Thinking inside the box:
Motion prediction in contained spaces uses simulation

Kevin A Smith (k2smith@mit.edu),[1] **Filipe de A B Peres** (filipe.peres@columbia.edu),[2]
Edward Vul (evul@ucsd.edu),[3] **Joshua B Tenenbaum** (jbt@mit.edu)[1]
[1] Department of Brain and Cognitive Sciences, MIT, Cambridge, MA
[2] Columbia University, New York, NY
[3] Department of Psychology, University of California, San Diego, La Jolla, CA

Abstract

Theories of the mental processes people use to perform physical reasoning often differ on whether they are based on simulation or on logical reasoning. Here we test how these different processes might combine in a motion-prediction task that can be solved either by simulation or by reasoning about the topology of the scene. Participants were asked to predict which of two goals a computerized ball would reach first, but in some of these scenes the ball was 'contained' in the same space as one goal but was topologically separated from the other. Even in these contained scenes, participants responded faster when they received motion information that would speed up simulation but not affect topological parsing. This suggests that simulation contributes to predicting short-range motion, even when alternate strategies are available.

Keywords: intuitive physics; simulation; topology; containers

Introduction

A long-standing debate in physical reasoning is whether people use simulation (Shepard & Metzler, 1971; Battaglia, Hamrick, & Tenenbaum, 2013) or symbolic reasoning (Forbus, 1983; Davis, Marcus, & Chen, 2013) to understand the world, and whether the mental representations that support physical reasoning are continuous or based on a qualitative analysis of the scene. Recently, models based on continuous simulation have had success explaining a range of human behaviors such as stability judgments (Battaglia et al., 2013), motion prediction (Smith & Vul, 2013), and causality judgments (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012). However, others have noted that simulation is an overly cumbersome process for many instances of physical reasoning where simple logical rules could suffice – e.g., if a ball is placed in a box and shaken, it seems easier to notice the topological relationship of containment and use the rule "An object in a closed container remains in the container" than to simulate the exact trajectory of the ball (Davis & Marcus, 2015).

In this paper we study how people make physical predictions in cases where either simulation or a logical parsing of the scene can provide an answer. As in the example of Davis and Marcus (2015), we focus on the topological relationship of containment: an object is contained in a portion of space if there exist other objects that surround it and prevent it from leaving that space. We choose to study containment for three reasons. First, Davis et al. (2013) demonstrate that topological containment can be parsed using simple rules of first-order logic for a rapid understanding of the scene. Second, people automatically and unconsciously process certain types

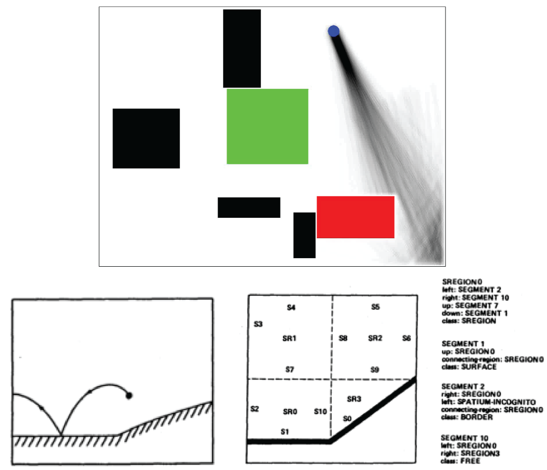

Figure 1: Diagram of continuous simulation (*top*) versus reasoning about kinematics using a qualitative scene parsing (*bottom*). Noisy simulation traces potential trajectories of objects through a continuous representation of the world. Qualitative physical reasoning segments the scene by topological regions and defines motion trajectories through a graph based on connections between those regions (bottom figure from Forbus, 1983).

of containment (Strickland & Scholl, 2015). And finally, in previous work we found that we could explain motion prediction using a model of noisy physical simulation across a wide variety of scenes, but that people made predictions more rapidly than would be expected under the simulation model in the handful of scenes where an object was topologically contained to make one outcome impossible (Smith, Dechter, Tenenbaum, & Vul, 2013). Together, this suggested that predicting motion in contained spaces would be a good candidate task for finding traces of both simulation and logical reasoning about scene topology.

