UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
When do people use containment heuristics for physical predictions?

Permalink
https://escholarship.org/uc/item/3426g875

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Brockbank, Erik
Vul, Edward
Smith, Kevin

Publication Date
2019

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

When do people use containment heuristics for physical predictions?

Erik Brockbank [1]
(ebrockbank@ucsd.edu)

Edward Vul [1]
(evul@ucsd.edu)

Kevin Smith [2]
(k2smith@mit.edu)

[1] Department of Psychology, 9500 Gilman Dr. La Jolla, CA
92093-109 USA
[2] Department of Brain and Cognitive Sciences, MIT Building 46-4053,
77 Massachusetts Avenue Cambridge, MA 02139

Abstract

Accounts of human physical reasoning based on simulation from a noisy physics engine have enjoyed considerable success in recent years. However, simulating complex physical dynamics can be a computationally expensive process, and it is possible that people use faster, cheaper shortcuts to make predictions and inferences in complicated physical scenarios. Here we asked people to predict the eventual destination of a ball on a 2D bumper table (in the style of Smith, de Peres, Vul, and Tenenbaum (2017)). We designed scenarios that we expected would modulate the use of heuristics and simulation: the bumper table provided varying degrees of containment to constrain future outcomes and to make a containment heuristic more useful, and could have more or less internal structure to vary the reliability of noisy simulation. As the containment heuristic becomes more useful, and as simulation becomes more expensive, we expected that people would switch from using simulation to rely more on rapid heuristic-based predictions and therefore respond faster. Instead, we found that even when containment was very predictive, people were progressively slower and less accurate as simulation complexity increased, indicating that they persisted in using simulation rather than containment heuristics.

Keywords: simulation; heuristics; physics

Introduction

In everyday life we are constantly tasked with making predictions about how physical objects will behave and interact, whether changing lanes in traffic or stacking dishes in the sink. Such inferences are so commonplace that we rarely think twice about them. However, the mechanisms by which we are able to make these inferences are far from obvious: at a minimum, they require a rich understanding of how things in the world tend to move and the ability to make rapid predictions based on this knowledge, both non-trivial achievements from a computational perspective.

Prior research has shown that a range of human physical inferences can be captured by *Intuitive Physics Engine* models that rely on simulations of physical outcomes performed with a probabilistic physics engine similar to those used in computer games (Battaglia, Hamrick, & Tenenbaum, 2013). By sampling from these simulations, probabilistic models can generate a reasonable representation of the physical world and make predictions accordingly (Ullman, Spelke, Battaglia, & Tenenbaum, 2017). Such models have been successful in reproducing human judgments across a range of tasks and domains, from predictions about object balance (Battaglia et al., 2013), mass (Hamrick, Battaglia, Griffiths, & Tenenbaum, 2016), and velocity (Smith & Vul, 2013) to liquid dynamics (Bates, Yildirim, Tenenbaum, & Battaglia, 2018) and causal attribution (Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017).

While simulation allows us to reproduce many features of human physical reasoning, there are also situations where people's behavior is inconsistent with the use of an intuitive physics engine (Smith, Battaglia, & Vul, 2018). Empirically, human behavior sometimes differs significantly from predictions made by simulation-based models, suggesting that we have sophisticated strategies for avoiding simulations when other forms of inference will suffice (Smith, Dechter, Tenenbaum, & Vul, 2013). In particular, research on errors in physical judgment have shown that people often hold a number of systematic biases which are inconsistent with even basic physical simulations (see Davis & Marcus, 2015 for an overview of some of these). Underlying this difference is a criticism of simulation as a computational account of all human physical reasoning: simulation of almost any sort, but particularly of complex physical phenomena, may require considering the interactions between a large number of objects over time. Because interactions between objects add uncertainty to predictions (Smith et al., 2013), in complex scenarios these simulations might therefore require keeping a large number of objects in mind and yet still produce very uncertain predictions. These sorts of considerations have led some to argue for a limited role of simulation in human physical reasoning (Davis & Marcus, 2016).

