Abstract

In this paper, we introduce a large-scale EXCALIBUR benchmark to encourage and evaluate embodied agents' ability to explore. We also collected human performance with Virtual Reality headsets through an immersive integration of our environment. In the supplementary material, we provide the following items that shed further insight into these contributions:

- *A An immersive interface to EXCALIBUR, that enables human annotators to transfer their knowledge from the real world to the virtual environment.*
- *B Objects in EXCALIBUR have a large diversity in terms of categories, colors, materials, sizes, and weights. Within each category, there are different handcrafted realistic object models.*
- *C A large diversity of procedurally generated houses.*
- *D Question filtering process that automatically chooses the most challenging questions for all scenes.*
- *E Handcrafted templates that are used for generating questions.*
- *F An explanation of Eq. 1 and how the constant is found based on human performance.*
- *G Detail description of the baseline model we used.*
- *H Discussion on detailed evaluation of agents and humans performance.*

A. Immersive Human Performance Collection for EXCALIBUR

In the supplementary material package, we provide a sample of a human episode (`human_traj.mp4`). Here, we mainly discuss the technical details of designing the VR interface for AI2THOR.

A.1. Design choices of VR interface

The most straightforward reason to use VR for data collection is that controlling the arm and grasping with keyboard input can be non-intuitive, while VR makes it possible to directly apply the sensorimotor experience. The agent's arm movement replicates human arm movement in all spatial dimensions, developing hinge joints for elbows, ball-and-socket joints for shoulders, and synovial joints for the wrist. These simulations are done using the Oculus hand controllers, facilitating a smooth interface for human performance. A user can pick up and drop objects in the VR, using the Oculus Grip Button, wherein the intensity with which the grip button is pressed determines the force with which the arm will exert to pick up an object in VR. Since different objects in the room have different masses, we require different amounts of force to pick up different objects. This will map a human's intuition of how heavy an object is to that of the agent's arm, making it a nearly exact emulation of the human mind coupled with actual physics.

This realistic interface also provides a great challenge for comparing human performance with agents. With OpenXR, we can use the exactly same scenes in both VR and agents' experiments. However, the VR interface has a continous action space (except for opening drawers, fridges and closets), while agents' action space is discrete. To make the comparison as fair as possible, we also count humans' actions as discrete time steps. All participants completed the data collection while sitting and try to only controller to move and rotate. The grasp force is discretized: whenever the force changes more than 0.05 kg, we count it as a timestep. The same applies to joint movements. This underestimates humans' performance, but we still find the huge gap between humans and best model as shown in Tab. 2.

A.2. Data collection procedure

We conducted the whole human performance collection in the following steps:

1. 7 college students with no prior experience of using VR headsets are recruited as participants;

2. To train each participant, we use a hand-crafted ARCHITECTHOR scene, which is not in the test set. After the participants wear the headset, we ask them to freely explore the houses for as long as they want, and to try out all of the actions on the controller. We look at a mirrored screen and give instructions whenever a collision happens. A training session ends when the participant lifts and drops every pickupable object and correctly answers 20 questions consecutively. The training sessions last about one hour per participant.

3. Each participant does 4 scenes with replay and 5 scenes without replay or 5 scenes with replay and 4 scenes without replay. The difference between with and without replay is whether the participant can look at the replay of their own

recording of Phase I and Phase III in Phase II and Phase IV. Before a participant enters the scene, we would tell them whether this episode is with replay or without. When a participant is exploring in the first Phase, we look at the live number of time steps they consumed. Whenever it reaches 2,500 steps, they are forced to stop. In practice, most of the participants spend less than 2,500 steps before deciding to stop themselves. One episode typically lasts about one hour per participant.

B. Full object list

See Table 4 for a listing of objects in AI2-THOR and their default properties (*e.g.* mass, volume, ||).

C. House distribution

Alg 1 shows the procedure to randomize houses, agent spawn locations and object physical properties.