Here we test whether and when topological reasoning about a scene occurs before simulation, versus when simulation supports prediction. We use a similar paradigm to Smith et al. (2013), in which participants observe a ball bouncing around a computerized screen and predict which of two 'goals' the ball will reach first. To make predictions using simulation, people would need to form a representation of the scene then step the motion of the ball forwards in time until it reaches a goal, but do not need to recognize the combined spatial relationship of objects: e.g., that a set of walls delineates one part of the space from another. Thus the time it takes to form simulations and make a prediction should be proportional to the path length the ball travels (Moulton &

Kosslyn, 2009), and so the amount of time to produce a response using simulation should be affected by the motion of the ball: if the ball is moving towards a goal, the relatively shorter path should require less time to simulate, whereas motion away from the goal would produce longer simulation paths and proportionally longer response times. Conversely, to predict that the ball will reach one of the two goals because it *cannot* reach the other requires representing the scene, logically reasoning about whether the walls form distinct regions, then deciding whether the ball rests within a region with one but not both of the goals. However, this reasoning process does not require moving the ball forward in time, and so should be insensitive to any differences in motion information.

In this experiment, we therefore asked participants to make a single prediction and measured their response times. We further varied the motion information provided: the ball could move towards the goal, away from the goal, or have no observed motion. A facilitation effect in which motion towards the goal produces faster responses is evidence that people are at least in part relying on simulation, while absence of this effect points towards topological processing. Finally, we vary the way in which the ball can be contained in order to test for the limits of topological processing.

Experiment

Procedure

We recruited 100 participants from Amazon Mechanical Turk using psiTurk (Gureckis et al., 2016) to take part in this experiment. The experiment lasted ~10-12 minutes, for which participants were compensated $1.20.

On each trial of the experiment, participants would observe a scene with a ball that could move in a straight line but bounce off of walls (such as the one in Figure 2) and were asked to predict whether the ball would reach the green goal or the red goal first. The colors of the goals were randomly switched to avoid any color biases, and responses were adjusted in switched-color trials for consistency of analysis. In two thirds of the trials, participants would observe the ball in motion for 500ms; in the remaining third of trials, participants would observe no motion but were instructed that the ball would move in a direction not known to them until after they made their prediction.

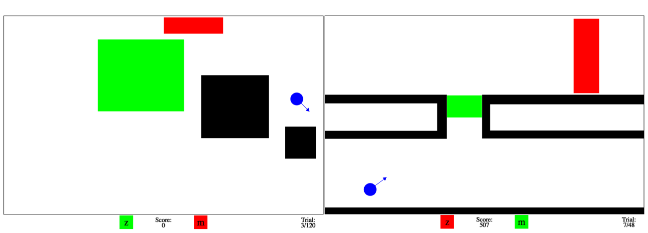

Figure 2: Diagram of experimental trials (*Left:* non-topological, *Right:* topological). Participants would either observe the ball in motion or a static ball, and would be asked to indicate whether they believed the ball would reach the green or red goal first. The arrow was not displayed but indicates the direction of motion.

Participants registered which goal they believed the ball would reach by pushing either the 'z' or 'm' buttons on the keyboard. The mapping between the key and goal color was randomized across participants to control for any directional effects. To ensure that participants observed the full motion path and to equate for processing time, the prediction could not be registered until either the 500ms of motion had stopped in the motion condition, or after the response buttons flashed after 500ms in the no motion condition.

After participants registered their response, the ball would travel along its trajectory until it reached a goal, and participants would be assigned a score between 0 and 100 points, with faster reaction times (up to 300ms) earning more points. If participants made an incorrect prediction, they always lost 10 points. If a participant did not respond for 2500ms, the trial would end and the participant would be awarded no points. These points were used to incentivize rapid responses and as motivation, but did not affect compensation.

On each trial we recorded both reaction time (between when a response could be indicated and when the button was pressed)[1] and the goal the participant predicted the ball would reach.

Stimuli

Participants observed 120 trials throughout the experiment. Of these trials, 96 were 'non-topological' trials that were randomly constructed such that the ball would reach a goal within 15 seconds, but were not hand designed with topological relationships. These uncontained trials were used to ensure that participants did not develop a deliberative, top-down strategy of judging topology.