In light of the challenges posed to a simulation-based account of human physical reasoning, what alternatives can account for people's ability to make diverse predictions about physical interactions in the world around them? A large body of research supports the idea that humans are adept in their use of heuristics and other simplified qualitative prediction strategies, including in the domain of physical predictions (Gigerenzer & Todd, 1999). Prior work has shown that people can represent certain topological relationships like containment using only first-order logic (Davis, Marcus, & Frazier-Logue, 2017). Given the large number of strategies available to reasoners and the flexibility with which we navigate the physical world, it has been proposed that humans selectively utilize a toolbox of prediction techniques, including simulation, qualitative reasoning, and logical inference, as well as analogical and rule-based strategies (Davis & Marcus, 2015).

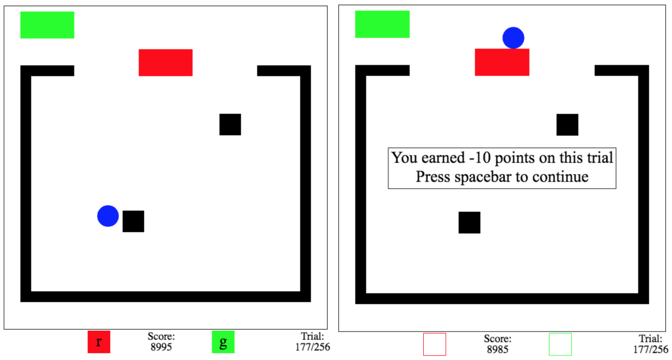


Figure 1: A simple trial with partial containment and two obstacles. At left is what participants see when prompted to guess a target after 2.5s of animated ball movement. At right is feedback after guessing "green" and seeing the ball animated on the remainder of its path.

The idea that humans are able to balance simulation-based prediction with alternative prediction strategies is intuitively appealing because it offers a way to unify simulation-based accounts with complementary accounts of physical inference based on e.g. topological and visual features. However, it raises a number of additional questions. If humans are able to flexibly recruit different strategies for making physical predictions, what determines the choice of one strategy over the other? How and when do we switch between fine-grained simulation methods and more coarse qualitative analyses? The exact mechanisms for such decisions remain poorly understood. For example, when novel but reliable and visually salient heuristics are available, people often fail to use them unless the existence of such heuristics are made explicit (Callaway, Hamrick, & Griffiths, 2017). A simple hypothesis is that compared to simulations, topological predictions are faster, lower fidelity, and less generally applicable; consequently, topology ought to be used when the scenario makes topological cues particularly useful, and renders simulations particularly imprecise and costly by complex scenarios. In other words, if computationally expensive simulations are unlikely or unable to produce a confident prediction, while topology can, a rational agent should make a guess based on simpler heuristics or visual features rather than waste resources on repeated simulations.

In the present study, we tested the hypothesis that people balance the precision and cost of simulation against the applicability of topological analysis when making physical predictions. Our experiment builds on prior research in several important ways. First, we examine people's reasoning about containment scenarios because prior research has shown that containment relationships can be expressed propositionally and that intuitive inferences about containment can be made with such knowledge-based reasoning even with very little information (Davis et al., 2017). As such, it is an ideal simplified model for physical inference. Second, containment relationships can in some cases be visually processed rapidly and automatically (Strickland & Scholl, 2015). Finally, prior research has used a similar paradigm to explore the degree to which people simulate or use topological inference when making physical predictions in scenarios involving containment relationships. Smith et al. (2013) modeled inference on a prediction task using noisy simulation but found that people's predictions were more rapid than the model predicted in scenarios involving containment. Building on these results, Smith et al. (2017) presented participants with similar tasks in which a containment heuristic was available but found evidence for simulation across all the tasks. However, in the tasks presented to participants, the simulation required was fairly straightforward and temporally limited. Therefore, insofar as simulation and topological processing happened in parallel or participants reasoned that simulation was a consistently viable strategy, they may have failed to leverage a more coarse containment-based judgment out of habit or convenience (Smith et al., 2017). We hypothesize that when topological predictions are available *and* simulation proves intractable or uncertain, participants will be more likely to make their predictions based on topology. In line with this hypothesis, Davis & Marcus (2015) argue that simulation is most effective on relatively short time scales and small spatial scales such that simulation is straightforward and reliable. Here we violate this condition by including trials in which the number of obstacles (complexity level) makes simulation both more uncertain and potentially longer. We expect that participants, faced with predictions involving unreliable simulations, will pursue alternative strategies for prediction: an agent that rationally trades off the advantages of simulated inference with the computational costs should select more favorable knowledge-based inference strategies when conditions support them.

