SUPPLEMENTARY MATERIAL FOR CAN NEURAL NETWORKS LEARN IMPLICIT LOGIC FROM PHYSICAL REASONING?

Anonymous authors
Paper under double-blind review

A SUMMARY OF DATASETS

Here is a brief recap of all of the datasets that we use in the main paper. All datasets contain 5,000 training examples. The Two-Cup Task test is 100 test examples.

Dataset Name	Introduced in	Expt. 1	Expt 2.	Expt 3	Visualized in
Two-Cup Task	§3.1	Test	Test	Finetuning	Main paper, Fig. 1
Physical Reasoning	§3.1.2	Training	Training	Pretraining	Appendix, Fig. 1
Two-Cup Ablations	§3.1.2	Training	Training	n/a	Appendix, Fig. 2, Sec G
Invisible Displacement	§3.1.2	n/a	Training	Pretraining	Main Paper, Fig. 2

Table 1: Quick summary of the datasets and where to find them

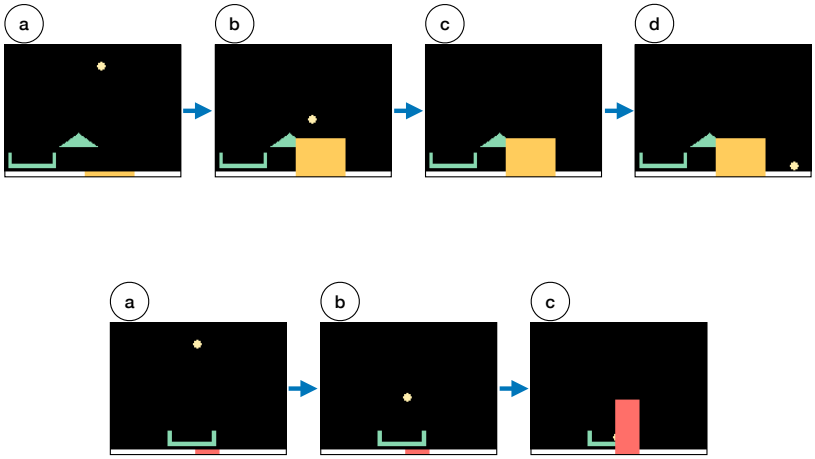

Figure 1: Frames from two example videos of Physical Reasoning Schema training data. The location of the ball and occluders is random throughout the scene, and cups and the wedge are centered around at a random x-location according to one of several templates.

B ADEPT

ADEPT maintains a set of k particles $P^t = \bigcup_{i=0}^{k} p_i^t$, where each p_i^t is a representation of the complete scene at timestep t. The next step p_i^{t+1} of each particle is then generated using an internal physics simulator. The internal physics simulator is noisy, and at each step a small amount of random noise drawn from a Gaussian with mean 0 and standard deviation σ_{loc} is added to the position of the

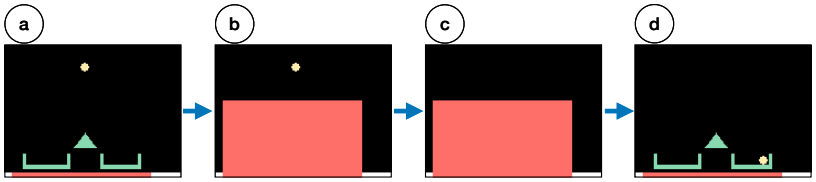

Figure 2: Frames from an example video from the Two-Cup Ablation dataset. This is a modification of the events in the Two-Cup Test scene. There is no reasoning by exclusion inference in this scene; while there is ambiguity at frame `c` about which cup the ball has fallen in, the only information provided by the scene is frame `d`, where the ball is directly observed.

ball. Every N steps, the particles are resampled proportional to their likelihood, where likelihood is generated in the following way: first, a score for each particle is generated based on the overlap between the bounding boxes it predicts $p_i^t \in P^{t+1}$ and those that are observed x^{t+1}. Then, each particle is given a penalty to its score if there is a mismatch between the observation and the beliefs represented by that particle (e.g. the observation reveals empty space where the particle represented the ball). The likelihood of each particle is then calculated as the negative log likelihood of its score.

