

A. Full Set of Background Training Tasks

Like the evaluation tasks, each background training task consisted of eight familiarization trials and one test trial. Each task type is described below.

a. Occlusion: Object Collection As shown in Figure A.1a, an agent navigates efficiently, avoiding walls and a rotating spinner, to reach a goal object, which changes color at impact. A grey square-shaped occluder appears under all elements in familiarization and above at the last trial.

b. Occlusion: Bouncing Object As shown in Figure A.1b, a spinner rotates and comes into contact with a passive object. At impact, the object which was previously still begins moving in a straight line perpendicular to the spinner. The object stops moving when it hits a wall or when it contacts another goal object, which changes color at impact. A grey square-shaped occluder lays under all elements in familiarization and on top of them at test time.

c. Occlusion: No-Navigation Preference As illustrated in Figure A.1c, an agent consistently reaches one of two proximal target objects, with varying locations across trials, demonstrating a preference to a goal object, instead of a goal location. Upon contact, both objects change color, eliciting cues for causal efficacy that distinguish goal-approach behavior from a pro-social approach. A spinning object is in the scene and do not interact with other elements. A grey square-shaped occluder lays under all elements in familiarization and on top of them at test time.

d. Occlusion: No-Navigation Bouncing Object As shown in Figure A.1d, a rotating spinner propels an element toward one of two target objects, which change color upon contact. The target object varies between trials, depending on the spinner's initial rotation and the element's position, illustrating that passive movement driven by another object does not necessarily signify goal preference.

e. Single Object Collect An agent navigates through obstacles to reach a single target object, which changes color upon contact (Figure A.1e). This task demonstrates obstacle navigation and goal pursuit. In some but not all episodes, the agent returns to the starting point after interacting with the object.

f. Moving Object As shown in Figure A.1f, an agent approaches a goal object and picks it up, which appears as the agent above and in the center of the object. The agent and object pair then travel together to a different location. After they stop moving, the agent drops off the object. This task shows how an object can be picked up and moved around by an agent in the environment.

g. & j. No-Navigation Approach: Instrumental & Social

This task involves three agents and a goal object. In the familiarization phase, an agent consistently approaches one of two stationary agents in proximity, demonstrating a social affiliation (left panels of Figure A.1g&j). At test time, the two previously stationary agents each move in a unique movement. Following that, the previously approaching agent also performs a movement. In No-Navigation Approach: Instrumental (right panel of Figure A.1g, the movement is identical to one of the other two agents which might or might not have been approached before. In addition, the agent interact with a goal object at the end of the move, causing it to change color. The route taken is the only efficient path to the goal object, cueing a goal-directed behavior. In No-Navigation Approach: Social (right panel of Figure A.1j), the previously approaching agent moves in the same way as the approached agent in familiarization.

h. & k. No-Relocation: True & False Belief This task features an occluder that might or might not appear between a goal-directed agent and its goal (Figure A.1h&k). Specifically, it is always in between on the first familiarization trial, obstructing the agent from identifying the side of the room where the goal is located. The agent randomly enters a room and either succeeds in or fails to find the goal object, giving machines an opportunity to learn that the occluder interferes with the perception of the rational goal-directed agent. If the agent finds the goal object in the first familiarization trial, it would consistently go to the same side of the room to find the object in the later familiarization trials, where the occluder randomly shows up between the agent and object or behind the environment. If the agent did not find the goal object in the first familiarization, it would later consistently go to the other side of the room to find the goal object. This task communicates two assumptions: 1) the agent believes a goal object is always at where it was last seen. 2) if the goal object is absent on one side of the room, it must be on the other side. At test time, a different agent moves the goal object around within the same room, either in the absence (False Belief, Figure A.1h) or the presence (True Belief, Figure A.1k) of the first agent. The task then shows the initial agent returning to the same room to retrieve the goal object.

j. & i. Trapped Agent: Helping & Hindering This task (Figure A.1j&i) employs a similar setup as the Helping task and the Hindering task in evaluation (Figure 3). However, the locations of the goal-directed agent and the goal object are swapped. In Helping (Figure A.1j), a social agent move away an obstacle to help the main agent approach its goal object while the other social agent does nothing. In Hindering (Figure A.1i), a social agent blocks the goal-directed agent from approaching its goal while the other social agent watches. At test, the goal-directed agent approaches the agent who helps, or the agent who does nothing when the other agent hinders.

