Cognition 254 (2025) 105995

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full length article

Blending simulation and abstraction for physical reasoning

Felix A. Sosa [a,c,*], Samuel J. Gershman [a,b,c], Tomer D. Ullman [a,c]

[a] Department of Psychology, Harvard University, 52 Oxford St, Cambridge MA 02138, USA
[b] Center for Brain Science, Harvard University, 52 Oxford St, Cambridge MA 02138, USA
[c] Center for Brains, Minds, and Machines, MIT, 43 Vassar St, Cambridge 02139, USA

ARTICLE INFO

Keywords:
Intuitive physics
simulation
heuristics
abstract reasoning

ABSTRACT

How are people able to understand everyday physical events with such ease? One hypothesis suggests people use an approximate probabilistic simulation of the world. A contrasting hypothesis is that people use a collection of abstractions or features. While it has been noted that the two hypotheses explain complementary aspects of physical reasoning, there has yet to be a model of how these two modes of reasoning can be used together. We develop a "blended model" that synthesizes the two hypotheses: under certain conditions, simulation is replaced by a visuo-spatial abstraction (linear path projection). This abstraction purchases efficiency at the cost of fidelity, and the blended model predicts that people will make systematic errors whenever the conditions for applying the abstraction are met. We tested this prediction in two experiments where participants made judgments about whether a falling ball will contact a target. First, we show that response times are longer when straight-line paths are unavailable, even when simulation time is held fixed, arguing against a pure-simulation model (Experiment 1). Second, we show that people incorrectly judge the trajectory of the ball in a manner consistent with linear path projection (Experiment 2). We conclude that people have access to a flexible mental physics engine, but adaptively invoke more efficient abstractions when they are useful.

1. Introduction

From catching a phone as it slides off an arm-rest, to tossing keys to a friend, to building a tower of blocks with a child, people adeptly, quickly, and automatically reason about the physics of everyday objects. This 'intuitive physics' is a core component of commonsense reasoning (Kubricht, Holyoak, & Lu, 2017; Lake, Ullman, Tenenbaum, & Gershman, 2017), and much of it is early-developing or potentially innate, cross-cultural, and shared with non-human animals (Bremner, Slater, & Johnson, 2015; Spelke, 2022; Spelke & Kinzler, 2007).

The subjective effortlessness of intuitive physics masks the complexity of the underlying computations. Despite a great deal of progress over the past decade, state-of-the-art artificial intelligence systems still struggle to reason about everyday physical phenomena, including stability (Buschoff, Akata, Bethge, & Schulz, 2023; Zhang, Wu, Zhang, Freeman, & Tenenbaum, 2016), collisions (Smith et al., 2019), trajectories (Chang, Ullman, Torralba, & Tenenbaum, 2016), and physical puzzles (Cherian, Peng, Lohit, Smith, & Tenenbaum, 2023). The inability of our current best engineered AI systems that aim to mimic human intelligence and match human-level intuitive physics shows that the problem is non-trivial, and that we still lack a comprehensive model of how humans engage in intuitive physics.

There have been several different approaches to modeling the computations that underlie intuitive physics. Each approach has had success in explaining different phenomena. One approach posits that people reason about physics through an approximate mental simulation of the world (Battaglia, Hamrick, & Tenenbaum, 2013; Hamrick, Battaglia, Griffiths, & Tenenbaum, 2016; Hegarty, 2004; Smith et al., 2024). By contrast, alternatives to mental simulation include reasoning based on visual or physical features (Baillargeon, 2002; Fragkiadaki, Agrawal, Levine, & Malik, 2015; Lerer, Gross, & Fergus, 2016; Mottaghi, Bagherinezhad, Rastegari, & Farhadi, 2016), heuristics (Gilden & Proffitt, 1989; Nusseck, Lagarde, Bardy, Fleming, & Bülthoff, 2007; Proffitt, Kaiser, & Whelan, 1990), logical rule-like structures (Davis, 1990; Siegler & Chen, 1998), qualitative reasoning (Forbus, 1988), deep learning (Piloto, Weinstein, Battaglia, & Botvinick, 2022), and pre-Newtonian theory-like structures (McCloskey & Kohl, 1983). While these alternatives to mental simulation differ significantly in their details, they share a notion of abstracting away from some of the details necessary for a simulation. We briefly consider strengths and weaknesses of mental simulation and these alternative accounts, and consider previous work to unify the two modes of reasoning. Then, we turn to our proposal of how these accounts of physical reasoning can

* Corresponding author at: Department of Psychology, Harvard University, 52 Oxford St, Cambridge MA 02138, USA.
E-mail address: fsosa@fas.harvard.edu (F.A. Sosa).

https://doi.org/10.1016/j.cognition.2024.105995
Received 1 July 2024; Received in revised form 16 October 2024; Accepted 17 October 2024
Available online 4 November 2024

be unified into a single model, and our experiments that test this new "blended model" against mental simulation and alternative models.

Models of intuitive physics based on mental simulation have quantitatively and qualitatively accounted for many aspects of physical judgment, including judgments about object motion (Smith & Vul, 2013), task-relevant and goal-oriented object properties (Bramley, Gerstenberg, Tenenbaum, & Gureckis, 2018), collisions (Sanborn, Mansinghka, & Griffiths, 2013; Smith, Battaglia, & Tenenbaum, 2023), mass (Hamrick et al., 2016), fluids (Bates, Yildirim, Tenenbaum, & Battaglia, 2019), causality (Gerstenberg & Stephan, 2021; Sosa, Ullman, Tenenbaum, Gershman, & Gerstenberg, 2021), and stability (Battaglia et al., 2013). Simulation also captures important aspects of the development of intuitive physics, and has been used to explain physical reasoning in pre-verbal infants (Téglás et al., 2011), build computational models that pass developmental benchmarks (Smith et al., 2019), and explain how children adapt cognitive strategies with respect to life experience (Allen et al., 2021) and goal-directed behavior (Bramley & Ruggeri, 2022). Mental physical simulation has also found support from recent work in neuroscience that points to a "physics engine" in the brain (Fischer, 2021; Fischer, Mikhael, Tenenbaum, & Kanwisher, 2016; Schwettmann, Tenenbaum, & Kanwisher, 2019).

Despite the success of mental simulation accounts, they have been criticized on both empirical and theoretical grounds (Kubricht et al., 2017). Empirically, people sometimes deviate from the predictions made by simulation. For example, people may systematically deviate from the predictions of an accurate simulation when they reason about the trajectories of moving objects (McCloskey, 1983; McCloskey, Caramazza, & Green, 1980), predict trajectories past collision (DiSessa, 1982), judge the relative probabilities of events happening (Ludwin-Peery, Bramley, Davis, & Gureckis, 2021), evaluate the mechanics of rotation (Proffitt et al., 1990) and pulley systems (Hegarty, 1992), and so on. Some of these issues can be resolved through a mental simulation that uses principled approximations (Bass, Smith, Bonawitz, & Ullman, 2021; Ullman, Spelke, Battaglia, & Tenenbaum, 2017), or by using stimuli that are more familiar (Kaiser, Jonides, & Alexander, 1986), or more realistic/dynamic (Kaiser, Proffitt, & Anderson, 1985; Kim & Spelke, 1999) but the general point stands. Mental simulation has also been challenged on theoretical grounds (Ludwin-Peery et al., 2021; Marcus & Davis, 2013). One of the principal theoretical challenges is that simulating the motion of every objects in a scene is slow, computationally expensive, memory taxing, and unnecessary compared to simpler alternatives.

While the non-simulation approaches to intuitive physics are not a unified camp, they often point to simplicity and speed as advantages. For example, according to models based on a bottom-up analysis of perceptual features (Barsalou, 1999), people assess the impending collapse of a teetering tower by considering features like its height or top-heaviness. Such calculations are seemingly done faster than a program that constructs the scene and runs it forward in time. Also, more recent iterations of feature-based accounts leverage deep neural networks to operationalize learning physical reasoning from experience. This includes learning to reason about stability (Conwell & Alvarez, 2019; Lerer et al., 2016), trajectory prediction (Piloto et al., 2022; Zhang et al., 2016), and the inference of physical variables like mass (Wu, Yildirim, Lim, Freeman, & Tenenbaum, 2015).