Experiment

In the present study, we tested the hypothesis that people would switch from using slower simulation to faster heuristics when simulation becomes less efficient. Specifically, we presented participants with a task which required them to make predictions about the path of a ball in a series of two-dimensional environments. We manipulated (a) how much the topography of the environment allowed a simple topological "containment" heuristic to identify the answer (degree of containment), (b) the complexity and uncertainty of simulations in the environment (degree of complexity). The core prediction is that participants would favor using simulation to obtain an answer when simulations were easy and topology was uninformative, but would switch to relying on containment, or other coarse topological cues when they were particularly effective, and simulation was particularly ineffective. Specifically, we rely on the assumption that using a fast containment heuristic would be more efficient than simulation, thus we predict that for high-containment scenarios, increasing complexity would decrease response times.

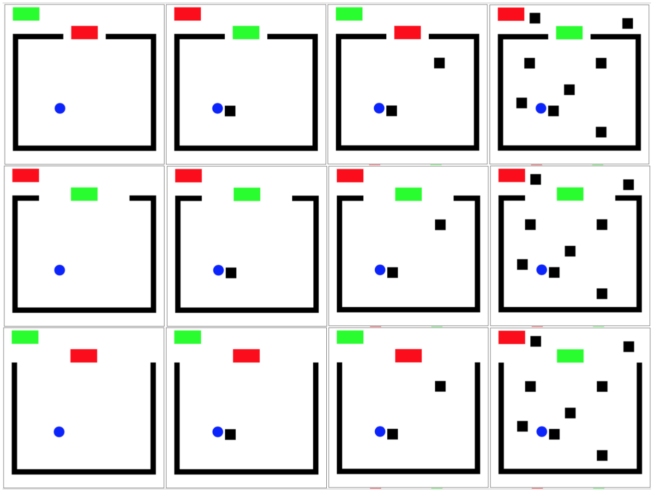


Figure 2: Twelve trials for Scenario 1, increasing in *simulation complexity* in the horizontal direction and *containment* in the vertical direction (highest containment at top). The high containment, high complexity trials offer a simple topological prediction without needing to simulate the ball's interaction with the walls and obstacles.

Participants

Participants were 81 undergraduates from the University of California, San Diego who received course credit for participation. The experiment lasted approximately 25 minutes.

Methods

We used a task that is very similar to Smith et al. (2017). During the experiment, participants were shown a series of trials depicting a blue ball on a flat surface (600 pixels by 600 pixels). The ball was surrounded by walls and square obstacles that the ball could bounce off. Each trial contained a red and a green target and the goal of the task was to determine whether the ball would hit the red target or the green target first (see Figure 1). Before making a guess, participants were shown 2.5s of the ball's movement, after which the ball paused in its trajectory and participants pressed either the "R" or "G" key to indicate their guess for the *red* or *green* target. After participants made their guess (or 10s elapsed), the ball would resume its movement until it hit one of the targets. At the end of each trial, participants received points based on their accuracy and their response time: -10 for an incorrect guess, 0 for no guess, variable points for a correct guess based on time to respond (see Figure 1). The points for a correct guess were allotted based on an exponential decay function of response time so participants were rewarded for guessing quickly if they could generate an accurate guess rapidly, but the penalty for longer response times quickly diminished. To illustrate, participants received 100 points for responses at 250ms, 71 points at 1000ms, and 45 points at 2000ms.