Algorithm 1 House randomization procedure

Require: ProcTHOR houses $\mathcal{S}$
  for $S \in \mathcal{S}$ **do**
    $\mathcal{P} \sim$ reachable locations in S
    while Agent collides with objects at location $\mathcal{P}$ **do**
      $\mathcal{P} \sim$ reachable locations in S
    end while
    for object in $\mathcal{H}$ **do**
      Sample scale factor $\alpha \sim [0.8, 1]$, and mass factor $\beta \sim [0.5, 1.5]$
      object.size $= \alpha \times$ object.size
      object.mass $= \beta \times$ object.mass
    end for
  end for

D. Question filtering

Alg. 2 shows the procedure to filter questions based on the distributions of answers. For each question, we calculate the answer distribution among all houses. We choose the questions with the highest

Algorithm 2 Question filtering

Require: Template set $\mathcal{T}$, Houses $\mathcal{S}$, number of questions generated per scene N, maximum number of questions per template M
  $Q \leftarrow \emptyset$
  for $T \in \mathcal{T}$ **do**
    $Q' \leftarrow \texttt{Dict}[\texttt{question} \rightarrow \texttt{Counter}[\texttt{answer}]]$
    for $S \in \mathcal{S}$ **do**
      for $q \in \text{Generate}(T, S)[: \texttt{N}]$ **do**
        if $\exists$ valid $a \leftarrow \text{Answer}(S, q)$ **then**
          $Q'[q][a] \leftarrow Q'[q][a] + 1$
        end if
      end for
    end for
    $Q' \leftarrow \text{Rank}_{q \in Q'}$ Ent $\leftarrow$ Shannon-Entropy $Q'[q]$
    $Q \leftarrow Q \bigcup Q'[: \texttt{M}]$
  end for