While useful, feature-based accounts face two major challenges. First, they do not capture the flexibility of human physical inferences. To go back to the example of block-towers: people can determine the stability of a tower of blocks, but also predict the tower's falling direction, anticipate the scattering distance of the blocks, imagine possible outcomes when two blocks are glued, consider the impact of striking the tower at an angle, and generally answer a host of other questions about the way the scene could unfold (Battaglia et al., 2013). Such generalization is handled by accounts that construct and simulate scenes. By contrast, feature-based models often require retraining for each new query, and exhibit poor generalization from one query to another. A second challenge for feature-based accounts is that quantitative comparisons to human performance often show low correlations between people and these models when tested outside their training sets (e.g. Bear et al., 2021).

Given the successes of, and challenges to, both simulation and non-simulation accounts, it has been proposed that humans have the ability to use both flexible simulation models and efficient abstractions in tandem (Smith et al., 2023). We refer to non-simulation accounts broadly as abstractions due to non-simulation accounts often utilizing less information to perform prediction compared to simulation accounts, for example by relying on high-level visual features rather than low-level variables such as instantaneous velocity, mass, or others required for computing dynamic forces. By utilizing less information to perform prediction, the abstraction often happening in non-simulation accounts occurs at the algorithm or process level. To motivate the synthesis, consider the following thought experiment: place a straight straw vertically above a cup, put a small round bead into the top of the straw, and let go of the bead. Will the bead fall into the cup? The answer is an almost immediate 'yes'. Now suppose the straw is twice as long. The answer is still an obvious 'yes', and it does not take twice as long to answer that. On its own, this is an argument against simulation. However, now consider dropping the bead into a convoluted straw, the kind that might be handed out as a party favor. The answer becomes non-obvious, as the bead might get stuck in the twists and turns. The time to come up with an answer now depends on how convoluted the straw is. But how does one switch between simulation and abstraction in a flexible, efficient manner?

Previous work has suggested that one way by which people might arbitrate between simulation and abstraction is by applying either simulation or abstraction on a per-scenario basis, based on the expected value of that mode of reasoning for that scenario (Smith et al., 2023). Considering our previous thought experiment about straws: the model in Smith et al. (2023) would suggest that for the straight straw, we apply a simple rule that says the bead will fall out the other end of the straw, and for the convoluted straw we would use simulation to determine whether the bead gets stuck in the loops of the straw or passes through and ultimately reaches the cup, and that there is some sort of cost-driven control mechanism outside of the model that determines when to apply one or the other. While this model is able to explain people's behavior on physical reasoning puzzles, we believe that a model of intuitive physics must be able to explain how people select between which mode of reasoning to use in real time, not on a per-scenario basis. For example, what if we had a semi-convoluted straw with only one small loop whose tails are long and straight? Do we use simulation or abstraction for the whole straw or do we selectively use simulation and abstraction where it is best fit for the straw, e.g. abstraction for the straight portions and simulation for the loop? While the previous work discussed would say it is either/or, we suggest people flexibly balance the inferences made by both modes of reasoning for a single scenario in real time.

In this work, we formalize the suggestion that people adaptively trade off simulation and abstraction based on threshold conditions that are computed in real-time, allowing for our model to use simulation and abstraction on-the-fly as it sees fit (Fig. 1). We test our proposal experimentally using established experimental designs from previous work in intuitive physics (e.g. Allen, Bakhtin, Smith, Tenenbaum, & van der Maaten, 2020; Ludwin-Peery et al., 2021). The blended model suggests that when people mentally simulate the motion of objects, they periodically check whether certain threshold conditions are met if simulation were to be replaced by an appropriate abstraction at that point in time. If the conditions are met, the person's belief of the next state of the scene is adjusted according to the appropriate abstraction, rather than being updated according to simulation. This real-time trade-off short-circuits simulation and saves on computational resources. If the conditions are not met, mental simulation continues as usual. Importantly, it is *not* the case that our model completely

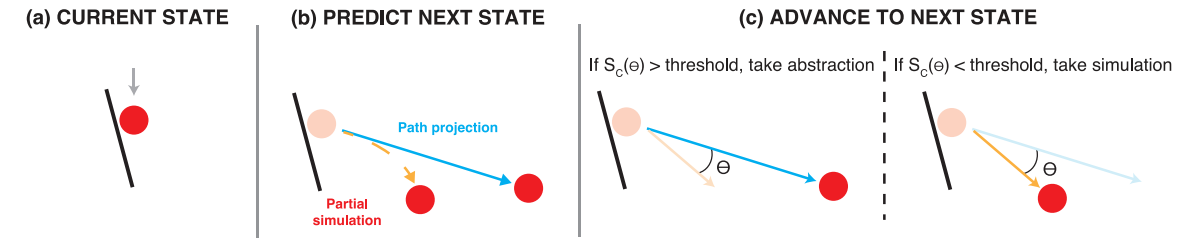

Fig. 1. Single pass through the blended model: (a) The model is given the current state as input, including the position and velocity of the ball, (b) The model predicts two future states, one using N steps of partial simulation (orange line), and one using a path-projection abstraction of length D (blue line), (c) the model determines the similarity of the two possible states (cosine similarity between resulting translation vectors). If the similarity $S_C(\theta)$ is above a threshold E, path projection determines the next state. If similarity is below the threshold, partial simulation determines the next state. The next state is handed back to the model as the new current state and the process is repeated until the end state of the scenario. Importantly, if the amount of partial simulation steps N is smaller than the length of the path-projection D, this can result in plausible cognitive resource-saving. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

simulates a scene, also considers an abstraction, and then trades off the two. Such a model would be both cognitively implausible and not result in any resource savings over simply simulating the scene. Rather, as we detail below and in the Modeling section, the blended model computes a possible next state of the scene according to a partial simulation of N steps, and computes an abstraction of path length D *in parallel*. Computational and cognitive savings emerge from this process in the event that abstraction is favored over simulation for a given inference step, and if the path traveled by the abstraction D is greater than the path traveled by the N steps of simulation.

While many different abstractions might be used to get around simulation, here we focus on an abstraction we term *linear path projection*. The abstraction is blended with simulation as follows: At any given moment t in a mental simulation, the blended model computes two possible future states: one according to abstraction, and one according to partial simulation. In our case, the future states are predicted by translating an object in a straight line path that starts at the moving object's current location x_t, y_t, and extends for some specified distance D in the direction of the object's velocity. This leads to a prediction of the next state according to abstraction. At the same time, a second possible future state of the object is predicted by mental simulation to a minimal degree of N ticks in parallel with abstraction. This partial simulation is what ensures that abstraction can lead to computational savings, as discussed above. If the similarity of the future trajectories from abstraction and partial simulation is higher than a threshold, the blended model updates its internal belief of the next state of the scene according to the prediction made by abstraction. If similarity is below a threshold, the partial simulation is used to update the blended model's state instead. The process is then repeated (see the Modeling section for more information). To the degree that the straight line projection moves the object past the point of partial simulation (i.e., D>N), it can save significantly on the computational costs associated with simulation. However, such computational cost-cutting comes with a price: it can lead to systematic deviations from the true trajectory.

In two experiments, we contrasted the predictions of the blended model with those of two models of pure simulation and two models of pure abstraction. Both experiments asked participants to view 2D scenes in which a ball was dropped under gravity, and to rapidly judge whether the ball will eventually reach a goal location. Experiment 1 found that response times monotonically increased in simulation time, as predicted by simulation models. However, as predicted by the blended model, response times were shorter for judgments with straight paths compared to judgments without straight paths, even when the true simulation time was the same for both. Experiment 2 found that people give systematically wrong answers, as predicted by the use of linear path projection. Taken together, these results suggest people flexibly combine simulation and abstractions to solve physical reasoning problems, as captured by a blended model that uses a combination of simulation and a visuo-spatial abstraction.

2. Methods

2.1. Transparency and openness statement

The experiments and analyses presented here were not pre-registered. We make all data, models, and analyses publicly available in the project's Github repository: https://github.com/flxsosa/physics_abstraction.