Participants read a brief set of instructions and completed three practice trials before doing the experimental trials. Each participant completed all trials in the experiment: 48 trials representing each complexity and containment level across four scenarios, with 64 "distractor" trials (discussed below) for a total of 256 trials. The order of the trials was randomized for each participant, as was the selection of the red and green target for reach trial.

Stimuli

The trials were grouped into four qualitative scenarios, and within each scenario they were parametrically manipulated along two dimensions that modified the uncertainty of simulations and the availability of topological predictions.

Scenario: Each trial belonged to one of four possible scenarios corresponding to the containment structure that the targets were placed in. For example, one scenario placed the ball inside variants of a box where one of the targets was placed in the opening, while another had the ball traveling down a right-angled tunnel with a target at one end. (see Figure 5).

Containment: Each scenario had three distinct *containment* levels that varied how much the ball and one of the targets were contained by the set of walls in the scenario. In the high containment trials for each scenario, the ball was virtually guaranteed to hit one of the targets because the ball and that particular target were almost entirely contained by the walls. In the low containment trials, the walls provided only minimal containment for the ball and one of the targets, rendering topology and containment fairly uninformative.

Simulation Complexity: For each scenario and containment level, there were four *complexity* levels which varied the degree of uncertainty involved in simulating the ball's path. This was accomplished by placing an increasing number of square obstacles throughout the scene: simulation therefore required accommodating the growing possibility of the ball bouncing off one or more obstacles before hitting one of the targets, making simulation results less certain. The lowest complexity levels for each scenario and containment level had no such obstacles, while the highest complexity levels had eight obstacles spread throughout the scene (see Figure 2).

Each unique scenario, containment, complexity combination was rotated 90, 180, and 270 degrees to allow for more trials and to prevent the scenarios from being too predictable. In addition, there were 64 *distractor* trials that were identical to the high containment trials in each scenario, except that both targets were placed inside or outside the containing structure. These were added to prevent participants from adopting a strategy of assuming that every trial would have a containment structure or other topological best guess once they had seen a number of trials in which that was the case.

For each trial, we captured participants' accuracy (correct or incorrect) and response time. Previous results using the same target task have provided evidence that participants are likely to make simulated inferences for this task across a range of scenarios and further that response time is correlated with time required to simulate the outcome (Hamrick, Smith, Griffiths, & Vul, 2015; Smith et al., 2017). We expected response time to be a reasonable measure of participants' reliance on simulation for the inference in the task: as the complexity of the simulation required to make a prediction increased, so too should the response time. In contrast, predictions made via topological inference should show little change in response time as complexity of the scene increased. When one of the targets was clearly contained in the same space as the ball, the uncertainty or duration of the ball's simulated path should not have had any bearing on judging which target the ball would hit first if participants were taking advantage of this containment information. Therefore, we expected to see a relationship between simulation complexity and response time which held for trials in which participants made a prediction by simulation but failed to hold for trials where participants were instead using visual cues which facilitated more coarse topological predictions.

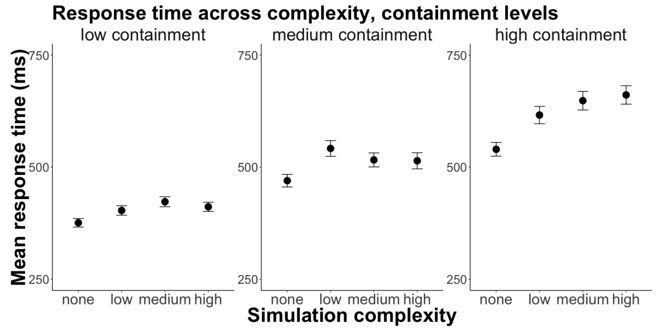
Results

Two of the 81 participants were excluded from analysis due to technical difficulties logging their data. For each participant, we excluded data from the 64 distractor trials. These were included in the experiment to prevent the inference that all trials would have a more and a less contained target. However, the data from these trials is not relevant to the present analyses. All subsequent analyses were therefore conducted with data from 79 participants over 192 trials (twelve trials for each of the four scenarios, rotated each of 0, 90, 180, and 270 degrees). For all analyses, response times were log-transformed to account for their skewed distribution (Whelan, 2008) but transformed back for reporting and display.

