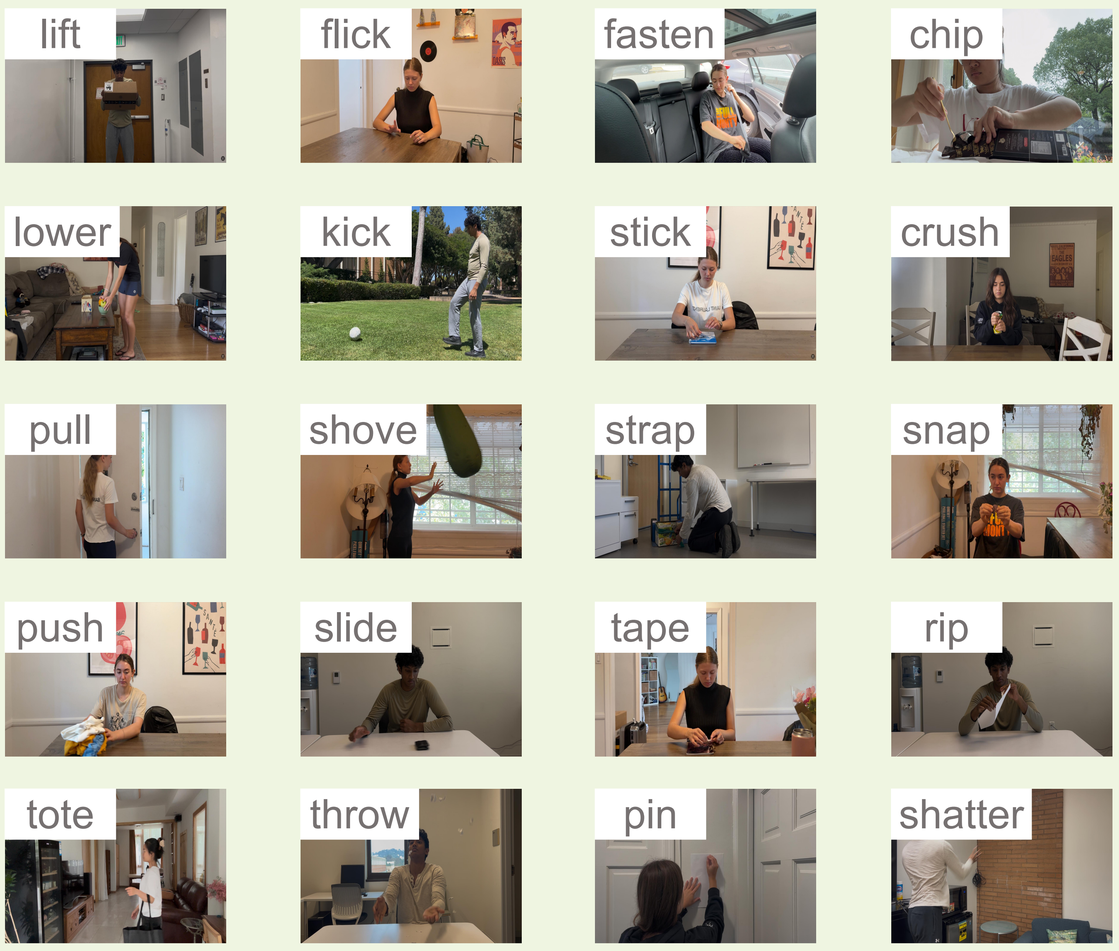
**No double-object construction:** Continuous application of force to an object to keep it moving

**Hypothesis:**
1. Humans categorize action videos based on the physical and causal attributes revealed by verb argument structure.
2. These attributes form the building blocks of human thought, hard-coded into grammar.