C Model Training

The models were implemented in PyTorch and took roughly 6 hours to converge on TitanV, TitanRTX, and QuadroRTX GPUs. We used the original model parameters from the OPNet paper for the experiments– 256 neurons for the hidden dimension in the "who to track" layer, and 512 neurons for the learned attention model. We experimented both with smaller models (64 and 128 neurons respectively) and with much larger models (1024 and 4096 neurons respectively), as well as with stacking LSTMs, and did not find that they affected our final results. We stopped models after they had trained for 30 epochs without hitting a new minimum loss on the dev set.

D Controls

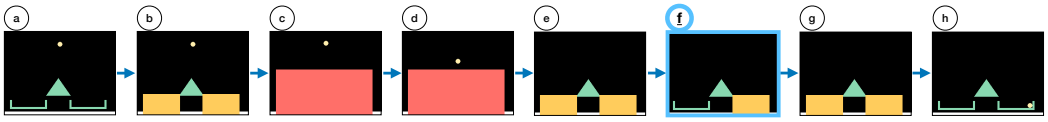

Figure 3: Figure from the paper of frames from the Two-Cup task test video. Frames `a-d` set up the scene and create ambiguity between the two possible hiding locations for the ball, and frames `e-h` are the essential part of the task on which we evaluate the model. The goal is to determine the location of the small yellow ball, which drops at frame `d`. At frame `e`, the subject should know that the ball is either in the left cup or the right cup, but its exact location is unknown. When the left occluder drops at frame `f`, the subject should know that the ball is *not* in the left cup, and *must* be in the right cup, despite not seeing it. This is the critical frame.

D.1 Two-Cup Freebie

This control is the same as the Two-Cup Task, except the ball is in the cup that is revealed in frame `f` of Fig. 1. This control is an important contrast to the Two-Cup Task because success helps establish that there is nothing about the environment in the setup sequence `a` to `e` that is causing the model's

failure on the Two-Cup Task other than the ambiguity of the ball's location between cups in the frame e.

D.2 CONTROL EXAMPLES: REORDERING EVENTS

The setup sequence of the Two-Cup Task contains four events: Small occluders rising (frame b), large occluder rising (frame c,), the ball falling (frame d), and the large occluder falling (frame e). Altering the order of these frames can give us insight into the model's performance.

The large occluder must fall before the small occluders so that each event is observed by the model, so frame (e) cannot be moved in the sequence.

There are six combinations of the remaining events in the setup sequence small rise (S), large rise (L), and ball fall (B):

- SLB: This is the Two-Cup Task.
- LSB: The large occluder rises before the small occluders do and then drops after the small occluders. Because S is never seen, this is equivalent to LB, which is an item in the Two-Cup Ablations training set, and is not a control.
- LBS: Same
- BLS: Same, but this is equivalent to BL
- **BSL**: The ball falls first, and then the rest of the scene plays out as it does in the Two-Cup Task. This means that the model has full information on the ball's location before any of the occluders are on the screen. Success on this variation implies that if the model has full knowledge of the ball's location, then that persists even if the ball is occluded and the rest of the Two-Cup Scene plays out. (This is Known Specific Location in the paper).
- **SBL**: The small occluders rise, then the ball falls, then the rest of the scene plays out as it does in the Two-Cup Task. This means that the model will see which cup the ball is in, but it will not observe exactly where in the cup the ball lands. Success on this variation implies that if the model knows the ball's location generally (although with some uncertainty), then that persists as the rest of the Two-Cup Scene plays out. (This is Known Cup Location in the paper).

Thus we include SBL and BSL as controls.

