Draping an Elephant: Uncovering Children's Reasoning About Cloth-Covered Objects

Tomer D. Ullman (tomeru@mit.edu)
Psychology, Harvard University

Eliza Kosoy (eko@mit.edu)
Psychology, UC Berkeley

Ilker Yildrim (ilkery@mit.edu)
Brain and Cognitive Sciences, MIT

Amir A. Soltani (arsalans@mit.edu)
Brain and Cognitive Sciences, MIT

Max Siegel (maxs@mit.edu)
Brain and Cognitive Sciences, MIT

Joshua B. Tenenbaum (jbt@mit.edu)
Brain and Cognitive Sciences, MIT

Elizabeth S. Spelke (spelke@wjh.harvard.edu)
Psychology, Harvard University

Abstract

Humans have an intuitive understanding of physics. They can predict how a physical scene will unfold, and reason about how it came to be. Adults may rely on such a physical representation for visual reasoning and recognition, going beyond visual features and capturing objects in terms of their physical properties. Recently, the use of draped objects in recognition was used to examine adult object representations in the absence of many common visual features. In this paper we examine young children's reasoning about draped objects in order to examine the develop of physical object representation. In addition, we argue that a better understanding of the development of the concept of cloth as a physical entity is worthwhile in and of itself, as it may form a basic ontological category in intuitive physical reasoning akin to liquids and solids. We use two experiments to investigate young children's (ages 3–5) reasoning about cloth-covered objects, and find that they perform significantly above chance (though far from perfectly) indicating a representation of physical objects that can interact dynamically with the world. Children's success and failure pattern is similar across the two experiments, and we compare it to adult behavior. We find a small effect, which suggests the specific features that make reasoning about certain objects more difficult may carry into adulthood.

Keywords: intuitive physics, cloth, cognitive development, object recognition, analysis-by-synthesis

Introduction

Imagine draping an elephant. What shape do you see? Probably not an exact silhouette, but a rough outline with a coarse bottom (Figure 1). This mental image also likely changes as you imagine draping an elephant placed on its side, or turned upside down. This simple feat of the imagination is quite remarkable. Imagining an elephant on its own may involve reactivating a learned representation or a visual memory of an elephant, but imagining an elephant draped by a cloth means 'seeing' a new object (the reader with extensive experience of draped elephants is free to imagine some other animal here). How do we come to this new image? One possible account is that we run a mental simulation and examine the outcome under noisy dynamic laws. But such a simulation requires object representations that go beyond representing image patches. Under this account, objects are represented as three-dimensional bodies, and the mental simulation is able to imagine the transformation and variation of the object under different processes. By examining people's ability to reason about the outcome of draping or to perceive draped object, we examine people's ability to reason visually without most of the traditional visual features that are assumed to play a part in recognition (Yildrim et al., 2016).

Recently, Yildrim, Siegel, and Tenenbaum (Yildrim et al., 2016) have investigated adult reasoning about cloth-covered objects as part of a larger examination of people's object representation as physical entities with the properties necessary for physical interaction. These studies showed that adults can reliably reason about the identity of covered objects in a match-to-sample task, even when the distractor object is within the same category type as the target object. Adult responses were best captured by a model based on a physics and graphics engine, which formalize the proposal that adults base their recognition and reasoning in part on a physical model of objects and a causal dynamical model of their interactions with the world.

More broadly, the Mental Physics Engine proposal suggests that the representations underlying much commonsense physical and visual reasoning are similar to those of modern game engines, software that is useful for quickly rendering an approximate simulation of a physical environment (see e.g. Battaglia et al., 2013; Gerstenberg et al., 2012; Smith & Vul, 2013; Hamrick et al., 2016; Ullman et al., 2017). Such game engines have also been proposed as an essential part of machine intelligence for commonsense reasoning (see e.g. Wu et al., 2015; Lake et al., 2017; Chang et al., 2017). While such a physics engine proposal predicts adult recognition and perception better than neural-network models based on visual image features, it is possible that adults come to this sophisticated understanding of objects and physics over time. Much less is known about children's representation of objects as physical objects for recognition. Here, we propose to examine young children's reasoning about draped objects as a way of examining the development of understanding objects as physical bodies, and of the causal processes that determine the behavior of objects.