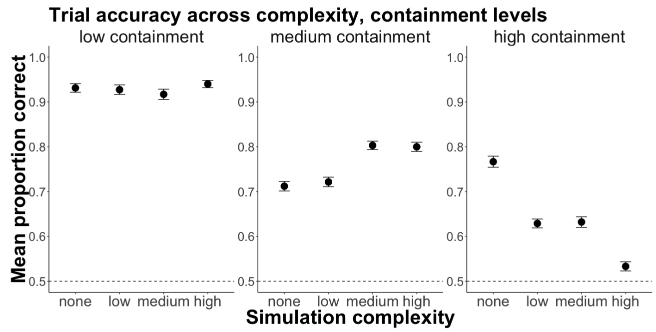
Response times

To assess whether participants were avoiding costly simulations when simulations were particularly uncertain and topological conditions supported more efficient predictions, we examined average response times across each level of complexity and containment. The results are illustrated in Figure 3a. We were interested in comparing response times in low-containment trials to high-containment trials, where an efficient topological prediction about which target the ball would hit was available. Rather than a stabilization or even a decrease in response time as complexity increased in high containment trials (signaling a switch to topological prediction), Figure 3a shows that response times increased progressively as containment increased from low to high and within each containment level as complexity increased from none to high. Moreover, the high-containment trials were slower, and less accurate (Figure 3b), than low-containment trials.

In a repeated measures ANOVA, response times vary with containment and complexity, ($F(2, 156) = 55.63$, $p < 0.001$ and $F(3, 234) = 8.87$, $p < 0.001$, respectively). However, consistent with the fact that participants are not treating complexity differently in high containment trials, there is no containment-complexity interaction ($F(6, 468) = 0.487$). Participants relying on topological information to infer which target the ball would hit in high containment trials would have



(a) Response times across conditions



(b) Accuracy across conditions

Figure 3: (a) Mean response time across containment and complexity levels. Despite the availability of simple topological predictions in the high containment, high complexity trials, response time is highest. (b) Mean accuracy across containment and complexity levels. Accuracy steadily decreases at higher containment levels, even though more contained trials would seem to make prediction more certain.

been able to do so quickly. As complexity increased, so too would the time required to simulate the ball's possible outcomes. Therefore, predictions made via topological analysis in high containment, high complexity trials could potentially be done in less time than required for prediction by simulation in trials with the same degree of complexity but lower containment. Even with complexity levels which make simulation difficult and topological information which makes prediction simple, participants showed no sign of using a containment heuristic.

Accuracy

In light of our findings that response times both increased as complexity increased within each containment level and also increased across containment levels, one interpretation is that this pattern was a result of a speed-accuracy tradeoff. Insofar as additional complexity in a given scenario made simulation more difficult and uncertain, participants may have spent more time confirming their predictions without any other change in their simulations or prediction strategies. To test this, we looked at each participant's accuracy in a given containment and complexity level (there are 16 trials in a given containment and complexity level for each participant). The