D.3 CONTROL: TRAIN ON TWO-CUP

We set up the training data carefully to exclude a) the events in the Two-Cup Task that include the critical frame, disambiguating sequence, and the frames in between, and b) any scene that can be solved by making an inference as to the ball's location. These constraints are important to ensure that the inference at the critical sequence of the Two-Cup Task is out of distribution, but the elements of the scene (the ball, cups, wedges, occluders) are not.

It is thus important to determine whether the Two-Cup Task can be solved by a model if it is within its training distribution.

For this control, we create a new training set of 5000 videos of the Two-Cup Task. This test set contains examples of the Two-Cup Task and examples from the Two-Cup freebie control. We then train two new models: one trained on Two-Cup alone, and one trained on both Two-Cup and Physical Reasoning. (This is done so that in that setting the model also views examples of the ball interacting in the environment beyond the Two-Cup Task.)

D.3.1 CONTROL: TWO-CUP TASK, BUT THE BALL FALLS OUTSIDE

It may be the case that a model trained on the Two-Cup Task only expects the ball to be in one of the two cups after the critical frame no matter where it sees the ball fall. We include two separate controls: one where the ball falls outside the cup and occluder structure and is visible for the entire sequence, and one where it falls outside of the cup and occluder structure but is obscured by its own occluder until the critical frame. Note that there is no "correct" cup on this task, as both cups

Train on Two-Cup task, test on Two-Cup task			
Dataset	Correct	Incorrect	Elsewhere
Two-cup task	0.81	0.19	0.00
Train on Physical Reasoning + Two-Cup Ablations, test on Two-Cup task			
Dataset	Correct	Incorrect	Elsewhere
Two-cup task	0.48	0.50	0.02
Freebie	1.00	0.00	0.00
Known Specific Location	1.00	0.00	0.00
Known Cup Location	0.94	0.03	0.03
Outside (visible)	n/a	0.00	1.00
Outside (occluded)	n/a	0.12	0.88

Table 2: Results on the Two-Cup task and the control datasets. The models' predictions before the critical frame are always split roughly evenly between the two cups, when relevant. In the Outside experiments, "elsewhere" is the correct answer, because the ball is outside of both cups.

are incorrect. Failure on either of these controls would indicate that the model has memorized the Two-Cup Task and cannot generalize to the full physics of the environment.

E RESULTS

Results are visible in Table 1.

Two-cup: Our experimental model fails on the Two-Cup Task. Before the critical sequence, it expects the ball to be in either cup with about 50% accuracy, and after the critical sequence, it does not update their predictions in accordance with the inference that is possible. (A small number of predictions move from the incorrect cup to elsewhere; however, these predictions move by mere pixels, and are simply moving on the border of the incorrect cup.)

Our control model that includes Two-Cup examples in its training data sees a performance gain on the Two-Cup Task after the critical frame is shown. This is a gain of roughly 31%, totaling 81%. In the example where the model predicts the ball "elsewhere", it moves its prediction *near* the correct cup, but outside of it on the edge of the occluder (just far enough to not be counted as "in" the cup). In the remaining 19% of examples, upon inspection, the model continues to pick the incorrect cup, refusing to update its prediction as the Physical Reasoning Schema+Two-Cup Ablations model does.

Two-cup freebie: The experimental models correctly predict the location of the ball both before and after the critical frame.

Known Specific Location: The experimental models correctly predict the location of the ball both before and after the critical frame.

Known Cup Location: The experimental models nearly always correctly predict the location of the ball both before and after the critical frame.

Two-cup ball outside (visible ball): Both experimental models and control models always correctly predict the location of the ball both before and after the critical frame.

Two-cup ball outside (occluded ball): The Physical Reasoning + Two-Cup Ablations model performs the worst at this dataset out of all of the control sets. Upon inspection of these videos, the prediction of the ball's location changes to be in one of the cups. This is likely because the training data biases the model towards predicting that the ball will always land in a cup when it observes a Two-Cup Scene. When we trained a model on Physical Reasoning without Two-Cup ablations, it always succeeded at this dataset.