Beyond this, we suggest that examining the development of intuitions about cloth is of interest in and of itself. This is because *cloth* (in the sense of a mesh or sheet of connected point masses, which can capture entities such as blankets, towels, and clothes) may be a basic ontological category in intuitive physical reasoning, akin to *rigid body* or *fluid*. At a high level, game engines separate physical entities into several broad classes based on their expected behavior, and the computational resources necessary to simulate them. This

high-level division is limited to only a few classes, and one of the common classes in modern engines is *cloth*, required specialized modular simulation Gregory (2009), and suggesting this may form a basic mental category as well. So, while it may initially seem that there are a large number of intuitive physical categories that can be investigated, of which cloth forms only a small subset, the success of the game engine approach to mental reasoning motivates us to focus on the small number of broad categories that have proven useful for engineers.

While cloth exists as a separate category in modern game engines, it is not obvious that an understanding of cloth has its origin in childhood. On one hand, by their first year many children have extensive experience with clothes, blankets, towels, tissues, and so on. A general mental physics engine with the right computational primitives may use this experience to generate the cloth category. On the other hand, our core knowledge physical reasoning is shared with many other animals and is believed to have a long evolutionary past (Spelke & Kinzler, 2007). Cloth, unlike liquids and rigid bodies, is a relatively recent category, and early human ancestors would not have needed to reason about it on a daily basis. Thus the mental physics engine may lack the right primitives to quickly construct this category.

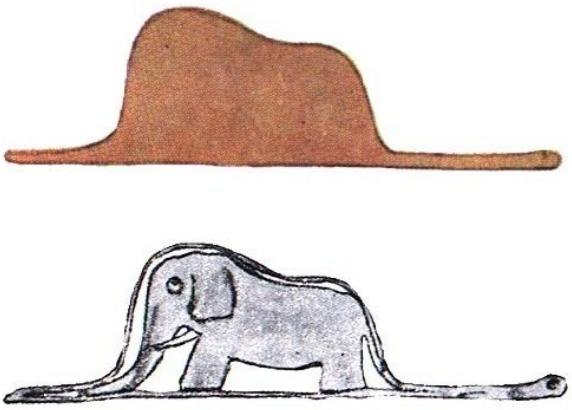

Figure 1: "My drawing was not a picture of a hat. It was a picture of a boa constrictor digesting an elephant. Then, I drew the inside of the boa constrictor, so that the grown-ups could see it clearly. They always need to have things explained." *The Little Prince*, by Antoine de Saint Exupéry.

In this paper, we probe young children's ability to reason about cloth using two basic tasks: reasoning from an uncovered object to a covered image (Experiment 1), and reasoning from a covered object to an uncovered image (Experiment 2). These tasks do not span the full space of the possible behavior of cloth, but they are meant to establish the existence (or lack) of basic competency. We consider an age range of 3–5 years, when children have for the most part not started a formal education, yet possess a sufficiently large vocabulary to understand the language used in the task. We find that children perform above chance in both tasks, and use an adult comparison to examine their patterns of success and failures. In the General Discussion, we consider the implication for generative vs. feature-based models, and the extension of cloth studies to infants.

Experiment 1: Uncovered → Covered

Participants

Sixteen individuals ($N = 16$, 5 female, median age 3.9 years, range 3.2-4.8) were recruited at the [City] Children's Museum. The size of the sample was pre-specified, based on a pilot study which indicated medium-to-large effect sizes can be expected.

Materials and methods

Participants were tested in a designated area in the [City] Children's Museum. Parents gave their informed consent, and advised not to encourage responses from their child.

Participants were presented with a touch-screen device (iPad), and told that they were going to play a game. Participants first played two warm-up rounds, in which they were shown a test-object on top of the screen (e.g. a bird), and asked to match it with one of two possible objects below (e.g. a bird and a horse). The warm-up round was meant to familiarize the participants with making a forced choice between two items based on a target item. By the second warm-up all participants correctly selected the matching object.