Type	# Unique	Properties	Volume	Size	Mass	Salient Colors	Salient Materials
AlarmClock	31	CanPickup	0.001m^3-0.007m^3	0.11m-0.27m	0.80kg		Glass, Metal, Plastic
AluminumFoil	1	CanPickup	0.001m^3	0.24m	0.90kg		Metal, Paper
Apple	31	CanBeSliced, CanPickup	0.000m^3-0.002m^3	0.08m-0.15m	0.20kg	Green, Orange, Red, RedOrange, Yellow, YellowGreen	Food
ArmChair	57	Moveable, Receptacle	0.24m^3-1.98m^3	0.66m-1.53m	28.30kg		Fabric, Leather, Metal, Wood
BaseballBat	5	CanPickup	0.000m^3-0.007m^3	0.43m-0.96m	0.90kg	Black, Brown, Gray, Orange, Silver, Tan	Metal, Plastic, Wood
BasketBall	2	CanPickup	0.01m^3-0.01m^3	0.24m-0.24m	0.60kg		Rubber
Bed	38	CanBeDirty, Moveable, Receptacle	0.86m^3-7.34m^3	1.65m-2.52m	30.00kg-40.00kg		Fabric, Metal, Wood
Blinds	1	CanOpen, Static	0.006m^3	1.10m	1.00kg		
Book	35	CanOpen, CanPickup	0.000m^3-0.009m^3	0.19m-0.43m	0.50kg		Paper
Boots	1	CanPickup	0.03m^3	0.52m	0.70kg		Plastic
Bottle	1	CanBeFilled, CanBreak, CanPickup, CanSeeThrough	0.003m^3	0.25m	0.20kg		Glass
Bowl	31	CanBeDirty, CanBeFilled, CanBreak, CanPickup, CanSeeThrough, Receptacle	0.002m^3-0.006m^3	0.15m-0.25m	0.47kg	Gray, Orange, Tan, Violet, White	Ceramic, Glass, Plastic
Box	34	CanOpen, CanPickup, Receptacle	0.010m^3-0.159m^3	0.26m-0.69m	0.14kg-0.30kg	Brown, Tan, White	Paper
Bread	30	CanBeSliced, CanPickup	0.004m^3-0.010m^3	0.22m-0.39m	0.70kg	Brown, Gray, Tan	Food
ButterKnife	2	CanPickup	0.000m^3-0.000m^3	0.24m-0.27m	0.08kg		Metal
CD	1	CanPickup	0.000m^3	0.10m	0.01kg		Plastic
Cabinet	6	Static	0.19m^3-0.80m^3	0.85m-0.98m	1.00kg		
Candle	9	CanPickup, CanToggleOnOff	0.000m^3-0.004m^3	0.08m-0.38m	0.20kg	BlueGreen, Bronze, Gold, Gray, Pink, Tan, White	Metal, Wax
Cart	1	Receptacle, Static	0.49m^3	1.28m	1.00kg		
CellPhone	9	CanBreak, CanPickup, CanToggleOnOff	0.000m^3-0.000m^3	0.11m-0.18m	0.16kg	Black, Blue, Orange, Red, White	Glass, Metal, Plastic
Chair	75	Moveable, Receptacle	0.11m^3-0.58m^3	0.67m-1.24m	10.00kg-15.00kg		Fabric, Metal, Plastic, Wood
Cloth	12	CanBeDirty, CanPickup	0.001m^3-0.002m^3	0.21m-0.25m	0.05kg	Black, BlueGreen, Green, Pink, Red, Violet, White, Yellow	Fabric
ClothesDryer	2	Moveable, Receptacle	0.47m^3-0.52m^3	0.96m-1.14m	1.00kg		
CoffeeMachine	30	CanToggleOnOff, Moveable, ObjectSpecificReceptacle, Receptacle, isHeatSource	0.06m^3-0.12m^3	0.42m-0.54m	5.00kg	Black, Blue, Gray, Red, Silver, Tan, White	Plastic
CoffeeTable	54	CanSeeThrough, Moveable, Receptacle	0.05m^3-0.77m^3	0.57m-1.77m	11.00kg-13.00kg		Glass, Metal, Wood
CounterTop	33	Receptacle, Static	0.35m^3-8.78m^3	0.94m-3.07m	1.00kg		
CreditCard	4	CanPickup	0.000m^3	0.09m	0.01kg		Plastic
Cup	31	CanBeDirty, CanBeFilled, CanBreak, CanPickup, CanSeeThrough, Receptacle	0.001m^3-0.003m^3	0.10m-0.22m	0.40kg		Ceramic, Glass, Paper, Plastic
Desk	63	Moveable, Receptacle, Static	0.25m^3-3.37m^3	0.83m-2.53m	34.00kg-45.00kg		Metal, Plastic, Stone, Wood
DeskLamp	15	CanPickup, CanToggleOnOff, Moveable	0.008m^3-0.203m^3	0.31m-0.85m	2.06kg		Fabric, Metal
Desktop	2	CanBreak, CanToggleOnOff, Moveable	0.12m^3-0.12m^3	0.88m-0.88m	2.54kg		Plastic
DiningTable	48	CanSeeThrough, Moveable, Receptacle	0.41m^3-2.15m^3	0.79m-2.81m	85.00kg		Fabric, Glass, Metal, Plastic, Wood
DishSponge	1	CanPickup	0.000m^3	0.14m	0.03kg		Sponge
DogBed	5	Moveable, Receptacle	0.06m^3-0.26m^3	0.73m-1.18m	0.98kg		Fabric
Doorframe	20	Static	0.18m^3-0.61m^3	2.07m-2.13m	1.00kg		
Doorway	20	CanOpen, Static	0.21m^3-1.02m^3	2.07m-2.13m	1.00kg		
