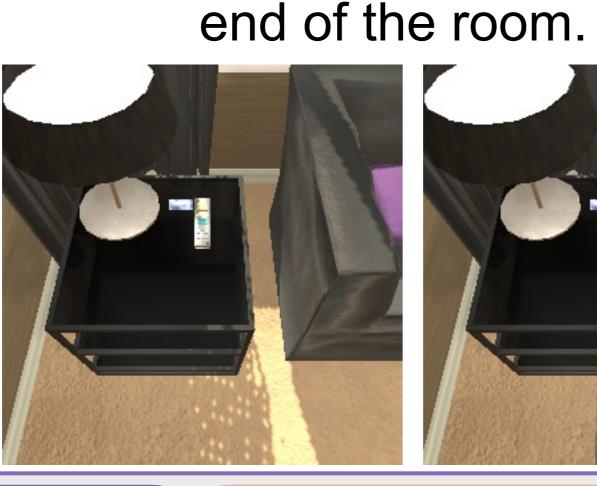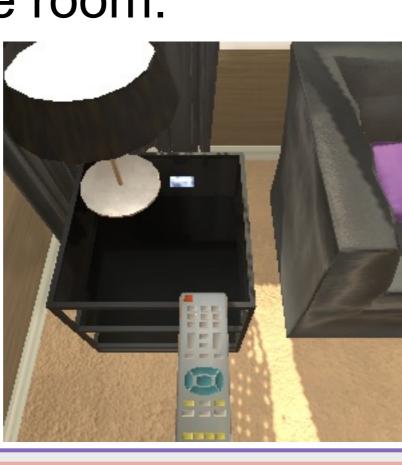**Generated Captions**

(w/ Commonsense)
1. Turn around and walk to the coffee table.

(w/o Commonsense)
1. Turn around and walk over to the end of the room, then turn right and walk over to the end of the room.

Explicit knowledge

<AtLocation> **house**
<Causes> **turn around**
<xWant> to sit down

Implicit Knowledge:

and **grab** a couple of candles and sit down

(w/ Commonsense)
2. *Pick up* the remote from the couch.
3. Turn around and walk to the coffee table.
4. Put the remote on the coffee table.
5. Pick up the remote from the coffee table.
6. Turn around and walk to the chair.
7. Put the remote on the shelf.

(w/o Commonsense)
2. Pick up the book on the bed.
3. Turn around and walk over to the couch.
4. Place the phone on the bed.
5. Pick up the book on the table.
6. Turn around and walk over to the wooden desk.
7. Place the book on the bed.

Generated Captions

(w/ Commonsense)
1. Turn around and walk to the stove.
2. Pick up the pan with the knife in it.
3. Turn around and walk to the sink.

(w/o Commonsense)
1. Turn around and walk to the fridge.
2. Pick up the pan on the counter.
3. Turn around and walk to the sink.

Explicit knowledge

<Causes> turn on water
<Desires> turn around
<xWant> to **wash the dishes**

Implicit Knowledge:

for more soap.

(w/ Commonsense)
4. Put the pan *in the sink*.
5. Turn around and walk to the stove.
6. Put the pan on the counter.

(w/o Commonsense)
4. Pick up the pan with the spoon in it.
5. Put the pan on the counter.
6. Put the pan on the counter.

Generated Captions

(w/ Commonsense)
1. Gymnast is seen standing on a horse and begins to do a routine on the bars.

(w/o Commonsense)
1. Gymnast is seen standing on a horse and begins performing a gymnastics routine.

Explicit knowledge

<AtLocation> **gym**
<Has A> **gymnastics**
<xNeed> to be **in a gym**

Implicit Knowledge:

of the arena.

(w/ Commonsense)
2. Gymnast is seen performing a routine on the bars and begins to do a routine *in the gym*.

(w/o Commonsense)
2. Gymnast is seen standing on a horse and begins performing a gymnastics routine.

Generated Captions

(w/ Commonsense)
1. Woman is seen speaking to the camera while holding a large bowl and leads into her speaking to the camera and leads into her hands and leads into her hands holding up a bowl.
2. Woman is standing in a kitchen and begins to make a juice.

(w/o Commonsense)
1. Woman is seen speaking to the camera and leads into her holding up a bowl.
2. Woman is standing in front of a counter and talking to the camera.

Explicit knowledge

<AtLocation> kitchen
<Causes> juice
<Has A> PersonX puts the juice **in the glass**

Implicit Knowledge:

of an orange and a small slice of lemon.

(w/ Commonsense)
3. Woman is seen speaking to the camera while holding *a glass* and leads into her speaking to the camera and leads *into the glass*.

(w/o Commonsense)
3. Woman is standing in a kitchen talking to the camera and leads into her pouring it into a bowl.