2.2. Participants

Participants were recruited using Prolific (www.prolific.co) for both Experiment 1 (n = 50, of which 20 identified as female, M_{age} = 29) and Experiment 2 (N = 49, of which 20 identified as female, M_{age} = 36). All participants provided informed consent before partaking in both experiments, and were compensated for their time ($12.00/h). Three participants were excluded from Experiment 1, and two participants were excluded from Experiment 2 due to failing comprehension checks.

2.3. Procedure

For both experiments, participants were told that they would be viewing videos depicting physical scenes containing a ball, a goal, and a random number of slides. Participants observed the ball falling for 1 s (64 steps of simulation), after which it vanished. Subsequently, participants were tasked with determining whether the ball would ultimately collide with the goal (see Fig. 2). The order of the stimuli were randomized across participants. Participants were told to use their keyboard to respond with "Yes" or "No", with the assigned keys for either response being randomized across participants. For each scene, a participant would watch a short 1-second video of the ball beginning to fall, as it fades out of scene. Participants were told to imagine the ball was still falling, and to respond as quickly as possible to the query.

Before the main experiment, participants were also shown four example stimuli, to familiarize them with the physical dynamics of the scenes, and to check their comprehension of the task. The four initial stimuli consisted of two videos in which the ball did not fade out (familiarization), followed by two videos in which the ball faded out after 1 s just as it would in the actual trials (comprehension). Comprehension was passed if participants correctly predicted whether the ball collided with the goal or not at the end of the video. Only after passing the comprehension trial were participants allowed to move on to the main experiment. Participants who failed the comprehension check more than two times were removed from further analysis (N = 3 for Experiment 1, N = 2 for Experiment 2).

2.4. Stimuli

Experiment 1 used 48 stimuli, and Experiment 2 used 58 stimuli. Order of presentation for stimuli were randomized across participants for both experiments. All stimuli were designed using the 2D rigid-body physics engine Pymunk.[1] Each stimulus showed a single scenario with a ball (a randomly-colored circle), goal (a randomly-colored rectangle), and slides (narrow, rigid black rectangles) set atop a gray background. After generating the stimuli, a random subset of stimuli was flipped about the horizontal axis before being shown to participants. All stimuli can be viewed in the project's GitHub repository (see our Transparency and openness statement above).

2.4.1. Physics engine parameters

The objects in Pymunk are rigid bodies defined about a set of vertices. In our stimuli, the ball is defined as a rigid circle with radius of 20 and a mass of 10, and the slides are defined as rigid polygons with infinite mass. Additionally, Pymunk also accepts values for global variables that govern the dynamics of simulations, such as gravity, object elasticity, and friction. For our work here, we set these values such that simulations appeared realistic, and note that the values themselves are unitless. Gravity set was to a value so that objects fell at realistic speeds (i.e., $(0, 300)$), elasticity was set to a value that produced realistic bouncing when a ball collided with a slide (i.e., 1.0). Friction is replicated via damping in Pymunk, which was also set to a value that produced realistic effects of friction on objects (i.e., 0.6). The settings of these parameters and their effects on the dynamics of the scenes in our experiments can be observed in the projects Github repository under `experiments/experiment1/img/stimuli/comprehension`. The URL for the repository is noted in our Transparency and openness statement above.

2.4.2. Experiment 1

The stimulus set for Experiment 1 consisted of 48 scenarios. The object placements in each stimulus were randomly generated according to a 3 x 2 x 2 design (Simulation Time x Collision x Trajectory), with 4 unique instances of each combination setting. Simulation time varied how many steps the physics engine needed to take to complete the simulation, and was either Short (0–150 steps), Medium (150–300 steps), or Long (300–450 steps). Collision was a binary outcome, whether the ball and goal collided according to pure simulation ("Yes" or "No"). Trajectory was a binary property of the scenario, indicating whether the ball's trajectory formed a straight line or not ("Yes" or "No").

2.4.3. Experiment 2

The stimulus set for Experiment 2 consisted of 58 scenarios. In order to generate response time curves that qualitatively differ between the different models, the stimuli were designed with a multi-step process: We first sampled random scenario configurations (random starting positions the ball and goal, random number and position of slides. We next simulated the trajectory of the ball according to ground-truth simulation. We then picked 29 evenly-spaced positions along the simulated trajectory of the ball, and generated response time predictions for each, which together gave response-time-curves for each scenario. We chose 29 positions as this number led to visually distinct stimuli, where each goal placement was sufficiently visually different (i.e., if we used more placements, stimuli would begin to look similar). We repeated this process until the predicted response time curves from pure simulation and the blended simulation and abstraction model qualitatively differed, at which point we accepted the stimuli created by the generative process. We performed this procedure twice, leading to 58 generated scenarios.

[1] www.pymunk.org.

2.5. Modeling

We consider five models: pure simulation, pure simulation with velocity damping, pure abstraction, pure abstraction with a collision constant, and a blended model that combines pure simulation and pure abstraction. Each model is defined by a transition function that moves the physical scenes from the current state to the next state. In the experiments reported in this paper, the only dynamic object is the ball. The state corresponds to the x–y ball position, and the state evolution equation specifies the ball dynamics. To generate judgments about whether the ball will hit the goal object, the evolution equation is run forward until the ball contacts either the goal object or the ground. In the context of pure abstraction with a collision constant and pure simulation with velocity damping, collisions between the ball and the slides are also recorded and used at inference time according to the respective model's transition function. After a given model generates a judgment for a scene, the execution time of the model, τ, is then transformed into the human response time by a noisy linear transformation:

$$\hat{\mathrm{RT}} = \beta_0 + \beta_1 \tau + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \lambda)$ models random noise in the response generation process. The coefficients (β_0, β_1) and noise variance (λ) are fit to human response time data, as described below. In addition to the noise inherent to the response, for each model considered, we assume noise over the initial position of the ball, corresponding to imperfect perception or memory, as is standard in many intuitive physics models (c.f. Battaglia et al. (2013)). We assume this positional noise is a zero-mean 2-dimensional Gaussian with standard deviation σ, which is a free parameter fit to data as described below.

Pure Simulation. The state evolution equation for the pure simulation model uses a deterministic, 2-dimensional, rigid-body physics engine (Pymunk, www.pymunk.org) that updates the ball position according to Newtonian dynamics, approximated in discrete time using Euler integration

Pure Simulation With Velocity Damping. We develop another model of pure simulation where we add uncertainty over post-collision velocity on top of the zero-mean Gaussian noise described above. Uncertainty over the velocities resulting from collisions has been suggested as a source of additional uncertainty that might contribute to people's systematic errors in physical reasoning (Sanborn et al., 2013; Smith & Vul, 2013). Here, we consider a model that approximates these sources of uncertainty by sampling post-collision velocities from a Beta distribution whose parameters are fit to empirical response time. The state evolution equation for the pure simulation model with velocity damping uses the same deterministic, 2-dimensional, rigid-body physics engine as pure simulation. Upon collision between the ball and any other object in the scene, a damping value δ is sampled from a beta distribution with parameters α and β, which reduces the ball's resultant velocity post-collision proportional to δ. For example, a $\delta = 0.1$ reduces the ball's velocity post-collision by 10%.

Pure Abstraction. The state evolution equation for the pure abstraction model is based on a linear projection from the ball's current position. The abstraction model uses the simulation engine to determine the ball's current heading (in this weak sense, it is not truly "pure"). It then translates the ball in that direction until either a maximum distance D is reached, or the ball collides with another object. In the case of a potential collision, the ball is projected up until that point of collision, rather than the full length D so as to prevent physically impossible predictions, such as objects teleporting through each other. The state evolution equation has 1 free parameter (D).

Pure Abstraction With Collision Constant. Previous work demonstrated that response time in physical prediction tasks increases with the number of collisions as well as the ground-truth simulation time (Hamrick, Smith, Griffiths, & Vul, 2015). As an alternative to pure abstraction that is in line with this proposal, we developed a derivative

of pure abstraction that includes a constant for the number of collisions encountered during abstraction. The state-evolution equation for the pure abstraction with collision constant model is the same as the pure abstraction model, with an added coefficient in its noisy linear transform that accounts for the number of collisions between the ball and any object in a given scene. For this model, the noisy linear transformation that maps model-execution time to human-response time is:

$$\hat{RT}(s) = \beta_0 + \beta_1 \tau + \beta_2 C_A(s) + \epsilon \quad (2)$$

, where C_A is the number of collisions recorded while evolving the scene s according to the pure abstraction model A.