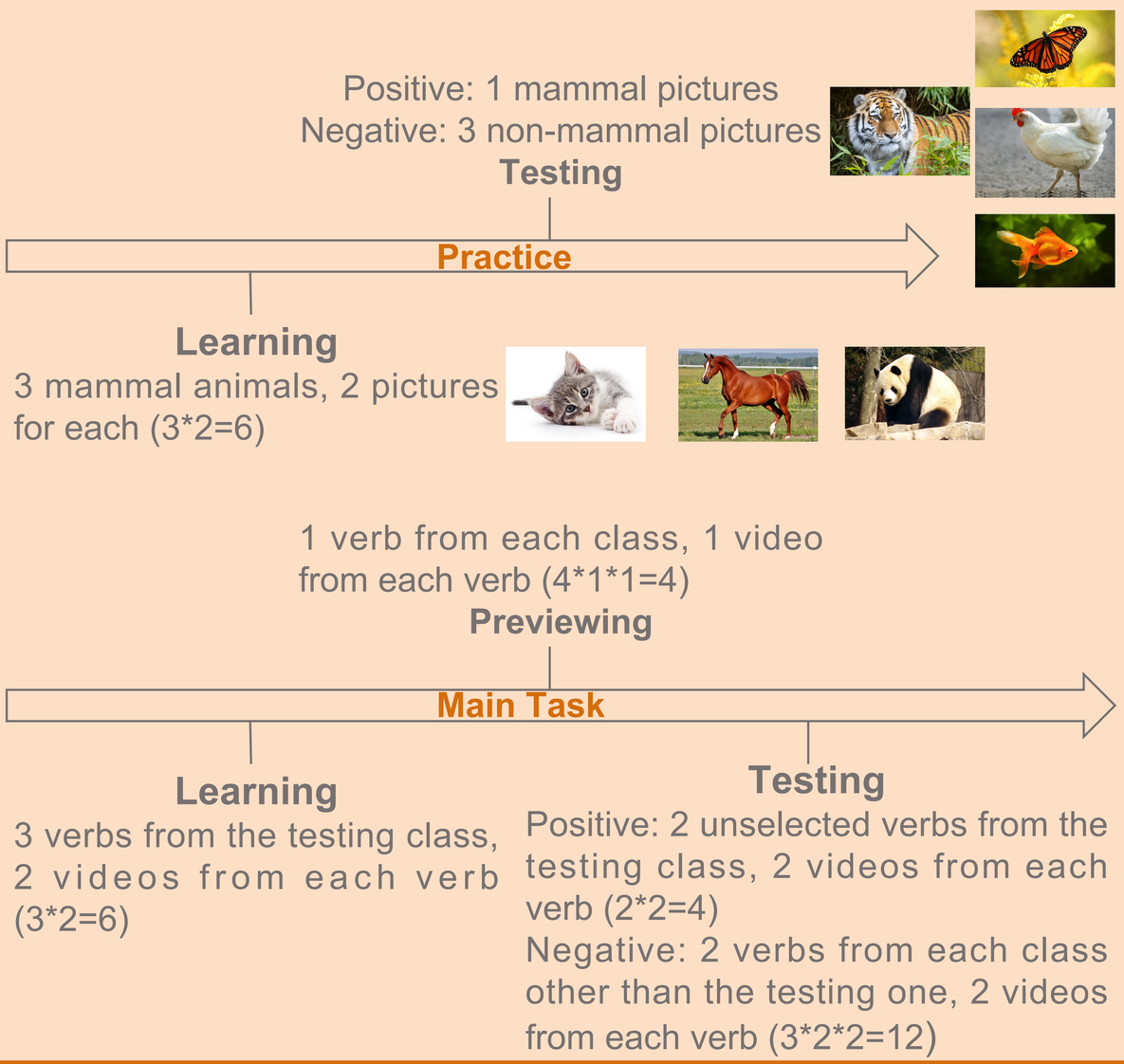
## Video Dataset

**4 Classes × 5 Verbs × 4 videos = 80 videos**

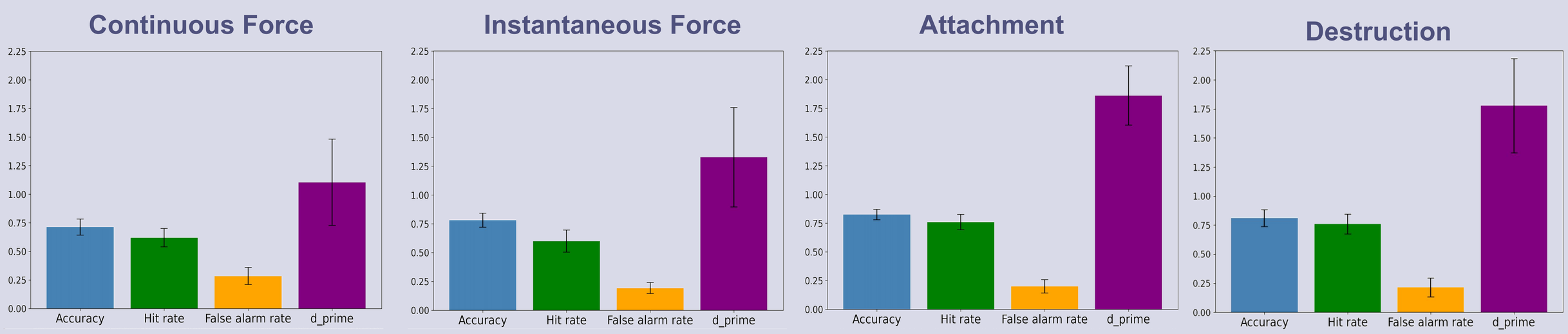
| Continuous | Instantaneouos | Attachment | Destruction |
|---|---|---|---|
| lift | flick | fasten | chip |
| lower | kick | stick | crush |
| pull | shove | strap | snap |
| push | slide | tape | rip |
| tote | throw | pin | shatter |



## Experiment Design




## Results & Discussion




**Discussion:**

1. **Successful few-shot event categorization:** Participants categorize unseen actions with ~70% accuracy from only a few positive examples.
2. **Variation across event categories:** Attachment and destruction are the easiest, while continuous vs. instantaneous force is often confused, with a bias to overgeneralize instantaneous as continuous.