During test, participants saw 6 trials in succession, in random order (see Figure 2, top). Each trial contained a pair of objects, for example a mug and a bench. One object in the pair was randomly selected as the test object. The test object was shown at the top of the screen, uncovered. Below the test object were the pair of objects, covered in cloth. Participants were told to imagine that the test object had been covered by a blanket, and asked to indicate what the resulting image would be. Participant choices were automatically stored. Participants were given general encouragement, but no indication of whether their choice was correct.

All the stimuli pairs used in the experiments are shown in Figure 3. Uncovered and covered stimuli images were created in Blender (Blender, 2015). Covered objects were created by draping the uncovered objects using a physical cloth simulation. Objects were chosen from a collection of available objects previously used in experiments with adults (Yildrim et al., 2016). The size of the objects was scaled such that they took up approximately the same amount of visual space when covered.

Results and analysis

Participants' responses were summed across the object pairs, and are shown in Figure 4 (left). The summation resulted in a labeling score going from 0 (no objects correctly identified) to 6 (all objects correctly identified), with chance performance at 3. On average, participants correctly labeled 4.14 objects (95% CI 3.54–4.68, bootstrapped with 10,000 samples). The

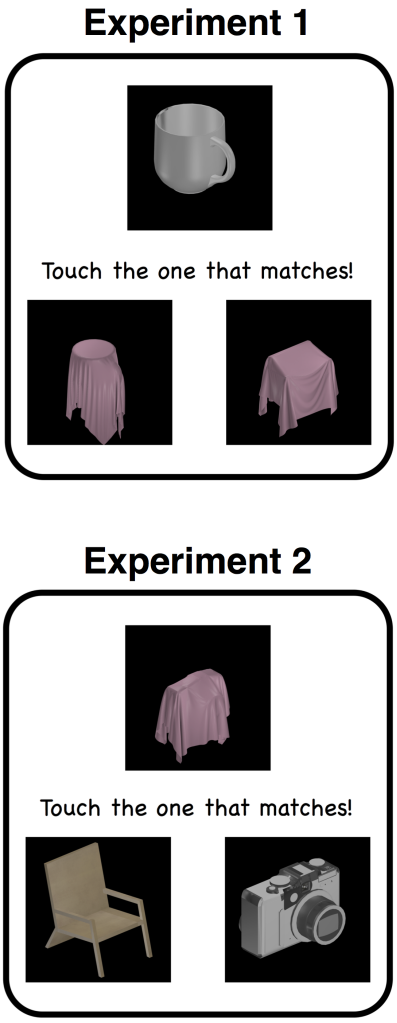

Figure 2: Schematic of example test trials in Experiments 1 and 2. At the top of a touchscreen is the target object (Uncovered in Experiment 1, covered in Experiment 2). Participants were asked to match the target object to one of the pair of objects at the bottom of the screen (Covered in Experiment 1, uncovered in Experiment 2).

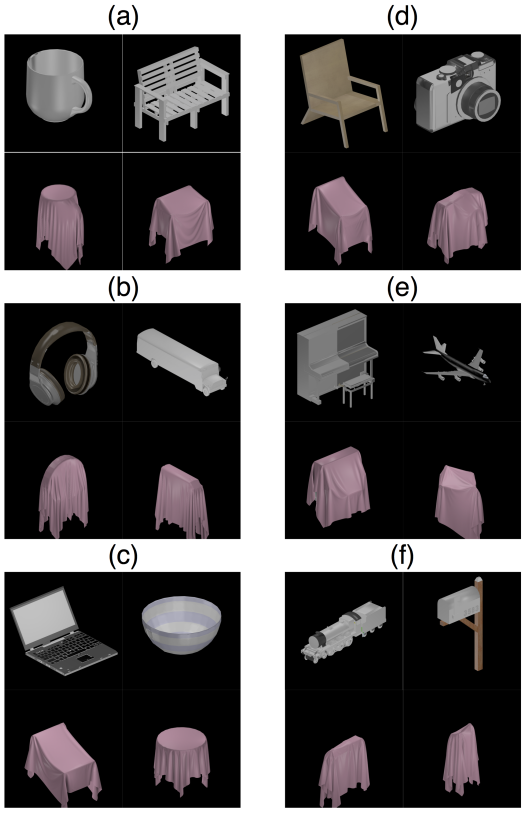

Figure 3: All stimuli pairs used in Experiments 1 and 2, uncovered and covered. In Experiment 1 participants saw one of the objects in the top row as the target, and matched it to the two items in the bottom row. In Experiment 2, participants saw one of the objects in the bottom row as the target, and matched it to one of the items in the top row.

confidence intervals are clearly above chance performance, and a standard two-sided T-test also indicates this result is statistically significant ($t(15) = 3.09, p < 0.01$).