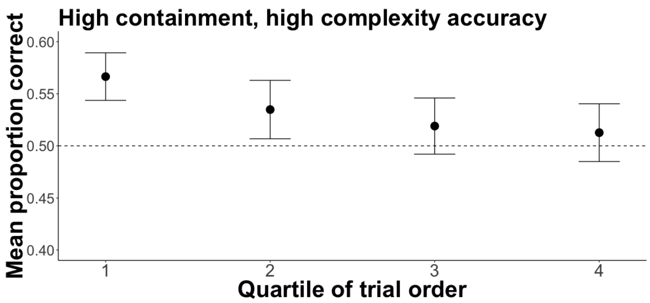


Figure 4: Performance of participants in *high containment, high complexity trials only* by trial order quartile. Accuracy remains close to chance and does not improve over the course of the experiment, suggesting that participants likely did not switch at any point to topological inference cues or other strategies that would have improved their accuracy.

mean accuracy proportions across all participants for each complexity and containment level are shown in Figure 3b.

In contrast to what would be predicted by a speed-accuracy tradeoff in which the containment and complexity levels that participants spent the most time on also have the highest accuracy, mean accuracy steadily decreases from low to high containment scenarios. In low containment trials, mean accuracy was above 90% across all complexity levels, while in high containment and high complexity trials, where participant response times were the largest, accuracy was only nominally above chance (95% CI 51.3% - 55.4%). In a repeated measures ANOVA, both containment and complexity accounted for a significant portion of the variance in accuracy ($F(2, 156) = 829.5$, $p < 0.001$ and $F(3, 234) = 15.7$, $p < 0.001$, respectively), as well as the interaction between them ($F(6, 468) = 61.79$, $p < 0.001$). As the containment and complexity of trials increased, participants spent more time making judgments and their accuracy decreased: these data are inconsistent with an account of prediction in which people process topological features to make the judgment as efficiently as possible. One alternative is that people persist in simulating outcomes in such trials even when alternatives are readily available. Under this account, participants would be expected to simulate more as complexity increased in order to overcome the uncertainty imposed by increases in complexity. They might do this even when increasing levels of containment made topological predictions simple.

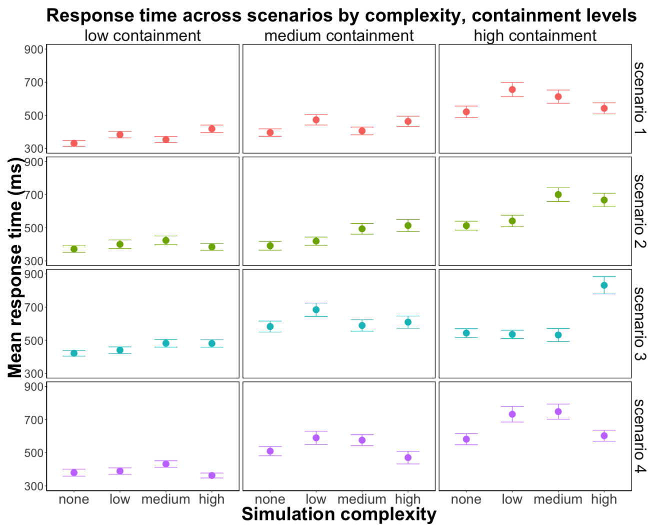
Strategy changes

Another interpretation of the current results is that people may have *eventually* switched to heuristic-based strategies in the more complex trials, but not right away. We predicted that the difficulty of simulation on high complexity trials would encourage participants to employ alternative inference strategies where available. But it may be that the complexity of a trial in and of itself is insufficient to induce strategy change. For example, participants might need to see several complex trials and infer that high complexity trials are likely to recur and are not “one offs”. Or, participants might overestimate the accuracy of simulation-based inferences: only after getting wrong answers on complex trials would they pursue alternative inference strategies.