E.1 INVISIBLE DISPLACEMENT CONTROL EXPERIMENT RESULTS

In the "roll-then-occlude" control setting, the model fails 15% of the time before the critical frame. These errors are mostly that the model expects the ball to continue rolling once it has been occluded. This is likely because it has seen the "critical frame" in extremely similar situations during training

	After critical frame		
Dataset	Correct	Incorrect	Elsewhere
Two-cup task	0.48	0.50	0.02
Two-cup Freebie	0.99	0.01	0.00
Invisible Displacement Control	1.00	0.00	0.00
Invisible Displacement Freebie	0.99	0.01	0.00
Known Specific Location	0.99	0.01	0.00
Known Cup Location	0.99	0.01	0.00
Outside (visible)	n/a	0.00	1.00
Outside (occluded)	n/a	0.10	0.90
Roll-then-occlude	0.95	0.02	0.03

Table 3: Control results for model trained on Physical Reasoning + Two-Cup Ablations + Invisible Displacement.

many times, and during training, there was ambiguity as to the ball's location; so it expects the ball to keep rolling.

F Training on Two-Cup Task

For most of the results in this work, we report that the OPNet model does not perform above 50% accuracy (above chance) at the two-cup task. However, the ADEPT model is able to perform at 100% accuracy. After training directly on the Two-Cup task, the model performs at 81% accuracy in our reported results in both Experiments 1 and 3 (§4.1 and §4.3).

We explored multiple hyperparameter configurations while training and testing on the Two-Cup task in an attempt to maximize performance at a held-out Two-Cup test set. We explored learning rates of 0.001, 0.0001, 0.00001, hidden dimensions 64, 128, 256, and 512, and 1 and 2 stacked LSTM layers, as well as training dataset sizes of 5,000 and 10,000. The best accuracy reported on the held-out test set was 85%.

The OPNet model does not reach 100% accuracy at the task likely because the LSTM model structure does not lend itself well to dramatically changing the bounding box that it outputs. In the Two-Cup task, cups A and B are rather far apart in the scene (because the ball must fall on the wedge between them to create the ambiguity). In the Two-Cup task, when one cup is revealed to be empty, the model must switch its prediction across the screen to the other cup. It is able to accomplish this roughly 80% of the time; in the remaining 20%, upon one occluder falling to reveal that one cup is empty, the model does not update its prediction. This is in contrast to Invisible Displacement, where by nature of the task locations A and B are right next to each other, as well as in contrast to the training data from the original OPNet paper. OPNet models trained on the Invisible Displacement task score 100% at a held-out Invisible Displacement test set. However, models trained on Physical Reasoning, the Two-Cup Task, and/or Two-Cup Ablations do not perform above chance on a held-out Invisible Displacement task.

G Two-Cup Ablations Templates

#C	#SO	#LO	Rise order	Fall order	Allow	notes
2	2	1	B L S	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.

2	2	1	L S B	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	2	1	L B S	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	2	0	B S	N/A	yes	hangs frame on e
2	2	0	S B	N/A	yes	hangs frame on e
2	1	1	S L B (fall in empty)	L S	yes	distinct from frames e and f
2	1	1	S B L (fall in empty)	L S	yes	distinct from frames e and f
2	1	1	B S L (fall in empty)	L S	yes	distinct from frames e and f
2	1	1	B L S (falls behind S)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	1	1	B L S (fall in empty)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	1	1	L S B (falls behind S)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	1	1	L S B (fall in empty)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	1	1	L B S (falls behind S)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	1	1	L B S (fall in empty)	S L	yes	This is equivalent to two-cup without small occluders. The small occluders rise and fall without the model seeing them. This means there is a longer delay between rise and fall of the large occluder.
2	0	1	L B	L	yes	two-cup without small occluders, indistinct from frame d
2	0	1	B L	L	yes	two-cup without small occluders, but you saw the ball land before the occluder went up. Indistinct from frame d