We did not predict nor find a significant effect of age on participant performance. A logistic regression of labeling score on age was not significant, and neither was a median split comparison. Given the small sample size, however, we do not take this to strongly indicate the non-existence of an age effect, but simply the lack evidence for it.

Considering the stimuli by pair, we found that the identity of the objects in a given pair had an effect on participants' labeling. That is, some pairs were harder to discern than others. Specifically, using a standard two-sided binomial test at the $p < 0.05$ level, participants correctly distinguished mug/bench, headphones/bus, and laptop/bowl (Figure 3 a, b, c). Participants were unable to distinguish mailbox/train, piano/airplane, and chair/camera (Figure 3 d, e, f). The exact number of participants correctly labeling the objects by pair is shown in Figure 5.

Experiment 2: Covered → Uncovered

We took the results of Experiment 1 to indicate pre-school children have a general ability to match objects to their cloth-covered representations, though they may have been using one of several different strategies to do so. We next examine whether children were able to go in the inverse direction, inferring the identity of an object hidden under cloth.

Participants

Seventeen individuals ($N = 16$, 5 female, median age 4.0 years, range 3.0-5.0) were recruited at the [City] Children's Museum. The size of the sample was pre-specified at 16 to match Experiment 1.

Materials and methods

Participants were tested in a designated area in the [City] Children's Museum. Parents gave their informed consent, and

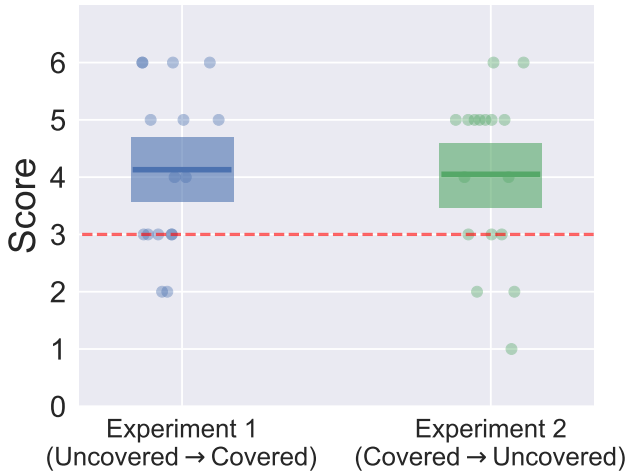

Figure 4: **Left**: Results of Experiment 1, seeing uncovered object and matching to cloth-covered image. **Right:** Results of Experiment 2, seeing cloth-covered image and matching to uncovered object. Score ranges from 0 (no trials correct) to 6 (all trials correct). Bold lines indicate mean score, and shaded colored area indicates 95% CI. Dashed red line indicates chance performance. Each dot indicates the response of one participant, jittered for visibility.

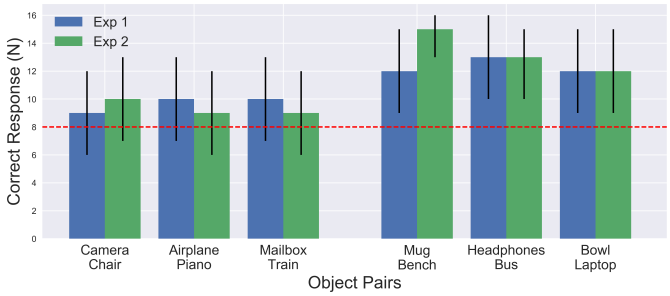

Figure 5: Results of Experiments 1 and 2 by object pair. The number of participants who correctly identified the target object is shown against specific object pairs. Black lines indicate 95% CI, dashed red line indicates chance performance. Children performed at chance or above chance levels for the same object pairs in both experiments.

advised not to encourage responses from their child.