Blended model. Pure simulation and pure abstraction are combined in the blended model by deciding at each iteration which state evolution equation to use (see Fig. 1). The exact state-evolution algorithm used by the blended model is as follows:

Algorithm 1 Blended model state evolution algorithm

$s_t \leftarrow s_0$ ▷ Get starting state of scene
while $s_t \neq$ End **do** ▷ Repeat until end state of scene is reached
  $s^{\pi}_{t+1} \leftarrow \pi(s_t;\ N)$ ▷ Infer future state according to partial simulation of N steps
  $s^{A}_{t+1} \leftarrow A(s_t;\ D)$ ▷ Infer future state according to abstraction of length D
  $\epsilon \leftarrow S_C([s_t, s^{\pi}_{t+1}], [s_t, s^{A}_{t+1}])$ ▷ Compute similarity between predictions
  if $\epsilon > E$ **then** ▷ Predictions are similar; set next state to abstraction
    $s_t \leftarrow s^{A}_{t+1}$
  else ▷ Predictions are not similar; set next state to simulation
    $s_t \leftarrow s^{\pi}_{t+1}$
  end if
end while

Unlike Pure Simulation, simulation in the Blended Model is limited by the parameter N: Simulation in the blended model evolves the state forward by N ticks, where N can possibly be less than needed to reach the end of the ground-truth trajectory. N is empirically determined from the data gathered in Experiment 1. Abstraction in the Blended Model evolves the state forward the same as it does for Pure Abstraction. If the cosine similarity S_C between the simulation and abstraction vectors is greater than a threshold E, abstraction (the computationally cheaper option) is used to project the ball forward a distance D. If the cosine similarity is less than the threshold, simulation is used for N iterations (i.e., N steps of Euler integration). Potential computational savings are born out of the difference between N and D. Additionally, this similarity metric acts as a regularizer for the blended model, preventing abstraction from making significant errors (e.g. projecting the ball infinitely upwards against gravity) unless those errors are within the bounds of the threshold of the similarity metric. The state evolution equation has 3 free parameters (E, D, N) and can be formalized with the following piecewise state evolution equation:

$$s_{t+1} = \begin{cases} \pi(s_t;\ N) & S_C([s_t, \pi(s_t; N)], [s_t, A(s_t; D)]) < E \\ A(s_t;\ D) & S_C([s_t, \pi(s_t; N)], [s_t, A(s_t; D)]) > E \end{cases} \quad (3)$$

where $\pi(s_t;\ N)$ and $A(s_t;\ D)$ are the pure simulation and pure abstraction state evolution equations, parameterized by N and D, respectively, and S_C refers to the cosine similarity metric. During inference in the blended model, the ball's velocity is updated according to whether simulation or abstraction sets the blended model's prediction. In the event that simulation is used to update the blended model's prediction of the next state, the ball's velocity is integrated according to the physics engine's Euler integrator. In the event that abstraction is used to update the blended model's prediction, the ball's velocity is simply preserved from the input state.

Collision handling for the blended model depends on the method being used: collisions encountered by simulation are addressed by the physics engine's collision handler and Euler integrator, and collisions encountered by the abstraction are handled via the short-cutting method described above.

Parameter estimation and model comparison. All parameters for all models were fit to response times averaged across participants for each scene, using maximum likelihood estimation (the computational cost of running the physics engine meant that it was not feasible to fit parameters to individual participants). Estimates of the state evolution parameters were found using grid search; estimates of the linear coefficients and response noise variance were found using ordinary least squares. The best fitting parameters for each model are summarized in Table 1. The maximized log-likelihood was used to calculate the Bayesian information criterion (BIC) for model comparison. Importantly, model fitting was only performed on data from Experiment 1.

3. Results

3.1. Experiment 1: A blended model of simulation and abstraction accounts for response times better than pure simulation or pure abstraction

In Experiment 1, we used response time data in a physical prediction task to discriminate between simulation, abstraction, and a blended model. Participants saw a series of 2D videos depicting Plinko-style scenes (see Fig. 2), which included a ball falling under gravity, a goal object, and physical barriers ('slides'). The videos showed the ball falling for 1 s, after which it faded away. Participants were told to assume the ball was still in the scene after it faded away, and tasked with judging whether the ball would eventually collide with the goal.

Across the videos, we varied the length of the ground-truth trajectory that the ball would have taken to get to the goal, according to an accurate simulation model. We were specifically interested in the relationship between the length of this trajectory, and the time it took participants to respond whether the ball would collide with the goal. Different models of physical reasoning predict different response time curves as a function of the ground truth trajectory (see Fig. 3): If people use a mental simulation that moves the ball step-by-step, we would expect response times to increase as the length of the ground truth trajectory increases. However, if people reason by abstracting away from a simulation, we would expect reaction time to depend on visual scene features, such as the existence of a straight path between the ball and the goal. Lastly, the blended model makes a more nuanced prediction, according to which response time should be a function of both the ground-truth trajectory length, and the existence of straight paths between the ball and the goal.

Experiment 1 found evidence for the blended model. In line with the simulation hypothesis, we found that response time overall increased with simulation time across the three levels of low, medium, and high simulation time (see Fig. 4a). The ordering of the response time from low to high was significant by a Mann–Whitney U test that examined relationships between all categories : short vs. medium $p < 0.001$, medium vs. long $p < 0.001$, $U = 1.99e^5$, short vs. long $p < 0.001$, $U = 1.00e^5$. However, deviating from the simulation model, we found that within the three simulation time levels, scenes with an observable straight path between the ball and the goal resulted in significantly shorter reaction times than scenes with non-straight paths (see Fig. 4b). Two forms of ordinary regression were performed comparing model and human data. The first ordinary regression was fit using only mean simulation time, and coefficient's of determination (R^2) were observed for each of the investigated models against human data (see Fig. 5). The blended model had the highest $R^2 = 0.71,\ p < 1e^{-}13$, followed by the pure abstraction model with $R^2 = 0.63,\ p < 1e^{-}11$, then the pure simulation model with $R^2 = 0.45,\ p < 1e^{-}7$, then the pure abstraction model with a collision constant with $R^2 = 0.15,\ p = 0.006$, and finally

Table 1
Parameter estimates. N is the number of simulation steps taken by the blended model (i.e., the amount of "minimal simulation" the blended model performs per forward pass). E is the cosine similarity threshold; if the threshold is met then the blended model chooses the prediction made by abstraction over simulation. D is the length or distance of the path projection abstraction. σ is the standard deviation of the zero-mean Gaussian noise injected into the starting position of the ball. λ is the noise variance of the response generation process (see Eq. (1)). $(\beta_0, \beta_1, \beta_2)$ are the coefficients of the linear map between model outputs and response time predictions (see Eqs. (1) and (2)).

Model	N	E	D	σ	λ	β_0	β_1	β_2
Blended	5	0.9	75	(0.01, 0.01)	0.069	794.93	6.04	N/A
Pure abstraction	N/A	N/A	75	(0.01, 0.01)	0.24	1189.64	0.073	N/A
Pure simulation	N/A	N/A	N/A	(0.04, 0.04)	0.13	908.98	1.35	N/A
Velocity damping	N/A	N/A	N/A	(0.0, 1.11)	0.83	1173.52	0.069	N/A
Abstraction + Coll	N/A	N/A	37.55	(0.22, 1.33)	0.56	1228.52	−1.72	3.46

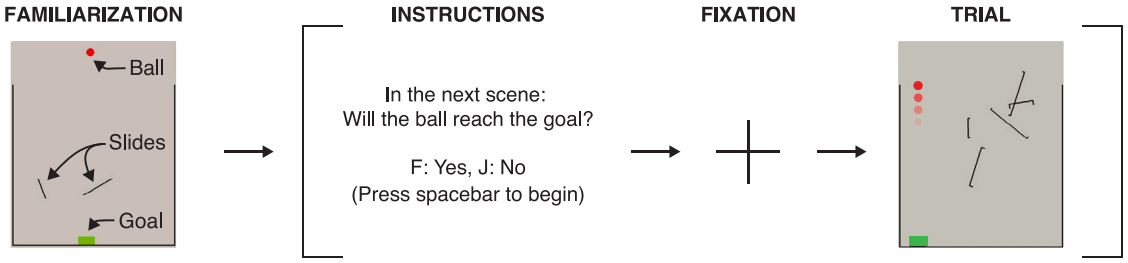

Fig. 2. Experiment design. Participants were first familiarized with the overall scene used in the upcoming trials. Participants were then instructed to determine as quickly as possible whether the ball will collide with the goal at the end of its trajectory. Participants then viewed dynamic scenes in which a ball fell under gravity and faded after 1 s. Participants could respond at any time.