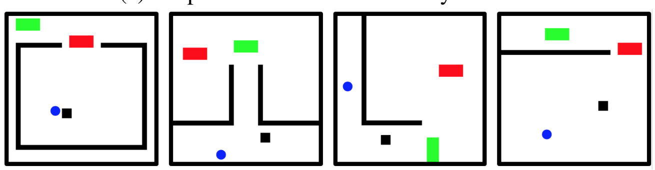
If participants were switching to heuristic-based strategies as a result of familiarity with the task or low accuracy on complex trials, we might expect a difference between high complexity, high containment trials encountered *earlier* versus *later* in the experiment. This difference would be revealed in changes in accuracy over the course of the experiment: if participants eventually ended up using a topological inference strategy for these high containment, high complexity trials, we would expect near perfect accuracy for any such trials. Figure 4 shows accuracy on *high containment, high complexity trials only*, arranged by the trial order quartile in which participants saw them. Participants performed relatively poorly on the high containment, high complexity trials at the outset. Critically, there is no sign of improvement over the course of the experiment: in an ANOVA with participants’ mean accuracy by quartile, accuracy did not vary significantly across quartiles ($F(1, 78) = 2.359$, $p = 0.129$). If participants had switched to a topologically based inference strategy, we would expect an increase in accuracy since high complexity, high containment trials enable a very confident containment-based solution. Figure 4 suggests that directionally, participants appeared to get worse on the high accuracy, high containment trials and remain fairly inaccurate throughout the experiment.

Scenario and rotation differences

A third account for the higher response times and lower accuracy as containment and complexity increased is that this overall pattern reflects a great deal of variance across scenarios. In a repeated measures ANOVA of response time that adds scenario on top of containment and complexity, there are significant main effects of containment and complexity (as described before) as well as scenario ($F(3, 234) = 64.89$, $p < 0.001$), reflecting the fact that participants’ response times seemed to vary across scenarios. In Figure 5, we show mean response times across containment and complexity levels but further broken down by scenario. The pattern of response times is fairly consistent for low containment trials but the directionality of response times as complexity increases in high containment trials varies across scenarios. In scenarios 1, 2, and 4, response times in high containment trials stabilize or diminish at higher complexity levels, which is qualitatively consistent with our hypothesis that participants would make faster predictions when topological conditions supported a coarse analysis and made simulation highly uncertain. Indeed, the effects of containment and complexity are not homogeneous across scenarios, revealed by significant interactions between scenario and containment ($F(6, 468) = 6.935$, $p < 0.001$), and scenario and complexity ($F(9, 702) = 4.662$, $p < 0.001$). The three-way interaction between scenario, containment, and complexity is weaker, but also significant ($F(18, 1404) = 1.626$, $p = 0.047$), indicating that the



(a) Response time broken down by scenario



(b) High containment example of each scenario (1–4)

Figure 5: (a) Response times are consistent across scenarios in lower containment and complexity levels but diverge considerably at higher containment and complexity levels. (b) A high containment (low complexity) trial for each scenario. Complexity was increased by adding more square obstacles.

pattern in scenario 3 is quite unusual. However, we cannot confidently conclude that any of the scenarios would reliably produce the sort of two-way interaction between containment and complexity that our hypothesis predicts.

Finally, it is worth noting that even though rotated versions of the trials were identical in configuration and ball movements, simply turned 90, 180, or 270 degrees, participants may have treated rotated versions of the trials differently. A repeated measures ANOVA of response time as a function of scenario and rotation found that rotation accounted for a significant amount of the variance ($F(3, 234) = 6.995$, $p < 0.001$), scenario was significant (as outlined above) and that there was a significant interaction between scenario and rotation ($F(9, 702) = 2.939$, $p = 0.002$). Whether this reflects some sort of bias towards e.g. the targets being at the top of the screen is unclear.

Conclusion

In this study we presented participants with physical prediction tasks that simultaneously varied the degree to which a simple containment heuristic could be used to make effective predictions and the complexity required to simulate outcomes instead. Our hypothesis was that as increasing complexity made simulations more and more uncertain and effortful, participants would pursue less costly topological prediction strategies. When conditions permitted such knowledge-based predictions, response times would reflect the rapid and efficient use of containment heuristics. We found no evidence of participants flexibly using heuristics when simulation was complex. In fact, participants spent the longest on trials that had the highest degree of containment; meanwhile, their accuracy was lowest on these same trials.