Prior to the touch-screen part of the experiment, participants were shown 6 images of covered items in succession (printed on paper), and asked what they thought was under the cloth covering. That is, participants provided a free-form verbal response. The experimenter did not provide feedback on whether the response was correct or incorrect.

The touch-screen part of the experiment was similar to Experiment 1. Participants were shown an iPad, and told that they were going to play a game. As in Experiment 1, participants first engaged in two warm-up trials, and by the second trial all participants correctly labeled the matching object.

The test phase was also similar to Experiment 1: participants saw 6 trials in succession, in random order. Each trial contained a pair of objects, using the same pairs as in Experiment 1. However, in this experiment, the test object in a pair was covered by cloth, and the two objects below it were uncovered (see Figure 2). Participants were asked to indicate which of the two objects was under the cloth. Participant choices were automatically stored. As before, participants were given general encouragement, but were not told whether their choice was correct.

Results and analysis

The verbal response of participants to the first part of the task (freeform response when prompted to guess what is under a cloth) is summarized in Table 1. We did not predict that participants would correctly guess what was under a cloth, rather we used this task to examine the range of possible guesses. Note that many of the participants responded 'table' as this was a salient object mentioned by the experimenter.

Participants' responses to the forced-choice part of the task were summed across the object pairs, and are shown in Figure 4 (right). The summation resulted in a score going from 0 (no objects correctly identified) to 6 (all objects correctly identified), with chance performance at 3. On average, participants correctly labeled 4.26 objects (95% CI 3.66–4.74, bootstrapped with 10,000 samples). The confidence intervals are above chance performance, and a standard two-sided T-test also indicates this result is statistically significant ($t(15) = 3.87, p < 0.01$).

As in Experiment 1, a logistic regression of labeling score on age was not significant, and neither was a comparison which split participants by median age. We again stress that while we did not expect an age effect, we also do not believe these results necessarily indicate a 'true null' (the non-existence of an age effect), merely a lack evidence for it.

The identity of the objects in a given pair again had an effect on participant labeling. Interestingly, the exact same pattern emerged when using a standard two-sided binomial test at the $p < 0.05$ level. That is, in Experiment 2 participants correctly distinguished mug/bench, headphones/bus, and laptop/bowl, but did not distinguish mailbox/train, piano/airplane, and chair/camera. Figure 5 shows the performance of participants by pair.

We considered two hypotheses regarding the observation of the same pattern of successes and failures in both experiments:

- H1: Children's performance on both tasks is unrelated
- H2: Children's cloth-related reasoning is affected by object properties due to underlying object features.

We captured hypothesis H1 by assuming children's response is the result of informed inference (a biased coin with weight $\theta = 0.8$) or a random guess ($\theta = 0.5$), and that there are 3 weighted coins and 3 random coins per each experiment, but they are unrelated across experiments. The value of the weighted coin reflects an average of participant performance across the two experiments. We captured hypoth-

Covered object	Verbal description
Chair	table (5), box (3), chair (2), monster (1)
Camera	table (3), box (1), present (1)
Bench	table (3), square box (2), tall present (1)
Mug	table (3), chair (2), box (1), mountain (1), couch (1), squiggle strips (1), circle (1)
Laptop	table (3), square table (1), box(1), dot (1), rectangle (1), square (1), bridge (1)
Bowl	table (3), round table (1), circle (1), chair (1)
Mailbox	table (2), box (1), ghost (1), cat (1), blaster (1), ice-cube (1), gate (1), boat (1)
Train	box (1), chair (1), fence (1), stepstool (1),
Airplane	table (2), present (1), cowboy (1), vacuum cleaner (1), surfboard (1)
Piano	house (3), box (2), ladder (1), table (1), chair or table (1)
Headphones	rainbow machine (1), dog (1), ball (1), mountain (1), band-aid (1), diamond (1), front of crib (1), chair (1), jelly-fish (1), table (1)
Bus	box (1), square (1), fountain (1)