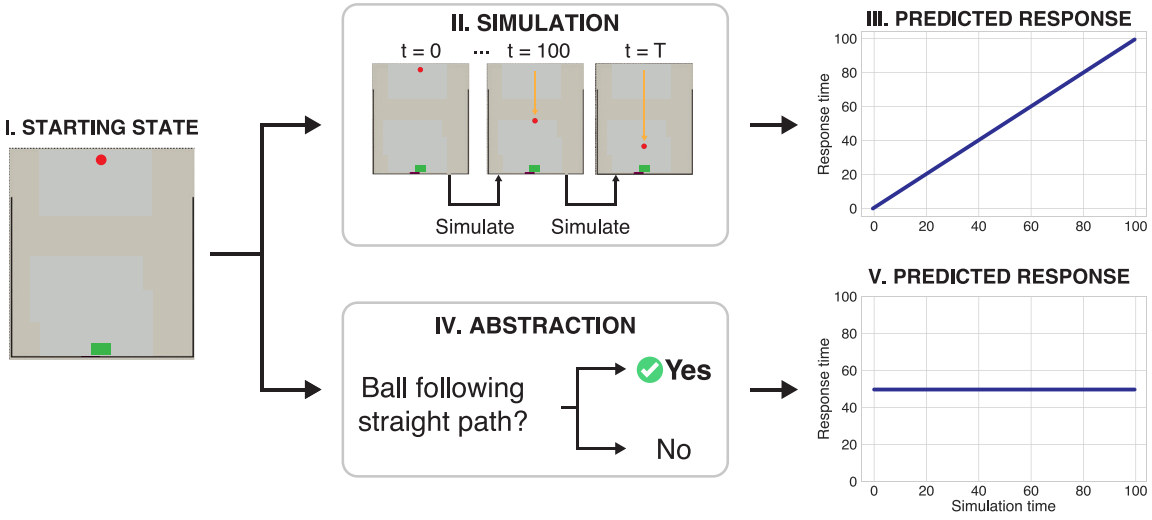

Fig. 3. Illustration of response time predictions. (I) The starting state of a scene. (II) Simulation updates the scene step by step. (III) Under simulation, response time to judge whether the ball will reach the goal scales with the number of simulation steps. (IV) Abstractions can be applications of rules such as "Is the ball following a straight path?". (V) The amount of time it takes to apply an abstraction to a scene is independent of simulation time, and should lead to a constant response time.

the pure simulation model with velocity damping with $R^2 = 0.12$, $p = 0.02$. The second form of ordinary regression lines were fit to the data using mean simulation time and path condition as a categorical variable in order to assess statistical difference in response time between scenes with and without straight paths. Analysis of covariance of the resulting regression lines indicated straight paths (blue line in Fig. 4b) contained both significantly different intercept terms ($p < 1e^{-10}$) and slopes ($p < 1e^{-11}$) than non-straight paths (orange line). These results are captured by the blended model, supporting the claim that participants used an integration of simulation and abstraction. Model comparison using the Bayesian information criterion (BIC) favored the blended model (BIC = 18.95) over the other models of pure abstraction (BIC = 76.89), pure simulation (BIC = 46.68), pure abstraction with a collision constant (BIC = 115.19), and pure simulation with velocity damping (BIC = 133.48).

3.2. Experiment 2: A blended model of simulation and abstraction accounts for accuracy better than pure simulation or pure abstraction

While abstractions may save on computation time compared to simulation, they can introduce errors in their deviation from the ground truth. In Experiment 2 we were interested in whether we could induce systematic errors based on abstraction. We introduced variations in the visual features of a different set of scenes from Experiment 1 by using an algorithm for generating stimuli described in the Methods, such as altering the horizontal distance between slides and the goal. These stimuli contained visual features designed to induce activation of the hypothesized abstraction system, in cases where it would produce the incorrect answer. For example, at first glance, it might appear that the ball in scene 3 of Fig. 6 will narrowly miss a nearby slide and fall directly to the goal. However, under pure simulation, the ball actually collides with the slide and misses the goal.

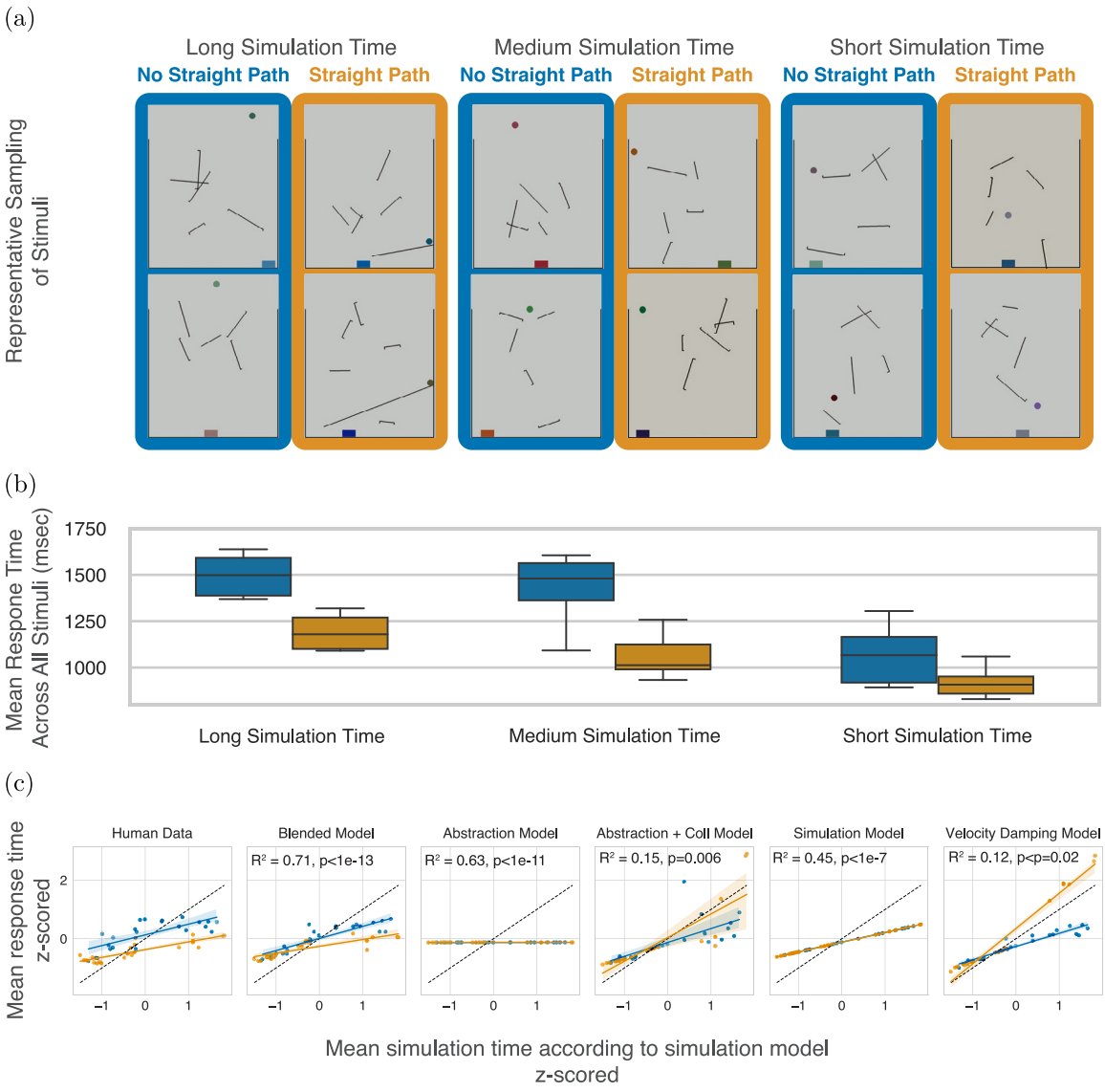

Fig. 4. Experiment 1 results. (a) Example stimuli for each Simulation Time and Trajectory condition setting. (b) Human response time in milliseconds (y-axis) for all stimuli across Simulation Time and Trajectory condition settings (x-axis). (c) z-Scored response time for humans and fitted models (y-axis) against z-scored ground-truth simulation time (x-axis). Each dot represents a single scene, the dashed line represents response time prediction of ground-truth simulation (without fitting to human data). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