The remaining 24 trials each participant saw were crafted to investigate one of four different dimensions of topological processing. There were six trial templates for each of the four dimensions (for a total of 24 templates), and each of these templates was adjusted to create three levels of containment across that dimension – each dimension started from the most contained, most simple, or smallest (level 1) and progressed to the most open, most complex, or largest (level 3). Thus there were 72 different topological trials used in the experiment, but each participant saw only one of the three levels formed from each template to avoid carry-over from similar trials. The dimensions of topological differences are described below.

Size This dimension was used to test whether topological parsing was performed by exploring the enclosed space at a constant rate, or whether topology is processed based on the configuration of the scene. If it is performed at a constant rate, then larger scenes with the same configuration should take longer to parse as topologically contained. We crafted these stimuli such that the smallest scene had dimensions that were 50% of the largest scene, while the middle scene had

[1] We also measured response time starting from when the stimuli came on the screen; results were qualitatively similar regardless of the choice of starting time.

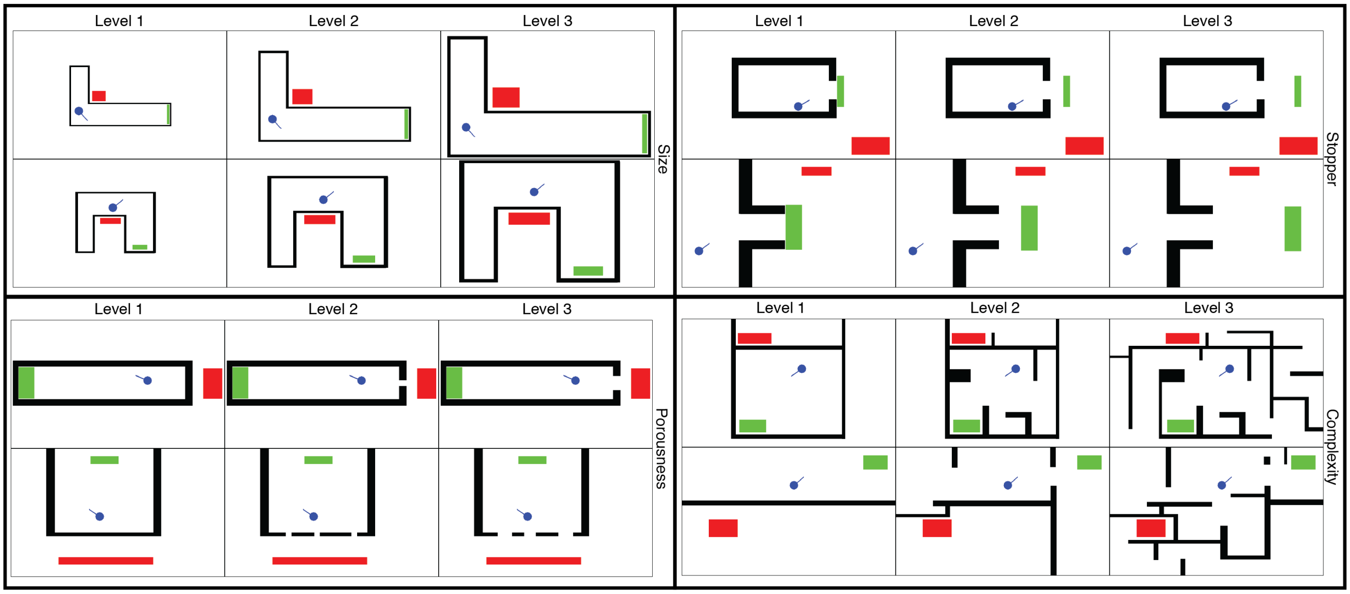

Figure 3: Sample trials from each of the topological dimensions along the three levels. The size trials (*upper left*) varied from small to large containers. The porousness trials (*lower left*) varied the size of gaps in the wall of the container. The stopper trials (*upper right*) varied the distance of the goal from making a seal with the rest of the container. The complexity trials (*lower right*) varied the internal and external structure of the containers. Balls in the "towards" condition moved in the direction indicated by the line, while balls in the "away" condition moved in the opposite direction. In all cases the ball ended at the green goal.

dimensions that were 75% as large (see Figure 3, upper left).

Porousness This dimension was used to test whether an object is topologically parsed as a container only if it is fully sealed, or whether containment is relative to the object inside. If topological processing is based solely on the container, then any gaps should prevent it from being seen as a container, whereas if it is calculated relative to the object inside, then it should still be observed as a container if gaps in its walls are smaller than the object it is holding. In the three levels along this dimension, one container was fully sealed with no breaks in its walls, one had gaps that were smaller than the ball, and the final level had gaps large enough for the ball to fit through (see Figure 3, lower left).