Dresser	46	Moveable, Receptacle, Static	0.14m^3-2.26m^3	0.70m-2.58m	57.00kg		Metal, Wood
Dumbbell	2	CanPickup	0.006m^3	0.33m	4.54kg		Metal
Egg	30	CanBeSliced, CanBreak, CanPickup	0.000m^3-0.001m^3	0.06m-0.12m	0.05kg	Brown, Tan, White	Food
EggCracked	30	CanBeCooked, CanPickup	0.000m^3-0.001m^3	0.11m-0.18m	0.05kg		Food
Faucet	37	CanToggleOnOff, Static	0.001m^3-0.076m^3	0.11m-0.90m	1.00kg		
FloorLamp	32	CanToggleOnOff, Moveable	0.06m^3-0.75m^3	1.19m-2.09m	1.00kg-3.93kg		Fabric, Metal, Wood
Footstool	10	CanPickup, Receptacle	0.02m^3-0.19m^3	0.28m-0.61m	1.00kg-4.00kg		Fabric, Leather, Wood
Fork	2	CanPickup	0.000m^3-0.000m^3	0.21m-0.25m	0.04kg		Metal
Fridge	30	CanOpen, Receptacle, Static, isColdSource	0.60m^3-2.58m^3	1.40m-2.19m	1.00kg		
GarbageBag	3	Moveable	0.15m^3-0.15m^3	0.56m-0.56m	3.50kg		Plastic
GarbageCan	31	CanPickup, Moveable, Receptacle	0.03m^3-0.17m^3	0.38m-0.68m	0.70kg	Black, Blue, Gold, Gray, Green, Red, Silver, White	Metal, Plastic
HandTowel	1	CanPickup	0.002m^3	0.40m	0.08kg	White	Fabric
HandTowelHolder	1	ObjectSpecificReceptacle, Receptacle, Static	0.002m^3	0.21m	1.00kg		
HousePlant	31	CanBeFilled, CanPickup, Moveable	0.001m^3-0.336m^3	0.13m-1.10m	3.00kg		Ceramic, Organic
Kettle	1	CanBeFilled, CanOpen, CanPickup	0.008m^3	0.21m	0.80kg		Metal, Plastic
KeyChain	3	CanPickup	0.000m^3-0.000m^3	0.13m-0.14m	0.08kg		Metal, Plastic
Knife	7	CanPickup	0.000m^3-0.000m^3	0.32m-0.34m	0.18kg		Metal, Plastic
Ladle	3	CanPickup	0.001m^3-0.001m^3	0.27m-0.31m	0.08kg	Black, Blue, Gold, Silver	Metal, Plastic
Laptop	31	CanBreak, CanOpen, CanPickup, CanToggleOnOff	0.02m^3-0.07m^3	0.32m-0.50m	2.30kg		Glass, Metal, Plastic
LaundryHamper	5	CanOpen, Moveable, Receptacle	0.10m^3-0.30m^3	0.60m-0.88m	1.55kg		Fabric
Lettuce	30	CanBeSliced, CanPickup	0.003m^3-0.008m^3	0.20m-0.27m	0.47kg	Green, YellowGreen	Food
LightSwitch	30	CanToggleOnOff, Static	0.000m^3-0.002m^3	0.12m-0.24m	1.00kg		
Microwave	30	CanOpen, CanToggleOnOff, Moveable, Receptacle, isHeatSource	0.07m^3-0.11m^3	0.61m-0.70m	7.00kg	Black, Bronze, Gray, Red, Silver	Glass, Metal
Mug	5	CanBeDirty, CanBeFilled, CanBreak, CanPickup, Receptacle	0.001m^3-0.001m^3	0.11m-0.14m	1.00kg		Ceramic
Newspaper	5	CanPickup	0.000m^3-0.002m^3	0.23m-0.38m	0.20kg		Paper
Ottoman	8	Moveable, Receptacle	0.17m^3-0.31m^3	0.77m-0.86m	10.00kg	Brown, Red, Violet, White	Fabric, Wood
Painting	26	Static	0.004m^3-0.032m^3	0.53m-1.23m	1.00kg		Paper, Plastic, Wood
Pan	30	CanBeDirty, CanPickup, Receptacle	0.003m^3-0.013m^3	0.27m-0.46m	0.67kg	Blue, Bronze, Gray, Orange, Red, Silver	Metal
PaperTowelRoll	1	CanBeUsedUp, CanPickup	0.001m^3	0.21m	0.22kg		Paper
Pen	5	CanPickup	0.000m^3-0.000m^3	0.14m-0.20m	0.01kg		Metal, Plastic
Pencil	8	CanPickup	0.000m^3-0.000m^3	0.09m-0.20m	0.01kg		Wood
PepperShaker	3	CanPickup	0.000m^3-0.001m^3	0.10m-0.20m	0.14kg-0.24kg		Glass, Metal
Pillow	32	CanPickup	0.01m^3-0.07m^3	0.33m-0.70m	0.70kg	Black, Blue, Brown, Gray, Green, Orange, Red, RedOrange, Tan, Violet, White, Yellow	Fabric
Plate	31	CanBeDirty, CanBreak, CanPickup, Receptacle	0.001m^3-0.002m^3	0.19m-0.29m	0.62kg	Black, Blue, Gold, Gray, White	Ceramic
Plunger	3	CanPickup	0.01m^3	0.54m	1.00kg	Black, BlueGreen, Brown, Red	Rubber, Wood
Pot	30	CanBeDirty, CanBeFilled, CanPickup, Receptacle	0.007m^3-0.036m^3	0.25m-0.54m	0.57kg	Black, Bronze, Gray, Silver	Metal
Potato	30	CanBeCooked, CanBeSliced, CanPickup	0.000m^3-0.002m^3	0.08m-0.15m	0.18kg	Brown, Gray, RedOrange, Tan, YellowOrange	Food