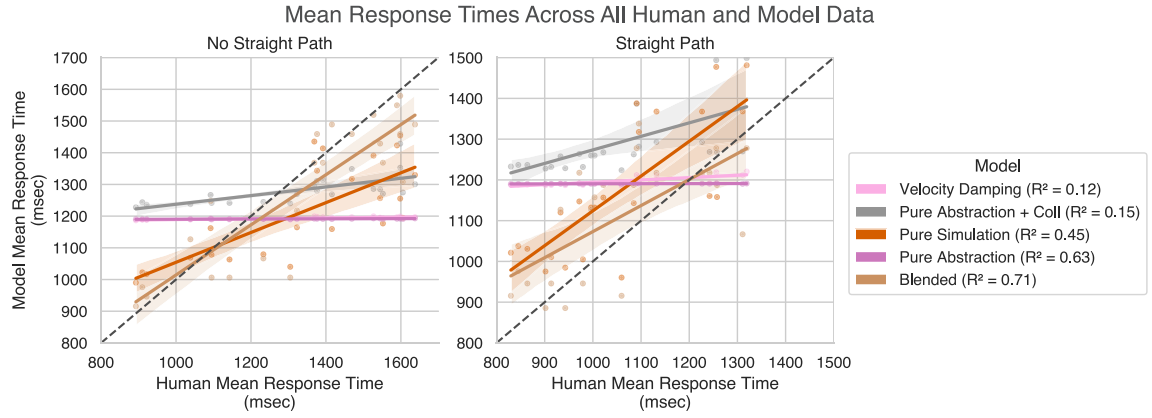

Fig. 5. Experiment 1 model comparisons. Mean human response times (x-axis) are plotted against mean fitted-model response times (y-axis). The black diagonal dashed line is a reference line for human response times; all human response times lie on the dashed line. Each dot represents a single scene. Left: mean response times on scenes with no straight paths. Right: mean response times on scene with straight paths. R^2 values are written for each model, demonstrating the blended model matches human response times best, followed by pure abstraction, pure simulation, pure abstraction with a collision constant, and finally pure simulation with velocity damping. Shaded regions surrounding model lines are 95% confidence intervals.

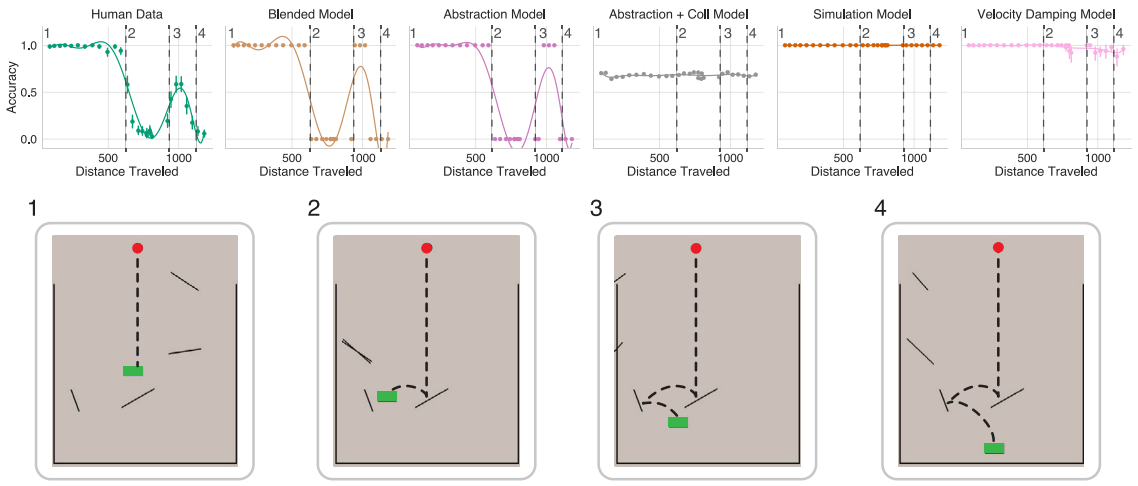

Fig. 6. Experiment 2 results. Top row: Accuracy (y-axis) as a function of the distance the ball traveled (x-axis), with each point showing mean accuracy across participants with 95% confidence intervals and spline fits. Spline fits are purely visual aids to demonstrate qualitative trends in the data. Participant accuracy ("Human Data") is a nonlinear function of the distance traveled by the ball across the ball's trajectory. Accuracy starts off near ceiling for the first group of stimuli in Section 1, in which distance is short and follows a straight path. Accuracy then takes a sharp dip for the second group of stimuli in Section 2, before rising slightly in Section 3, and dipping again in Section 4. These results are qualitatively captured by the blended model and a pure abstraction model, but not by the remaining models. Bottom row gives representative examples of the stimuli featured for each of the Sections 1–4 shown in the accuracy plots of the participants and the models (e.g., stimulus 1 is an example of the stimuli for Section 1, etc.). Dashed black line indicates the expected trajectory of the ball under pure simulation.

Participant accuracy was systematically affected by the distance between the ball and the goal (Fig. 6), following a non-monotonic pattern that mirrors the predictions of both the blended model and the abstraction model. The ability of our blended model to capture this trend suggests that people's performance is systematically affected by features of the task in similar ways to our blended model. An important note is that the model predictions for Experiment 2 were based on parameter values derived from fitting to data collected in Experiment 1; no model fitting was performed on the data from this experiment. The blended model captures this non-monotonic pattern as does the abstraction model, which leads us to believe most of the predictive power is coming from the abstraction model. In contrast, the simulation model predicts a flat accuracy curve that remains at ceiling for all distances. While both a pure abstraction and blended model capture the non-monotonic pattern observed from participants, model comparison using the Bayesian information criterion (BIC) favored the blended model (BIC = −41.60) over the other models of pure abstraction (BIC = −35.74), pure simulation (BIC = 81.18), pure simulation with velocity damping (76.94), and pure abstraction with a collision constant (61.35).

Taken together, the results support our hypothesis that people use a combined approach to reason about the physics in the scenes presented, and that a blended model which uses both abstraction and simulation can specifically account for the resulting behavior.

4. Discussion

How people reason about the physical world so easily is not so easily understood. Different computational accounts of intuitive physics have made very different assumptions and commitments about the underlying mental process. In this work, we broadly categorized the different accounts as either based on mental simulation (proposing that human physical reasoning is based on a noisy step-by-step process that recreates the objects and dynamics of the perceived scene), or not based on mental simulation (including models based on heuristics, logical rules, features, qualitative theories, and so on). Like previous work (e.g. Smith et al., 2023), we proposed a blended model, in which simulation gives way to cost-cutting abstractions when certain conditions are met. Unlike previous work, our model computes these conditions in real-time based on the state of the scenario, rather than once for the scenario as a whole, allowing us to model how people flexibly reason about physics in real time. To test our account, we ran novel experiments that presented people with physically controlled visual stimuli, and measured their accuracy and response time, while varying the scene configuration. We found that neither models of pure simulation with and without additional sources of uncertainty nor a pure abstraction model with and without additional coefficients for the number of collisions captured people's response time or accuracy, but that the blended model correctly accounts for both, supporting our hypothesis that people use both simulation and abstraction to perform physical reasoning. In the case of accuracy, we observe that a pure abstraction model captures participant accuracy as well, though has a lower log-likelihood compared to the blended model.