Table 1: Verbal responses of participants in Experiment 2. Numbers in brackets indicate the number of participants giving the preceding response. Numbers do not add up to the total number of participants as not all participants replied in all trials.

esis H2 by assuming the same set-up as hypothesis H1, but with the additional assumption that the weighted coins are matched with the same object pairs in both experiments. Assuming an uninformed uniform prior over both hypotheses, we can assess K, the Bayes factor of the two hypotheses, by estimating the ratio of the likelihood of the data under each hypothesis: $K = \frac{P(H2|D)}{P(H1|D)}$. The data under consideration is passing 3 binomial tests for each experiment, for the same object pairs. Using a bootstrap analysis in which 16 simulated participants have their behavior sampled from the coins described for H1 and H2, using 10,000 samples, we find a Bayes factor of $K = 21$, indicating strong evidence in favor of H2. Put briefly, the 'suspicious coincidence' that children are able to distinguish the same 3 pairs in both experiments is indicative of underlying features of the objects interacting with cloth-based reasoning.

Experiment 3: Adult comparison

While pre-school children were able to overall correctly reason about cloth-covered objects, they also made characteristic mistakes, indicating an underlying difficulty in reasoning about how particular objects will interact with cloth. Such difficulties may be due to simple lower-level feature interaction (for example, covering the mailbox and train both result in elongated rectangular shapes), or due to the end-result of a coarse draping simulation resulting in similar images, or a different reason altogether. Whatever the source of the difficulty, we wanted to examine whether it carried into adulthood. In the next experiment we examined the targeted prediction that adults would overall do worse on the trials that children failed.

Participants

One-hundred and twenty (N=120) participants were recruited online via Amazon Mechanical Turk. Eleven participants were discarded after failing to answer a catch question, and the remaining participants (N=109) are considered in the analysis below ($Median_{age}$ = 33 years, age range 20–70, 48 self-identified as female). We anticipated the task would be easy for adults, and based the number of participants on the expectation of small effect sizes.

Materials and methods

Participants were shown 6 trials, similar to Experiment 1. For each trial, participants were shown a target object and asked to imagine it covered with cloth. On a following page, participants were asked to select which of two covered objects matched the target object. The object pairs were the same as those used in Experiment 1. The order of presentation, the right/left location of the covered objects, and the identity of the target object were all randomized. Participants were asked to respond as quickly as possible. At the end of the 6 trials participants were asked to describe their task was in the study, and irrelevant answers (e.g. 'opinion', 'work', '0') led to discarding their data prior to analysis. Participants were also asked to provide information regarding their age and gender.

Results and analysis

Participants responded within about a second of presentation, with a median response of 1.1 seconds (95% CI 1.04–1.16) per trial. Participants also found the task relatively simple, with an average success rate of 98% (95% CI 96%–99%) per trial.

We considered the average correct response rate for the objects children found easier ('Children pass') and harder ('Children fail'). The average correct response rate by adults for the 'Children pass' trials was 99% (95% CI 98%–100%), whereas the correct response rate for the 'Children fail' trails was 96% (95% CI 94%–98%). The bootstrapped distribution over these variables and the response rate per object pair is shown in Figure 6.

The average correct response rate of adults per trial appears higher for the pairs that children found easier in Experiments 1 and 2, but this effect is very small as adults are nearly at ceiling.

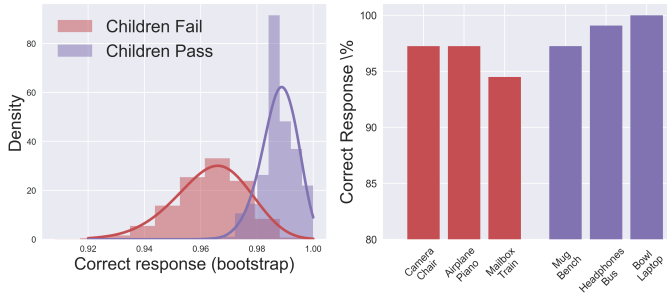

Figure 6: Results of Experiment 3. **Left**: bootstrap posterior distribution over the correct response rate aggregated by trial type ('Children pass' and 'Children fail'), 10,000 samples. **Right**: Correct response rate for each object pair, sorted by trial type.