Stopper Forbus (1983) suggests that scene and motion descriptions also take into account what sorts of motions are allowable within the scene. If this is the case, then topological processing might also be affected by whether the path an object takes to exit a container is implausible or impossible. We therefore tested whether participants would consider the ball to be contained if there were almost no conceivable physical paths that would allow it to escape, even if there exist simple paths outside that do not account for physical motion. Level 1 along this dimension was produced with one goal forming a seal with the rest of the container. Level 2 moved the goal away from the container so that the ball could fit through, but in a physically implausible way. Level 3 moved the goal even further so that it is easily possible that the ball could exit the container without hitting the goal under plausible kinematic motion (see Figure 3, upper right).

Complexity This dimension was used to test how topology and simulation interact – even in situations where the ball is fully contained will people use simulation if parsing the boundaries of the container is too difficult? Along this dimension, the levels included a simple configuration (e.g., the screen is split into two parts by a single wall), a moderately complex configuration with more internal and external structure, and very complex configuration (see Figure 3, lower right).

In all of the levels of all of the trials, participants would see either motion that is in the general direction of the goal within the container, motion in the opposite direction away from the goal, or no motion. To ensure each trial was novel, each participant only saw one type of motion for each trial, counterbalanced across participants. These motion conditions were tested because differences in velocity information should affect simulation but not topological judgments. If people are using simulation, we would expect that motion towards the goal should speed processing as compared to motion away, since each simulation will have a shorter distance to travel and fewer bounces before reaching the goal (Hamrick, Smith, Griffiths, & Vul, 2015). Similarly, if people predict the outcome of no motion trials by simulating paths the ball could take in any direction, because most potential paths would be longer than the paths created with motion towards the goal, predictions should also be slowed in this case. If containment is judged by parsing the topology without using information about velocities, then changing the type of motion information provided should not change the speed of this mental process. We can therefore test for the presence of simulation by the presence of faster reaction times in the "towards" motion condition compared to the "away" or "no motion" conditions.

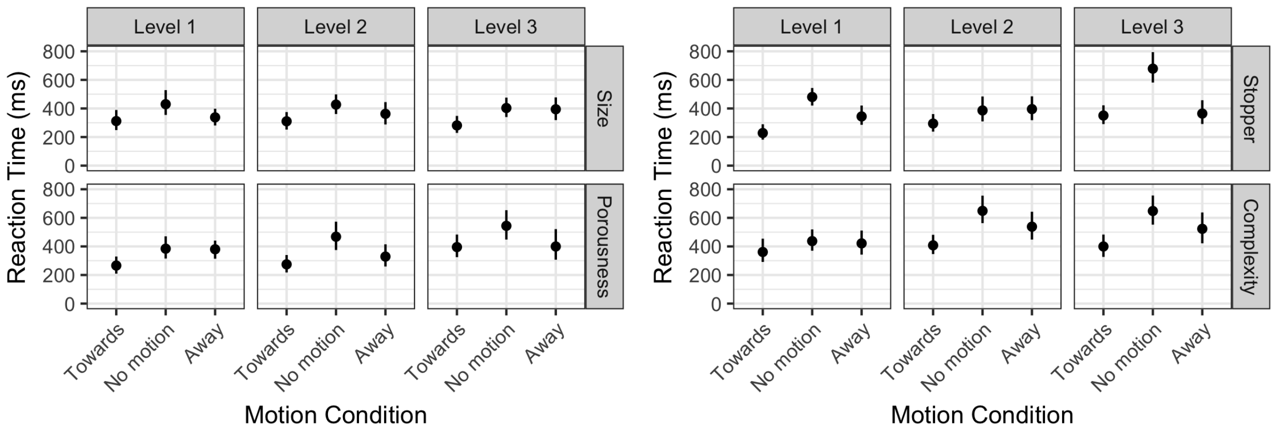

Figure 4: Geometric means of reaction times across topological dimensions and motion conditions. Bars indicate 95% confidence intervals bootstrapped from 500 samples. In all cases reaction times in the "towards" condition were faster than those in the other two conditions, indicating a use of simulation.