The blended model accounts for participant behavior in the limited settings we investigated, but we see our blended model as an instance of a more general computational framework, in which simulation and abstraction can be used in conjunction as instances of broad classes of inference methods for physical reasoning. The idea of using multiple modes of reasoning that balance efficiency and fidelity is a domain general phenomenon that extends beyond physical reasoning.

The framework we proposed here does not address the question of where abstractions originate. One possibility is that some modes of reasoning such as simulation are innate or early developing (Smith et al., 2019), and that various abstractions are then learned on top of this foundation. Our main claims in this work do not hinge on innateness or developmental claims, and the available evidence is not enough to distinguish whether it is simulation that is innate or early-developing, or various abstractions that are innate or early-developing. However, if simulation is the earlier of the two, both infants and adults could in principle learn abstractions via a process driven by compressing over previous simulation traces that led to satisfactory task performance. In other words, we can in principle explain abstraction learning as the process of actively attempting to reduce the cost of our innate or early-developing simulation engine, which can be done by caching the results of different simulations and clustering similar results into a category based on the features of the simulation, such as the existence of straight paths. Over time, these clusters could then be used as abstractions to reason about scenes or tasks that share their defining features.

This idea draws parallels with recent models of efficient concept learning, where the objective is to acquire a collection of maximally-compressed concepts that effectively address the tasks encountered

by the agent (Dechter, Malmaud, Adams, & Tenenbaum, 2013). In these models of "library learning", the emphasis is on minimizing the complexity of learned concepts while still solving the given tasks, striking a balance between model performance and computational cost (Ellis, Morales, Sablé-Meyer, Solar-Lezama, & Tenenbaum, 2018; Ellis et al., 2023). In the context of our framework, abstractions can be represented as subroutines integrated with simulation, as demonstrated in the blended model. These abstractions could then undergo iterative compression, ensuring a balance between computational efficiency and task performance. As we discuss below, we are actively working on models of how abstractions might be learned, and investigating how well these learned abstractions explain human physical reasoning in the context of our stimuli.

4.1. Limitations and future directions

People use more than one abstraction. The blended model considered here included only one kind of abstraction (path projection), but people likely use all kinds of additional abstractions. One example of a different useful abstraction for physical prediction is the notion of containment, which has been posited as one of the earliest abstractions children develop in order to reason about the physical world (Piaget & Inhelder, 1969). The more general computational framework proposed here treats abstractions, such as path projection or containment, as subroutines which can be used in place of simulation when certain task features are present, such as the visual presence of straight paths or containers. Current and future work is addressing this challenge by replacing our path projection abstraction with learned function approximators. Under certain environmental (learning) conditions, these function approximators may begin to form a repertoire of abstractions whose dynamics, such as if, when, and how they lead to cost-saving inferences or inaccurate predictions that mirror those found in people.

Threshold conditions do not work for all abstractions. Another limitation of the blended model used in this work is the specific algorithm used to trade off between simulation and abstraction, which is not general enough to handle abstractions other than path projection. Currently, the algorithm used to arbitrate between abstraction and simulation relies on the calculation of the cosine similarity between the resulting translation vectors between path projection and minimal simulation, and the parameters of this trade-off are fit to our specific implementation of path projection abstraction (see Methods). A more general method for arbitrating abstractions and simulation would be to learn a mapping between observable or latent (inferred) features of the current state of the scene and the choice of inference method (i.e., simulation or some particular abstraction). Such a method could be integrated with models of resource rationality: The features of the state of a scene would activate a potential set of relevant abstractions to use for a given task, while a decision model would select among those activated abstractions based on their expected cost savings against simulation. These abstractions could be partially evaluated to come to a final inference of the next state of the scene, conditioned on the cost of partially evaluating these abstractions in comparison with other available modes of reasoning, such as simulation. Such a decision model could be built from merging our current blended model with a Bayesian model of resource rationality, which would offer a parsimonious explanation of how people balance the resource costs of inference against expected task performance (Griffiths, Lieder, & Goodman, 2015; Ma & Woodford, 2020).

We are currently exploring a generalization of the algorithm used to arbitrate between simulation and abstraction that utilizes predicted uncertainty on the side of the abstraction. In this regime, when an abstraction is selected, it predicts a future state of the scene and assigns a computed uncertainty to that prediction. In the event that uncertainty is too high, control is given back to simulation for the time being. In concert with the learned abstractions we described above that we are investigating, we can determine the extent to which threshold conditions might serve as a suitable framework for arbitrating between modes of reasoning on-the-fly, as they are a computationally easy and cognitively plausible method of arbitration.

Richer measurements. The experiments presented here rely on response-time and accuracy measurements from online studies. While these data are valuable and support our framework of blended reasoning, we believe this support can be strengthened further with more high-resolution behavioral data such as eye tracking data. Eye tracking provides rich data on the mechanism and time course of cognitive processing (Beesley, Pearson, & Le Pelley, 2019), such as those processes that we investigate here in the context of physical reasoning. In visual stimulus tracking studies (e.g. being tasked with looking at target locations on a screen), rapid eye movements known as saccades have been shown to not only be reactive to visual stimuli when they appear, but also anticipate visual stimuli *before* they appear in the event previous visual stimuli are predictive of future visual stimuli (Polidora, Ratoosh, & Westheimer, 1957; Stark, Vossius, & Young, 1962). Overall, the literature on eye tracking suggests that a person's eye follows their mind's eye, looking toward where the person anticipates a target stimulus might appear, as well as reacting to where target stimuli have already appeared (Shelhamer & Joiner, 2003). Eye tracking has recently been used to investigate mental simulation in the context of counterfactual reasoning in intuitive physics tasks (Beller, Xu, Linderman, & Gerstenberg, 2022; Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017). We are currently conducting eye tracking studies using stimuli similar to those presented in Experiment 1 and Experiment 2 to test the predictions of the blended model.

5. Conclusion

Our ability to reason about the physical world with ease is both impressive and a mystery. Here, we suggest that one aspect of the perceived effortlessness of our physical reasoning is due to a form of clever laziness, where we adeptly decide where to simulate and where to abstract on the fly in order to make sense of the physical world. When something is potentially complicated, we might take our time to simulate what might happen next, but if it is familiar or simple, we take what happens next for granted and abstract, at a potential cost of accuracy. We developed a computational framework that trades off simulation and abstraction in a way that explains human response time and accuracy on a novel physical reasoning task in way neither models of simulation or models of abstraction can alone, suggesting that people trade off between similar modes of reasoning in a similar way.

The computational commitments of our general framework and the discussed extensions for future work are motivated by a fundamental question: How do people reason so rapidly and efficiently about everyday physics? Is it exclusively through mental simulation, or a different process altogether? Our theoretical proposal, substantiated by empirical findings, asserts that it is *both*—a blend of simulation and abstraction. While other proposals have also tried to provide a synthesis of approaches, they have tended to suggest a divide-and-conquer approach based on task: for two-dimensional static sketches of familiar scenes, it makes more sense to deploy static reasoning, whereas for realistic animations of novel and complex phenomena mental simulation might be more appropriate. By contrast, the proposed blended model dynamically combines simulation and abstraction for the same task, on the fly, offering a unique synthesis. Moreover, our proposed model serves as a specific instance within a broader framework, laying the groundwork for the development of more realistic and human-like models of intuitive physics.

CRediT authorship contribution statement

Felix A. Sosa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration,

Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Samuel J. Gershman:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization,. **Tomer D. Ullman:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization,.

Acknowledgments

We are grateful to Elizabeth Spelke for helpful discussions. This work was supported by the Center for Brains, Minds, and Machines (NSF STC award CCF-1231216), as well as the Harvard Hodgson Innovation Fund.

Data availability

The data and code is publicly available on GitHub.

References

Allen, K. R., Bakhtin, A., Smith, K., Tenenbaum, J. B., & van der Maaten, L. (2020). Ogre: An object-based generalization for reasoning environment. In *NeurIPS workshop on object representations for learning and reasoning*.