General Discussion

People can reason intuitively about how things drape, wrap, envelop, sag, and droop. Recent experiments (Yildrim et al., 2016) have shown that adults perform well in a task that requires matching a covered and uncovered object, and that this ability can be captured by a physics and graphics engine which approximately simulates the draping of an object. Motivated by this work, as well as by the general category of 'cloth' in current game engines, we examined whether pre-schoolers can also reason about the interaction of cloth and rigid objects, and found their performance to be above chance in two such tasks. Children's pattern of failure and success was similar across the tasks, and a comparative task with adults found a small effect, suggesting that they too find the same object pairs hard or easy.

The current studies warrant tentative conclusions regarding object representation and the use of dynamic mental simulation in children. Previous studies with adults (Yildrim et al., 2016) rotated the objects, in a way that prevented simple feature-matching and meant in part to examine whether the adults were relying on a generative model reconstructions of the object. We did not use such a rotation in our studies, and we see them as a first step to examine whether children have *any* competence with cloth-based reasoning. It is possible that children's abilities rely on relatively simple feature matching, while adult reasoning is based more on reconstructing a mental representation of the 3D object shape It is also unclear which of several proposals for a generative model of 3D objects (whether for children or adults) is the right one (and see for example Soltani et al. 2017, for a comparison of several such methods for recovered objects from silhouettes). Further studies will need to use object rotations and a wider array of object pairs to examine this question.

The dynamics of cloth go beyond draping objects. For example, cloth sags when objects are placed on top of it, to a degree dependent on internal parameters related to its stiff and stretch. Can children reason about the likely sag of a piece of fabric, based on seeing its motion and knowing an object's felt weight? Are children sensitive to the weight of cloth, or will they reason about it as a weightless 2 dimensional manifold that only interacts geometrically with objects?

Even if both adults and young children rely on similar representations for reasoning about cloth, it is possible that these representations develop late compared to the basic expectations that infants have about rigid bodies (which innate or extremely early developing) and about liquids (which develop over the first year of life). Looking time experiments with infants could test this possibility by familiarizing infants to either cloth or a rigid body of similar proportions and texture, followed by an interaction in which the cloth and rigid body collide with or drape rigid objects.

To wrap up, while many issues remain hanging, this work begins to uncover the origin of cloth-based reasoning, which may form a separate ontological category within intuitive physical reasoning. It opens the door to future research probing the richness and origins of children's reasoning about a human invention that is ubiquitous in human cultures, and that occupies an interesting middle ground between rigid objects and amorphous stuff.

Acknowledgments

We wish to thank the parents and children who participated in the research carried out in [location]. This material is based upon work supported by [center], funded by [funding].

References

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*, 18327–32.

Blender. (2015). Blender - a 3d modelling and rendering package [Computer software manual]. Blender Institute, Amsterdam. Retrieved from `http://www.blender.org`

Chang, M. B., Ullman, T., Torralba, A., & Tenenbaum, J. B. (2017). A compositional object-based approach to learn-

ing physical dynamics. In *Proceedings of the 5th annual international conference on learning representations.*

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society.*

Gregory, J. (2009). *Game engine architecture*. CRC Press.

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition*, *157*, 61–76.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Smith, K. A., & Vul, E. (2013). Sources of Uncertainty in Intuitive Physics. *Topics in Cognitive Science*, *5*, 185–199.

Soltani, A. A., Huang, H., Wu, J., Kulkarni, T. D., & Tenenbaum, J. B. (2017). Synthesizing 3d shapes via modeling multi-view depth maps and silhouettes with deep generative networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1511–1519).

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.

Ullman, T. D., Spelke, E. S., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in cognitive sciences*, *21*(9), 649–665.

Wu, J., Yildirim, I., Lim, J. J., Freeman, B., & Tenenbaum, J. (2015). Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning. In *Advances in neural information processing systems* (pp. 127–135).

Yildrim, I., Siegel, M. H., & Tenenbaum, J. B. (2016). Perceiving Fully Occluded Objects via Physical Simulation. In *Proceedings of the 38th annual conference of the cognitive science society.*