RemoteControl	4	CanPickup	0.000m^3-0.000m^3	0.19m-0.22m	0.15kg		Metal, Plastic
RoomDecor	3	Moveable	0.11m^3-0.19m^3	1.23m-1.45m	2.72kg		Ceramic, Organic
Safe	3	CanOpen, Moveable, Receptacle	0.03m^3-0.14m^3	0.33m-0.56m	10.00kg		Metal
SaltShaker	3	CanPickup	0.000m^3-0.001m^3	0.10m-0.20m	0.40kg-0.60kg		Glass, Metal
ScrubBrush	1	CanPickup	0.008m^3	0.48m	0.19kg		Plastic
ShelvingUnit	24	Moveable, Receptacle	0.07m^3-1.25m^3	0.69m-2.21m	20.00kg-22.00kg		Metal, Paper, Wood
ShowerCurtain	2	CanOpen, Static	0.10m^3-0.31m^3	1.39m-1.94m	1.00kg		
ShowerHead	2	CanToggleOnOff, Static	0.03m^3-0.15m^3	1.00m-1.81m	1.00kg		
SideTable	128	CanSeeThrough, Moveable, Receptacle, Static	0.03m^3-0.75m^3	0.39m-1.83m	11.00kg		Glass, Metal, Plastic, Wood
Sink	30	Receptacle, Static	0.02m^3-0.55m^3	0.46m-1.37m	1.00kg		
SoapBar	1	CanPickup	0.000m^3	0.09m	0.11kg		Soap
SoapBottle	30	CanBeUsedUp, CanPickup	0.001m^3-0.005m^3	0.17m-0.31m	0.40kg		Glass, Metal, Plastic
Sofa	54	Moveable, Receptacle	1.02m^3-4.65m^3	1.55m-2.86m	104.00kg		Fabric, Leather, Wood
Spatula	1	CanPickup	0.000m^3	0.30m	0.06kg		Metal, Plastic
Spoon	1	CanPickup	0.000m^3	0.21m	0.04kg		Metal
SprayBottle	9	CanPickup	0.001m^3-0.002m^3	0.25m-0.26m	0.11kg	Black, Blue, BlueGreen, Green, Orange, Violet, White, Yellow	Plastic
Statue	26	CanBreak, CanPickup	0.007m^3-0.015m^3	0.27m-0.59m	1.00kg	Black, Bronze, Brown, Gold, Gray, Green, Orange, Silver, White	Metal, Stone
Stool	15	Moveable, Receptacle	0.04m^3-0.17m^3	0.45m-0.76m	3.18kg		Metal, Plastic, Wood
StoveKnob	3	CanToggleOnOff, Static	0.000m^3-0.000m^3	0.06m-0.08m	1.00kg		
TVStand	20	Moveable, Receptacle	0.21m^3-0.76m^3	0.90m-1.90m	11.00kg		Glass, Metal, Plastic, Wood
TableTopDecor	2	CanPickup, CanSeeThrough	0.02m^3-0.02m^3	0.40m-0.40m	5.40kg		Glass
TeddyBear	2	CanPickup	0.07m^3-0.07m^3	0.47m-0.47m	0.90kg	Black, Brown, Tan, White	Fabric
Television	32	CanBreak, CanPickup, CanToggleOnOff, Moveable, Static	0.05m^3-0.63m^3	0.82m-1.70m	9.83kg		Glass, Metal, Plastic
TennisRacket	6	CanPickup	0.005m^3-0.007m^3	0.54m-0.65m	0.31kg	Black, Blue, Bronze, Green, Orange, Red, Silver, Tan, White	Metal, Plastic
TissueBox	1	CanBeUsedUp, CanPickup	0.003m^3	0.20m	0.35kg		Paper
Toaster	30	CanToggleOnOff, Moveable, ObjectSpecificReceptacle, Receptacle, isHeatSource	0.01m^3-0.03m^3	0.34m-0.47m	5.00kg	Black, BlueGreen, Brown, Gray, Green, Red, Silver, White	Metal
Toilet	2	CanOpen, Receptacle, Static	0.53m^3-1.23m^3	1.04m-1.16m	1.00kg		
ToiletPaper	2	CanBeUsedUp, CanPickup	0.000m^3-0.001m^3	0.12m-0.12m	0.20kg		Paper
ToiletPaperHanger	30	ObjectSpecificReceptacle, Receptacle, Static	0.000m^3-0.008m^3	0.19m-0.30m	1.00kg		
Tomato	30	CanBeSliced, CanPickup	0.001m^3-0.003m^3	0.10m-0.15m	0.12kg	Red	Food
Towel	3	CanPickup	0.01m^3-0.01m^3	0.59m-0.61m	0.57kg	Gray, Tan, White	Fabric
TowelHolder	1	ObjectSpecificReceptacle, Receptacle, Static	0.003m^3	0.66m	1.00kg		
VacuumCleaner	2	Moveable	0.11m^3-0.11m^3	1.01m-1.01m	6.80kg		Glass, Metal, Plastic, Rubber
Vase	20	CanBreak, CanPickup, CanSeeThrough	0.002m^3-0.009m^3	0.15m-0.29m	1.00kg	Blue, Gold, Gray, Green, Red, White, Yellow	Ceramic, Glass, Metal
WashingMachine	2	Moveable, Receptacle	0.47m^3-0.52m^3	0.96m-1.14m	1.00kg		
Watch	2	CanPickup	0.000m^3-0.000m^3	0.08m-0.08m	0.07kg	Brown, RedOrange, Silver, White	Glass, Metal, Plastic
WateringCan	1	CanBeFilled, CanPickup	0.008m^3	0.27m	1.00kg		Plastic
Window	21	Static	0.07m^3-1.32m^3	0.92m-2.44m	1.00kg		
WineBottle	1	CanBeFilled, CanBreak, CanPickup	0.003m^3	0.33m	1.20kg		Glass