Allen, K. R., Smith, K. A., Bird, L.-A., Tenenbaum, J. B., Makin, T. R., & Cowie, D. (2021). Lifelong learning of cognitive strategies for physical problem-solving: the effect of embodied experience. bioRxiv.

Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In *Blackwell handbook of childhood cognitive development* (pp. 47–83). Wiley Online Library.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

Bass, I., Smith, K. A., Bonawitz, E., & Ullman, T. D. (2021). Partial mental simulation explains fallacies in physical reasoning. *Cognitive Neuropsychology*, *38*(7–8), 413–424.

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. (2019). Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Computational Biology*, *15*(7), Article e1007210.

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*(45), 18327–18332.

Bear, D., Wang, E., Mrowca, D., Binder, F. J., Tung, H.-Y., Pramod, R., et al. (2021). Physion: Evaluating physical prediction from vision in humans and machines. In *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 1)*.

Beesley, T., Pearson, D., & Le Pelley, M. (2019). Eye tracking as a tool for examining cognitive processes. In G. Foster (Ed.), *Biophysical measurement in experimental social science research* (pp. 1–30). Academic Press.

Beller, A., Xu, Y., Linderman, S., & Gerstenberg, T. (2022). Looking into the past: Eye-tracking mental simulation in physical inference. *vol. 44*, In *Proceedings of the annual meeting of the cognitive science society*.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, *105*, 9–38.

Bramley, N. R., & Ruggeri, A. (2022). Children's active physical learning is as effective and goal-targeted as adults'. *Developmental Psychology*, *58*(12), 2310.

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of object persistence: The origins of object permanence in infancy. *Child Development Perspectives*, *9*(1), 7–13.

Buschoff, L. M. S., Akata, E., Bethge, M., & Schulz, E. (2023). Visual cognition in multimodal large language models. arXiv.

Chang, M. B., Ullman, T., Torralba, A., & Tenenbaum, J. B. (2016). A compositional object-based approach to learning physical dynamics. arXiv preprint arXiv:1612.00341.

Cherian, A., Peng, K.-C., Lohit, S., Smith, K. A., & Tenenbaum, J. B. (2023). Are deep neural networks SMARTer than second graders? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10834–10844).

Conwell, C., & Alvarez, G. A. (2019). Leveling the field: Comparing the visual perception of stability across humans and machines. *Journal of Vision*, *19*(10), 26a.

Davis, E. (1990). *Representations of commonsense knowledge*. Morgan Kaufmann.

Dechter, E., Malmaud, J., Adams, R. P., & Tenenbaum, J. B. (2013). Bootstrap learning via modular concept discovery. In *Proceedings of the international joint conference on artificial intelligence*. AAAI Press/International Joint Conferences on Artificial Intelligence.

DiSessa, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science*, *6*(1), 37–75.

Ellis, K., Morales, L., Sablé-Meyer, M., Solar-Lezama, A., & Tenenbaum, J. (2018). Learning libraries of subroutines for neurally–guided Bayesian program induction. *Advances in Neural Information Processing Systems*, *31*.

Ellis, K., Wong, L., Nye, M., Sable-Meyer, M., Cary, L., Anaya Pozo, L., et al. (2023). DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philosophical Transactions of the Royal Society, Series A*, *381*(2251), Article 20220050.

Fischer, J. (2021). The building blocks of intuitive physics in the mind and brain. *Cognitive Neuropsychology*, *38*(7–8), 409–412.

Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences*, *113*(34), E5072–E5081.

Forbus, K. D. (1988). Qualitative physics: Past, present, and future. In *Exploring artificial intelligence* (pp. 239–296). Elsevier.

Fragkiadaki, K., Agrawal, P., Levine, S., & Malik, J. (2015). Learning visual predictive models of physics for playing billiards. arXiv preprint arXiv:1511.07404.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science*, *28*(12), 1731–1744.

Gerstenberg, T., & Stephan, S. (2021). A counterfactual simulation model of causation by omission. *Cognition*, *216*, Article 104842.

Gilden, D. L., & Proffitt, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 372.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

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? The amount of mental simulation tracks uncertainty in the outcome. In *CogSci*. Citeseer.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1084.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*(6), 280–285.

Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, *14*, 308–312.

Kaiser, M. K., Proffitt, D. R., & Anderson, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*(4), 795.

Kim, I.-K., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*(3), 339–362.

Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, *21*(10), 749–759.

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*.

Lerer, A., Gross, S., & Fergus, R. (2016). Learning physical intuition of block towers by example. In *International conference on machine learning* (pp. 430–438). PMLR.

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on simulation approaches in intuitive physics. *Cognitive Psychology*, *127*, Article 101396.

Ma, W. J., & Woodford, M. (2020). Multiple conceptions of resource rationality. *Behavioral and Brain Sciences*, *43*.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360.

McCloskey, M. (1983). Intuitive physics. *Scientific American*, *248*(4), 122–131.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, *210*(4474), 1139–1141.

McCloskey, M., & Kohl, D. (1983). Naive physics: the curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *9*(1), 146.

Mottaghi, R., Bagherinezhad, H., Rastegari, M., & Farhadi, A. (2016). Newtonian scene understanding: Unfolding the dynamics of objects in static images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3521–3529).

Nusseck, M., Lagarde, J., Bardy, B., Fleming, R., & Bülthoff, H. H. (2007). Perception and prediction of simple object interactions. In *Proceedings of the 4th symposium on applied perception in graphics and visualization* (pp. 27–34).

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. Basic Books.

Piloto, L. S., Weinstein, A., Battaglia, P., & Botvinick, M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature Human Behaviour*, *6*(9), 1257–1267.

Polidora, V., Ratoosh, P., & Westheimer, G. (1957). Precision of rhythmic responses of the oculomotor system. *Perceptual and Motor Skills*, *7*(3), 247–250.

Proffitt, D. R., Kaiser, M. K., & Whelan, S. M. (1990). Understanding wheel dynamics. *Cognitive Psychology*, *22*(3), 342–373.

Sanborn, A., Mansinghka, V., & Griffiths, T. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*, 411–437.

Schwettmann, S., Tenenbaum, J. B., & Kanwisher, N. (2019). Invariant representations of mass in the human brain. *Elife*, *8*, Article e46619.

Shelhamer, M., & Joiner, W. M. (2003). Saccades exhibit abrupt transition between reactive and predictive, predictive saccade sequences have long-term correlations. *Journal of Neurophysiology*, *90*(4), 2763–2769.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, *36*(3), 273–310.

Smith, K., Battaglia, P., & Tenenbaum, J. (2023). Integrating heuristic and simulation-based reasoning in intuitive physics. PsyArXiv.

Smith, K., Hamrick, J., Sanborn, A., Battaglia, P., Gerstenberg, T., Ullman, T., et al. (2024). Probabilistic models of physical reasoning. In T. Griffiths, N. Chater, J. Tenenbaum (Eds.), *Bayesian models of cognition: reverse engineering the mind*. MIT Press.

Smith, K., Mei, L., Yao, S., Wu, J., Spelke, E., Tenenbaum, J., et al. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. *Advances in Neural Information Processing Systems*, *32*.

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

Sosa, F. A., Ullman, T., Tenenbaum, J. B., Gershman, S. J., & Gerstenberg, T. (2021). Moral dynamics: Grounding moral judgment in intuitive physics and intuitive psychology. *Cognition*, *217*, Article 104890.

Spelke, E. S. (2022). *What babies know: Core knowledge and composition*: *vol. 1*, Oxford University Press.

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

Stark, L., Vossius, G., & Young, L. R. (1962). Predictive control of eye tracking movements. *IRE Transactions on Human Factors in Electronics*, *3*, 52–57.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–1059.

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*(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. *Advances in Neural Information Processing Systems*, *28*.

Zhang, R., Wu, J., Zhang, C., Freeman, W. T., & Tenenbaum, J. B. (2016). A comparative evaluation of approximate probabilistic simulation and deep neural networks as accounts of human physical scene understanding. arXiv preprint arXiv:1605.01138.