2	0	1	L B	N/A	yes	hangs frame on d
2	0	1	B L	N/A	yes	hangs frame on d
1	2	2	S L B (rolls to side)	L S	yes	Although contains E, the scene should be entirely disambiguated.
1	2	2	S L B (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (There is no logical inference here)
1	2	2	S B L (rolls to side)	L S	yes	Although contains E, the scene should be entirely disambiguated.
1	2	2	S B L (lands in cup)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (We saw the ball go in the cup– there is no logical inference.)
1	2	2	S B L (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (There is no logical inference here)
1	2	2	B S L (rolls to side)	L S	yes	Although contains E, the scene should be entirely disambiguated.
1	2	2	B S L (lands in cup)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (We saw the ball go in the cup– there is no logical inference.)
1	2	2	B S L (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (There is no logical inference here)
1	2	2	B L S (lands in cup)	S L	yes	disambiguated, no small occluders
1	2	2	B L S (rolls to side)	S L	yes	disambiguated, no small occluders
1	2	2	B L S (lands in cup)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (We saw the ball go in the cup– there is no logical inference.)
1	2	2	B L S (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (There is no logical inference here)
1	2	2	L S B (rolls to side)	S L	yes	two-cup with no small occluders, the scene should be entirely disambiguated.
1	2	2	L S B (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated. (There is no logical inference here)
1	2	2	L B S (rolls to side)	S L	yes	two-cup with no small occluders, the scene should be entirely disambiguated.
1	2	2	L B S (rolls to side)	N/A	yes	Freezes on frame D, the scene should be entirely disambiguated
0	2	2	S L B	L S	yes	Ball rolls to side, is totally disambiguated
0	2	2	S L B	N/A	yes	Ball rolls to side, is totally disambiguated
0	2	2	S B L	L S	yes	Ball rolls to side, is totally disambiguated
0	2	2	S B L	N/A	yes	Ball rolls to side, is totally disambiguated
0	2	2	B S L	L S	yes	Ball rolls to side, is totally disambiguated
0	2	2	B S L	N/A	yes	Ball rolls to side, is totally disambiguated
0	2	2	B L S	S L	yes	Never sees small occluders, ball rolls to side, is totally disambiguated