Results

To ensure that we do not use data from participants who were not paying attention, we eliminated responses from trials where participants minimized or otherwise hid their browser screen (0.3% of trials) or where participants did not indicate a prediction in the allotted time (1.6% of trials). Finally, because participants observed the scene for 500ms before making a response, some responses could be anticipatory. To prevent these measurements from skewing the data, we removed all responses under 10ms (0.7% of all trials). For the purpose of all analyses, reaction times were log-transformed to account for long tails (Whelan, 2008) but transformed back for reporting and display.

We first test for overall differences in speed of processing across all topological dimensions, levels, and motion directions.[2] This analysis suggests that motion direction plays a pivotal role in explaining reaction times ($F(2,2152) = 61.8, p \approx 0$), with "towards" (321ms, 95% CI: [288, 357]) being faster than "away" (394ms, 95% CI: [354, 439]), which in turn is faster than "no motion" (487ms, 95% CI: [437, 543]) over all trial types.

The dimension of topology also affects reaction time ($F(3,58) = 14.2, p = 7.8 * 10^{-8}$), with the complex trials (481ms, 95% CI: [424, 546]) being slower than the size trials (359ms, 95% CI: [316, 407]), the porous trials (373ms, 95% CI: [329, 423]), and the stopper trials (377ms, 95% CI: [333, 428]).

Finally, the level of topology had an overall effect ($F(2,58) = 5.9, p = 0.0046$). Although the specific way in which trials changed with differences in level was not the same across topological dimensions, they were all ordered such that the first level was expected to produce the fastest predictions and the third the slowest. Here the simplest / most contained trials were the fastest (359ms, 95% CI: [319, 404]), followed by the intermediate trials (393ms, 95% CI: [349, 443]), followed by the most complex / least contained (436ms, 95% CI: [387, 490]).

[2] We modeled log-RT using a linear mixed effects model with random effects for participant, trial, and a trial-by-motion direction interaction.

Nonetheless, there was no statistically reliable effect of any interaction (all Fs <1.5, all ps >0.13). This suggests that the amount of speed-up from observing motion does not change with the type of topological trial, which in turn suggests that simulation is used across all topological trials.

To more directly test for the use of simulation across dimensions of topology, we can compare how fast people respond in the "towards" condition as compared to the "away" and "no motion" conditions. We calculated a simulation facilitation index as the ratio of the reaction times in the "towards" condition versus the average of the other two conditions. If this index is less than one, then we have evidence that participants were using simulation to make predictions in that condition. As can be seen in Table 1, across every condition the simulation facilitation index is numerically less than one, and in most conditions (9 of 12) the 95% confidence intervals do not include one either.

These simulation facilitation effects are not driven by a small set of outlier trials. Across all topological trials, reaction times of participants in the "towards" condition were faster than those in the "away" condition in 52 of 72 trials (binomial test, $p = 0.0002$), and faster than those in the "no motion" condition in 61 of 72 trials (binomial test, $p = 1.6 * 10^{-9}$).

We also consider whether the facilitation in the "towards" condition is truly facilitation, or whether this effect is observed because "away" motion slows down processing: in some cases the ball was moving away from the correct goal and towards the incorrect goal, and simple directional motion towards a goal might speed reactions for that goal and slow reactions for the incongruent goal. However, participants were still faster in the "away" motion condition than the "no motion" condition both in average reaction time and in 49 of 72 trials (binomial test, $p = 0.003$), so the differences in reaction time cannot be explained simply as a slowdown due to motion towards the incorrect goal.

Table 1: Simulation facilitation index for each of the topological dimensions and levels. Numbers in brackets indicate 95% confidence intervals. In all cases, there is a simulation facilitation advantage, and in all but three conditions the confidence intervals are below one. This suggests that simulation is used across all topological conditions, including the conditions with simple containers.