Table 4. Table of all object types along with summaries of the attributes of instances of those object types in AI2-THOR.

Index	Template
1	What number of [C2] [M2] [S2]s are <R> the [C1] [M1] [S1]?
2	Are there any [C2] [M2] [S2]s <R> the [C1] [M1] [S1]?
3	What color is the [M2] [S2] [that is] <R> the [C1] [M1] [S1]?
4	What is the material of the [C2] [S2] [that is] <R> the [C1] [M1] [S1]?
5	What is the [C2] [M2] object [that is] <R> the [C1] [M1] [S1]?
6	Is there anything else that has the same color as the [C1] [M1] [S1]?
7	Is there anything else that has the same material as the [C1] [M1] [S1]?
8	Do the [M1] [S1] and the [M2] [S2] have the same color?
9	Do the [C1] [S1] and the [C2] [S2] have the same material?
10	Do the [M1] [S1] <R> the [C3] [M3] [S3] and the [M2] [S2] have the same color?
11	Do the [C1] [S1] <R> the [C3] [M3] [S3] and the [C2] [S2] have the same material?

Table 5. Eleven different question families. Ci, Mi, and Si $(i-1,2,3)$ are one of the colors, materials, and object types listed in Tab. 4. [] denotes that word in the parenthesis can be omitted(Si $(i=1,2,3)$ can be changed to "object"). R denotes one of the relations specified in Tab. 6

Index	Relation / Property	Explanation
1	Color	the most salient one or two colors of the object
2	Material	the most salient one to three materials of the objct
3	ContainedBy	the relationship describing an object is located in a receptacle
4	AdjacentTo	two objects are within 0.5 meters and no objects are on the line connecting the two
5	OnTopOf	One object's bounding box collides with another's on the bottom and top respectively
6	HeavierThan	an object's mass is higher than the other's by at least 0.05 kg
7	LighterThan	an object's mass is lower than the other's by at least 0.05 kg
8	LargerThan	an object's longest, middle and shortest dimensions are all longer than the other's longest, middle and shortest dimensions respectively
9	SmallerThan	an object's longest, middle and shortest dimensions are all shorter than the other's longest, middle and shortest dimensions respectively
10	LongerThan	for a few object types, including a baseball bat, candle, ladle, and plungers, if one object's longest dimension is longer than the other's
11	ShorterThan	for a few object types, including a baseball bat, candle, ladle, and plunger, if one object's longest dimension is shorter than the other's

Table 6. Relations and properties in EXCALIBUR.