Why might participants have spent more time and been less accurate on trials where a simple containment-based prediction was available? First, it's possible that the structure of the task at the outset biased participants towards a simulation-based strategy in a way that might have been difficult to overcome, even when complexity of trials made simulation difficult. Earlier work that used static control stimuli in a similar task found evidence that people used simulation even with static stimuli (Smith et al., 2017). Therefore, it's possible that participants had a high "fixedness" when confronted with complex trials. Additionally, it has been shown that when explicitly instructed to apply distinct simulation strategies, participants show notable performance differences on mental rotation tasks (Flusberg & Boroditsky, 2011). In the present study, participants were not instructed to simulate or make a containment-based inference and were solving the problems as they naturally would, but future work might look at how instructions play a role in guiding more efficient strategies.

Alternatively, Smith et al. (2017) suggest that if simulation and alternative prediction strategies are running in parallel, detecting scenarios in which participants switch from a default simulation-based prediction to a more qualitative one that is quicker but more coarse might require enough time for simulation to *run out*. In the present study, average response times in the slowest high containment, high complexity trials were still less than one second on average (see Figure 3a). Perhaps participants, upon finding that they were not able to make an accurate simulation-based prediction on these trials, still did not spend long enough attempting an accurate answer to detect the containment relationship or make a prediction based on such a holistic topological feature. Insofar as the higher containment and complexity trials simply required longer to visually process the full scene, participants may have resorted to an even quicker and more general heuristic in order to respond quickly, such as the target that was the shortest Euclidean distance or seemed most directly along the ball's initial path irrespective of obstacles. Alternatively, participants may have simply persisted in slower and less efficient simulations on high containment, high complexity trials rather than pursue alternate strategies (Hamrick et al., 2015). Future research will need to carefully design stimuli in order to control for the many ways participants might make predictions and consider other hypotheses that allow for manipulation of the uncertainty of simulations during prediction.

Acknowledgments

We thank our reviewers for their thoughtful suggestions.

KAS is supported by CBMM funded by NSF STC award CCF-1231216, and ONR grant N00014-13-1-0333.

References

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. (2018). Modeling human intuitions about liquid flow with particle-based simulation. *arXiv:1809.01524 [cs, q-bio]*.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. In *Proceedings of the national academy of sciences.*

Callaway, F., Hamrick, J. B., & Griffiths, T. (2017). Discovering simple heuristics from mental simulation. In *Proceedings of the 39th annual meeting of the cognitive science society.*

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

Davis, E., & Marcus, G. (2016). The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, *233*, 60–72.

Davis, E., Marcus, G., & Frazier-Logue, N. (2017). Commonsense reasoning about containers using radically incomplete information. *Artificial Intelligence*, *248*, 46-84.

Flusberg, S., & Boroditsky, L. (2011). Are things that are hard to physically move also hard to imagine moving? *Psychonomic bulletin & review*, *18*(1), 158-164.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-Tracking Causality. *Psychological Science*, *28*(12), 1731-1744. doi: 10.1177/0956797617713053

Gigerenzer, G., & Todd, P. (1999). *Simple heuristics that make us smart*. USA: Oxford University 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.

Hamrick, J. B., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? optimal mental simulation tracks problem difficulty. In *Proceedings of the 37th annual meeting of the cognitive science society.*

Smith, K. A., Battaglia, P. W., & Vul, E. (2018). Different Physical Intuitions Exist Between Tasks, Not Domains. *Computational Brain & Behavior*, *1*(2), 101-118. doi: 10.1007/s42113-018-0007-3

Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (2013). Physical predictions over time. In *Proceedings of the 35th annual meeting of the cognitive science society.*

Smith, K. A., de Peres, F., Vul, E., & Tenenbaum, J. B. (2017). Thinking inside the box: Motion prediction in contained spaces uses simulation. In *Proceedings of the 39th annual meeting of the cognitive science society.*

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

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

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

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