0	2	2	B L S	N/A	yes	Ball rolls to side, is totally disambiguated
0	2	2	L S B	S L	yes	Never sees small occluders, ball rolls to side, is totally disambiguated
0	2	2	L S B	N/A	yes	Ball rolls to side, is totally disambiguated
0	2	2	L B S	S L	yes	Never sees small occluders, ball rolls to side, is totally disambiguated
0	2	2	L B S	N/A	yes	Ball rolls to side, is totally disambiguated
2	2	1	S L B	L S (one)	no	overlaps with frames e-f-g
2	2	1	S B L	L S (one)	no	overlaps with frames e-f-g
2	2	1	B S L	L S (one)	no	overlaps with frames e-f-g
2	2	1	B L S	L S (one)	no	overlaps with frames e-f-g
2	2	1	L S B	L S (one)	no	overlaps with frames e-f-g
2	2	1	L B S	L S (one)	no	overlaps with frames e-f-g
2	2	0	B S	S (one)	no	overlaps with frames e-f-g
2	2	0	S B	S (one)	no	overlaps with frames e-f-g
2	1	1	S L B (falls behind S)	L S	no	overlaps with "real thing" at frame f in frame sequence e->f->g
2	1	1	S B L (falls behind S)	L S	no	overlaps with "real thing" at frame f in frame sequence e->f->g
2	1	1	B S L (falls behind S)	L S	no	overlaps with "real thing" at frame f in frame sequence e->f->g
2	1	1	B L S (falls behind S)	L S	no	1) overlaps with "real thing" at frame f in frame sequence e->f->g. 2) occluders appear out of nowhere.
2	1	1	L S B (falls behind S)	L S	no	1) overlaps with "real thing" at frame f in frame sequence e->f->g. 2) occluders appear out of nowhere.
2	1	1	L B S (falls behind S)	L S	no	1) overlaps with "real thing" at frame f in frame sequence e->f->g. 2) occluders appear out of nowhere.
1	2	1	S L B (lands in cup)	L S	no	overlaps with frames e-f-g
1	2	1	S L B (lands in cup)	N/A	no	This contains a logical inference! The ball isn't rolling out– it must be in the unseen cup.
1	2	1	S B L (lands in cup)	L S	no	overlaps with frames e-f-g
1	2	1	B S L (lands in cup)	L S	no	overlaps with frames e-f-g
1	2	1	L S B (lands in cup)	S L	no	This contains a logical inference! The ball isn't rolling out– it must be in the unseen cup.
1	2	1	L S B (lands in cup)	N/A	no	This contains a logical inference! The ball isn't rolling out– it must be in the unseen cup.
1	2	1	L B S (lands in cup)	L S	no	occluders appear out of nowhere, and degenerates to e->f->g alike
1	2	1	L B S (lands in cup)	S L	no	This contains a logical inference! The ball isn't rolling out– it must be in the unseen cup.
1	2	1	L B S (lands in cup)	N/A	no	This contains a logical inference! The ball isn't rolling out– it must be in the unseen cup.
1	2	1	S L B (lands in cup)	L S	no	overlaps with frames e-f-g
2	2	1	S L B	S L	no	occluders appear to disappear
2	2	1	S B L	S L	no	occluders appear to disappear
2	2	1	B S L	S L	no	occluders appear to disappear
2	1	1	S L B (falls behind S)	S L	no	occluders appear to disappear
2	1	1	S L B (fall in empty)	S L	no	occluders appear to disappear

2	1	1	S B L (falls behind S)	S L	no	occluders appear to disappear
2	1	1	S B L (fall in empty)	S L	no	occluders appear to disappear
2	1	1	B S L (falls behind S)	S L	no	occluders appear to disappear
2	1	1	B S L (fall in empty)	S L	no	occluders appear to disappear
2	1	1	B L S (fall in empty)	L S	no	occluders appear out of nowhere
2	1	1	L S B (fall in empty)	L S	no	occluders appear out of nowhere
2	1	1	L B S (fall in empty)	L S	no	occluders appear out of nowhere
1	2	1	S L B (lands in cup)	S L	no	occluders appear to disappear
1	2	1	S L B (rolls to side)	S L	no	occluders appear to disappear
1	2	1	S B L (lands in cup)	S L	no	occluders appear to disappear
1	2	1	S B L (rolls to side)	S L	no	occluders appear to disappear
1	2	1	B S L (lands in cup)	S L	no	occluders appear to disappear
1	2	1	B S L (rolls to side)	S L	no	occluders appear to disappear
1	2	1	B L S (lands in cup)	L S	no	occluders appear out of nowhere
1	2	1	B L S (rolls to side)	L S	no	occluders appear out of nowhere
1	2	1	L S B (lands in cup)	L S	no	occluders appear out of nowhere
1	2	1	L S B (rolls to side)	L S	no	occluders appear out of nowhere
1	2	1	L B S (rolls to side)	L S	no	occluders appear out of nowhere
0	2	1	S L B	S L	no	occluders appear to disappear
0	2	1	S B L	S L	no	occluders appear to disappear
0	2	1	B S L	S L	no	occluders appear to disappear
0	2	1	B L S	L S	no	occluders appear out of nowhere
0	2	1	L S B	L S	no	occluders appear out of nowhere
0	2	1	L B S	L S	no	occluders appear out of nowhere