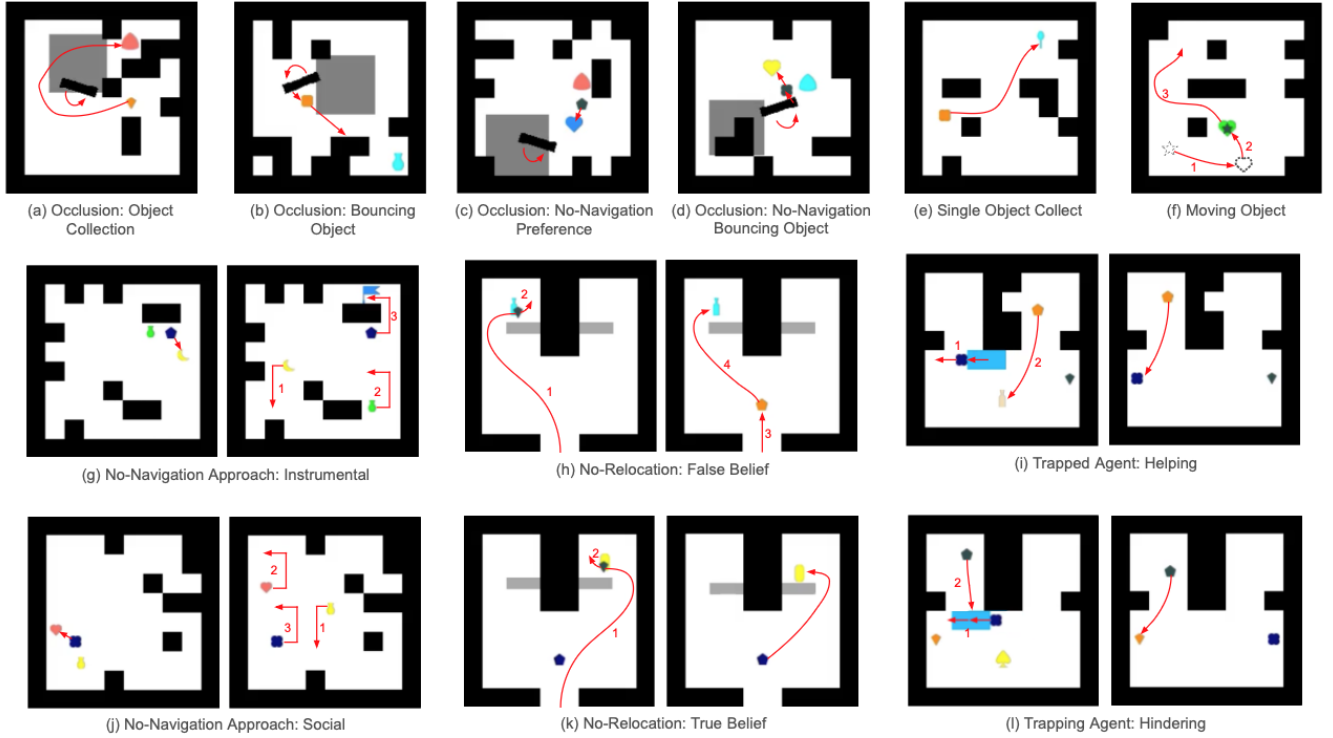


Figure A.1: **Overview of All Background Training Tasks.** AI models undergo training across various tasks to capture environmental dynamics and essential cognitive abilities for solving the benchmark. This figure illustrates the schematics of both familiarization and test trials. For tasks (a)-(f), a single panel depicts each task, where the familiarization trials and a test trial are drawn from the same distribution. For tasks (g)-(l), the left panel represents a familiarization trial, while the right panel illustrates a test trial. Red arrows indicate the direction of movement, and, where applicable, numbers indicate the orders at which the actions take place.

B. Model Hyperparameters

Each video frame is resized to 84×84 pixels and then split into 49 patches of 12×12 pixels. A three-layer CNN creates a 256-dimensional embedding for each patch, with sinusoidal positional encoding. The row, column, frame, and trial numbers are separately encoded and each takes up 64 dimensions of the positional encoding. The Transformer encoder and decoder each has eight attention heads and five layers. The model is trained on two A100 GPUs for 100 epochs with a batch size of 32. The learning rate and weight decay are both set at 1×10^{-4} .