	Level 1	Level 2	Level 3
Size	0.764 [0.593, 0.983]	0.82 [0.635, 1.06]	0.716 [0.556, 0.921]
Porousness	0.714 [0.554, 0.92]	0.716 [0.556, 0.922]	0.869 [0.676, 1.12]
Stopper	0.592 [0.46, 0.761]	0.704 [0.548, 0.904]	0.681 [0.529, 0.877]
Complexity	0.845 [0.655, 1.09]	0.7 [0.543, 0.903]	0.713 [0.554, 0.917]

Although motion towards the goal speeds reaction times across all trial types, we can test whether participants were simply cued to respond faster in general when the ball is moving towards the goal, or whether there is evidence that these judgments are based on simulation. If people are using simulation, then as the ball travels further to reach the goal we would expect mental simulations to also travel a longer path and thus take more time to produce (Moulton & Kosslyn, 2009). We therefore expect that reaction times should increase roughly in line with the time it takes the ball to actually reach the goal.[3] As can be seen in Figure 5, there is a relationship between the actual travel time of the ball and participants' reaction time on that trial across all of the topological trials ($r = 0.29, t(142) = 3.5, p = 0.00056$), but we do not have evidence that this relationship differs between the "towards" and "away" conditions ($F(1,141) = 0.76, p = 0.38$).[4] This relationship suggests that simulation was in general used to produce motion predictions in this task regardless of the direction of motion.

Finally, we considered two alternate explanations that might give rise to this pattern of data by chance. First, if participants were 'guessing' more in the towards motion trials, we might expect them to respond faster but be less accurate. Second, if participants changed the speed with which they responded over time, this could be a potential confound in our analyses. However, neither of these alternate explanations hold.

If participants are using a different speed-accuracy tradeoff across motion types, we might expect that the reduction in speed is counterbalanced by higher accuracy. Among the topological trials, participants were numerically most accurate in the 'no motion' condition (89.0%), followed by the 'towards' condition (86.7%), then the 'away' condition (81.3%). While there is an overall effect of motion direction on accuracy ($\chi^2(2) = 20, p = 4.5 * 10^{-5}$), this is driven by the 'away' condition being less accurate than the other two (vs. 'towards', $z = 2.6, p = 0.024$; vs. 'no motion', $z = 3.28, p = 0.003$) rather than by a difference between the 'towards' and 'no motion' conditions ($z = 0.74, p = 0.74$). Furthermore, a speed-accuracy trade-off cannot explain why people are both slower and less accurate in the 'away' condition than they are in the 'towards' motion condition.[5]

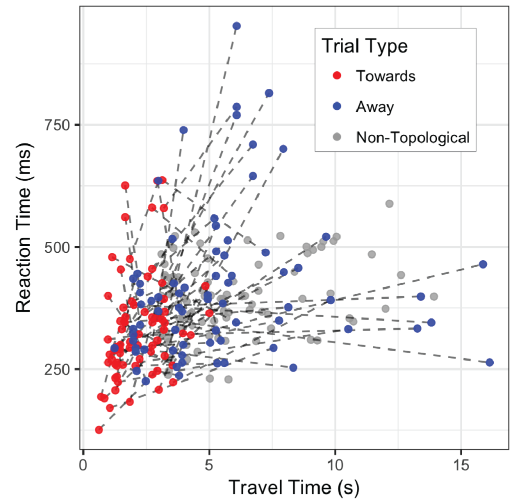

Figure 5: Comparison of the time it takes the ball to actually reach the end goal versus the geometric mean of participants' reaction time in that trial ($r = 0.29$). Topological trials are linked between 'towards' and 'away' motion with dashed lines.

We also tested for changes in the speed of response throughout the experiment. There was a minuscule effect of trial order on response speed that was not statistically reliable (each additional trial was 0.08% slower, 95% confidence interval = [-0.02%, 0.18%], $F(1,151) = 2.7, p = 0.10$). Therefore changes in response times over the experiment cannot explain the difference in reaction times across the different motion conditions.

Discussion

In this study we tested for situations where reasoning about topological containment preempts physical simulation across a wide variety of trials where both topological relationships and simulations could be used. We found that participants were using simulation across all types of topological trials, including the most simple cases of containment.

But why do we find evidence of simulation when a simple topological analysis alone would suffice? We consider five

[3]Because it was unclear how far the ball should travel in the "no motion" condition, we did not include those trials in this analysis.

[4]If we include the non-topological trials in this analysis, we still find a relationship between ball travel time and reaction time ($r = 0.29, t(238) = 4.6, p = 7.2 * 10^{-6}$) but do not find statistical evidence that the slopes differ between non-topological, topological towards, and topological away conditions ($F(2,236) = 1.01, p = 0.37$). Furthermore, if we remove topological trials that were not fully contained (Porous level 3 and Stopper levels 2 & 3), this relationship remains ($r = 0.34, t(106) = 3.7, p = 0.00037$) and we still do not find changes with motion condition ($F(1,105) = 0.78, p = 0.38$).