E. Question families

F. Exploration score

We consider a normal distribution prior to the exploration score, i.e. ExQA $\sim \mathcal{N}(\mu, \sigma)$, and maximize the likelihood of the scores of human annotators

$$k^* = \arg\max_{k} \max_{\mu,\sigma} \mathbb{E}_{\text{ExQA}}[-\frac{1}{2}\left(\frac{\text{ExQA}-\mu}{\sigma}\right)^2 - \log\sigma]. \tag{5}$$

G. Baseline Model

Our baseline model is a GRU [13] f_ϕ^{GRU} with parameters ϕ, which at each step t takes observation o_t, action at last time step a_{t-1}, and questions q as input, and outputs belief:

$$h_t = f_\phi^{\text{GRU}}(h_{t-1}, f_\psi^{\text{obs}}([\texttt{MLP}_1(\text{CLIP}(o_t)); \texttt{MLP}_2(\text{action-emb}(a_t)); \texttt{MLP}_3(\text{T5}^{\text{encoder}}(q)]])), \tag{6}$$

where f_ψ^{obs} and $\texttt{MLP}_i, i = 1, 2, 3$ are all three-layer MLPs, CLIP [48] encodes the observation frame at time step t, action-emb is an embedding layer, and the encoder of T5 [49] encodes a text string composed of questions $\mathcal{Q}$ after Phase II and outputs zero vectors in Phase II, since the questions are not available.

The belief h_t is used for two different purposes: (1) predict actions

$$a_t \sim \texttt{softmax}(f_\xi^{\text{action}}(h_t)), \tag{7}$$

where a_t is an action in the space shown in Fig. 2, and f_ξ^{action} is a three-layer MLP converting belief to action logits; (2) predict answers to questions

$$\text{answer}_i = \text{beam-search}(p_{\text{T5}}(a \mid [f_\theta^{\text{prefix}}(h_t), f^{\text{emb}}(q_i)])), \tag{8}$$

where f_θ^{prefix} is an MLP generating soft prompts for the frozen T5 similar to [60], and $f^{\text{emb}}(q)$ is the embedding layer of question.

H. Further Analysis of Results

To better understand the different failing patterns of humans and reinforcement learning agents, we study the following four metrics as well

1. TTI – number of time steps till touching an object for the first time,
2. Area – the ratio of regions covered by the agents,
3. % Objects Seen – the ratio of objects that are seen,
4. and % Objects Interacted – the ratio of objects that are touched.

Where do humans go wrong? Human failures mainly result from forgetting, and model failures mainly result from a lack of exploration. Through comparing humans w/o replay and w/ replay we can confirm the first part, and through comparing human area, we can clearly see that when exposed to an immersive setup, an average user covers more area, hence performing better. Also, when we compare Phase III to Phase I, the performance definitely improves as the humans now know what exactly to observe. Another reason why humans lag in the first phase is because humans tend to remember objects only when they have a pre-defined goal in mind. A general exploration, doesn't trigger the humans to learn every thing they see. Which also explains why Phase III performance is better than Phase I.

Relation of % Objects seen/interacted to Performance: We see an increase in the number of Objects Seen and Objects Interacted for human performance, which is again easier to do so in the VR environment. When the humans to able to interact with a greater number of items, they firstly make better idea of the things placed in the setup. Also, interaction stimulates the memory, making them remember which objects they interacted on each of the rooms. Overall, effect of these help them get a better spatial mapping and they tend to perform better. If we closely observe the results, better performance has a direct relation to the percentage of objects seen or interacted.

TTI v/s Accuracy: An interesting observation is that for human performance, the TTI in the third phase increases compared to the first phase as the humans know which part of the house to explore and which object to interact with. Due to this, human tends to avoid interacting with all objects and checks only the relevant ones. Also, TTI for humans after watching the replay is higher than without replay case because after watching the replay, re-interaction is required only for a few objects, which explains the increase in TTI. Furthermore, humans intelligently explore those parts of the house which they missed during Phase I exploration, thereby reducing the area covered in the re-entering phase.

	Phase I Exploration				Phase III Reentering			
	TTI	Area	% Obj Seen	% Obj Inter	TTI	Area	% Obj Seen	% Obj Inter
QA reward	303.5	11.2	23.4	**10.4**	202.3	21.2	15.1	24.9
Novelty reward	571.2	31.4	33.5	5.6	172.1	23.2	23.0	3.2
Novelty+QA	436.1	**32.8**	**37.9**	9.9	153.2	23.6	20.5	25.8
Human w/o replay	53.7	59.3	96.7	83.8	65.7	28.5	83.1	54.0
Human w/ replay	68.5	70.7	98.5	81.2	154.3	26.6	75.4	39.4

Table 7. Analysis on exploration and reentering behavior of agents and humans.