[5]This analysis considers all topological conditions, including the porousness and stopper trials in which simulation in theory could reach the incorrect goal. However, limiting this analysis to just the size and complexity trials (which were all fully contained) produces the same qualitative results.

possibilities.

First, the simulation facilitation effect could arise from a mixture of individuals, some of whom use simulation and others who use topological processing. Due to the small number of topological trials each participant saw in this experiment, we cannot precisely measure whether each participant individually had a simulation facilitation effect, only that this effect is found on average. Further research is required to investigate individual differences in the use of these two processes.

Second, because the majority of the trials could not be solved with topological reasoning, if people must choose between either simulation or topological reasoning, simulation would be the more general choice. Thus if there are cognitive costs for switching between different processing strategies, participants might constantly use simulation. Future work will study whether people continue to use simulation even when it is not as frequently required.

Third, Davis and Marcus (2015) suggest that "simulation is effective for physical reasoning when the task is prediction, when complete information is available, ... and when the range of spatial or temporal scale involved is moderate" – exactly the conditions of this experiment. Perhaps simulation is automatically activated in tasks that fit this description but not in others, and we happened to use a task that relied on simulation. This might also imply that the "no motion" trials involved a separate, logic-based process as opposed to the motion trials with complete information. Indeed there is a numerical pattern in these results that would support this interpretation: in Figure 4 the "away" reaction times are always slightly slower than the "towards" reaction times, but the difference between "towards" and "no motion" is more variable across conditions. Although there was not statistical evidence for such a difference, this pattern would be consistent with people using a separate process that requires a longer and more variable amount of time in cases where no motion was observed.

Fourth, people may be using simulation to gain information about containment. Liang, Zhao, Zhu, and Zhu (2015) explain human ratings of how well one object will be contained by another by simulating how often the first object will stay inside the second when dropped into it. This might suggest that for simple tasks our perception of containment is statistical (one would not expect this object to ever leave the container) rather than logical (the topology of the container entails the object inside will not leave).

Finally, making predictions may involve multiple processes running in parallel, including both simulation and topological parsing. In many of the topological trials participants observed – especially the "towards" trials – the ball did not have to travel far to reach the goal. If both simulation and topological reasoning are active at the same time, these might be the cases where simulation provides an answer quickly and wins out over topological processing. In Figure 5, the relationship between the time the ball actually takes to reach the goal and reaction time becomes more variable and flatter as the travel time takes longer. These longer trials might be cases where simulation fails to provide an answer before less continuous processes can, and so we do not see the same sort of relationship between path length and reaction time. Intriguingly, this relationship is reduced even for the non-topological trials that last this long, suggesting perhaps that simulation can only look a short time into the future, after which point we use more qualitative scene representations that could support either qualitative simulation or logical reasoning.

Although simulation appears to be active in simple tasks that require predicting the motion of objects, fully explaining human physical reasoning will require a better understanding of how simulation interacts and trades off with more qualitative methods of conceptualizing the world.

Acknowledgments

We thank the three anonymous reviewers for their comments on this manuscript. This work was supported by the Center for Brains, Minds & Machines, funded by NSF STC award CCF-1231216.

References

Battaglia, P., Hamrick, J., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Davis, E., & Marcus, G. (2015). The scope and limits of simulation in cognitive models. *arXiv preprint arXiv:1506.04956*.

Davis, E., Marcus, G., & Chen, A. (2013). Reasoning from radically incomplete information: The case of containers. In *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* (p. 273-288).

Forbus, K. D. (1983). Qualitative reasoning about space and motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (p. 53-73). New Jersey: LEA Associates, Inc.

Gerstenberg, T., Goodman, N., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Societyroceedings of the 34th.*

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., ... Chan, P. (2016). psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, *48*(3), 829–842.

Hamrick, J., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? Optimal mental simulation tracks problem difficulty. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Liang, W., Zhao, Y., Zhu, Y., & Zhu, S.-C. (2015). Evaluating human cognition of containing relations with physical simulation. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273–1280.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.

Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (2013). Physical predictions over time. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *35th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Smith, K. A., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*, *5*(1), 185–199.

Strickland, B., & Scholl, B. J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General*, *144*(3), 570-580.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475-482.