Humans predict liquid dynamics using probabilistic simulation

Christopher J. Bates, Ilker Yildirim, Joshua B. Tenenbaum, Peter W. Battaglia
{cjbates, ilkery, jbt, pbatt}@mit.edu
Department of Brain and Cognitive Sciences, MIT. Cambridge, MA

Abstract

Liquids can splash, squirt, gush, slosh, soak, drip, drain, trickle, pool, and be poured–complex behaviors that we can easily distinguish, imagine, describe, and, crucially, *predict*, despite tremendous diversity among different liquids' material and dynamical characteristics. This proficiency suggests the brain has a sophisticated cognitive mechanism for reasoning about liquids, yet to date there has been little effort to study this mechanism quantitatively or describe it computationally. Here we find evidence that people's reasoning about how liquids move is consistent with a computational cognitive model based on approximate probabilistic simulation. In a psychophysical experiment, participants predicted how different liquids would flow around solid obstacles, and their judgments agreed with those of a family of models in which volumes of liquid are represented as collections of interacting particles, within a dynamical fluid simulation. Our model explains people's accuracy, and their predictions' sensitivity to liquids of different viscosity. We also explored several models that did not involve simulation, and found they could not account for the experimental data as well. Our results are consistent with previous reports that people's physical understanding of solid objects is based on simulation, but extends this thesis to the more complex and unexplored domain of reasoning about liquids.

Introduction

From a glance at liquid flowing into a glass (Figure 1A), you can guess so much: it is pouring rapidly, likely from a spout or small opening; it is not viscous; little will likely splash out, although if the angle of the glass were lowered slightly, perhaps much more liquid would escape. These physical judgments are complex, and well beyond the capacity of current machine vision systems. How do humans, even young children (Figure 1B), reason about and manipulate liquids so effectively and so quickly? How sophisticated is people's implicit knowledge of liquid dynamics, and by what mechanisms is this knowledge applied to support people's broad competency?

A growing body of evidence supports the "Noisy Newton" hypothesis (Sanborn, 2014; Sanborn, Mansinghka, & Griffiths, 2013; Battaglia, Hamrick, & Tenenbaum, 2013; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012; K. Smith, Battaglia, & Vul, 2013; K. A. Smith & Vul, 2013; Hegarty, 2004), which suggests that humans have rich implicit knowledge of many aspects of everyday physics, which inform their predictions, inferences, and planning through a system of probabilistic reasoning. Battaglia et al. (2013) proposed a specific cognitive mechanism for physical scene understanding based on "approximate probabilistic simulation", in which objects' spatial geometry and physical attributes, as well as certain laws of mechanics, are represented approximately in a way that supports fast efficient probabilistic judgments over short time scales, by running a small number of simulations based on sampled estimates of the underlying

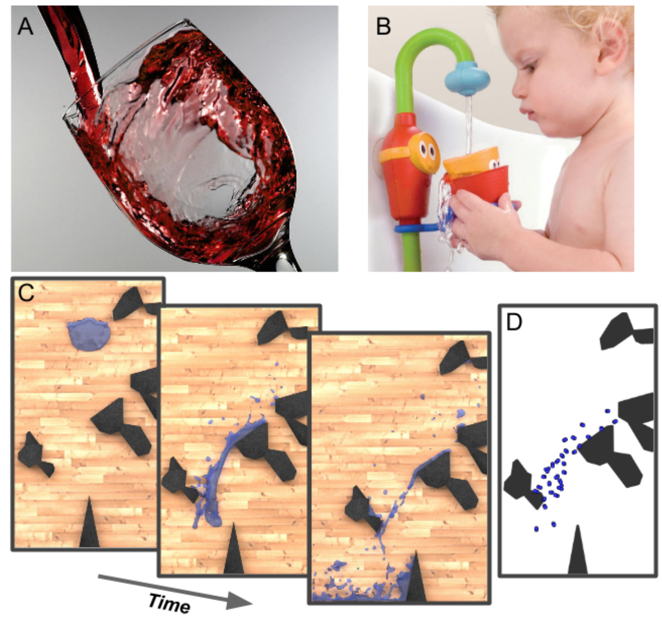

Figure 1: (*A*) Fluid dynamics are very complex, yet ubiquitous in everyday scenes. (*B*) Humans–even infants–can reason about and interact with liquids effectively. (*C*) Virtual rendering of liquid used for stimuli. (*D*) Our cognitive model hypothesizes that humans employ a particle-like representation of fluids.

world state. Their "intuitive physics engine" model explained people's physical predictions about stability and support relationships, and the motion of objects under gravity, across a wide range of rigid body scenes.

This recent literature may appear to contrast with earlier studies emphasizing conditions under which people's intuitions about physics do not seem consistent with Newtonian mechanics (McCloskey & Kohl, 1983). However, even in those earlier studies, a majority or plurality of participants often did give judgments consistent with Newtonian principles or approximate Newtonian simulations, and subsequent work suggested that people tend to show more accurate physical knowledge when tested in more natural perceptual or interactive sensorimotor tasks (K. Smith et al., 2013).

Here we present the first computational model extending the approximate probabilistic simulation approach to a family of dynamical scenes that are highly natural but have far more complex physical dynamics: We model people's perceptual intuitions about liquid dynamics, and present two experiments comparing this model and a number of alternatives with people's predictions about fluid flow under gravity in complex scenes. This complements previous research that examined people's physical intuitions about solid objects and

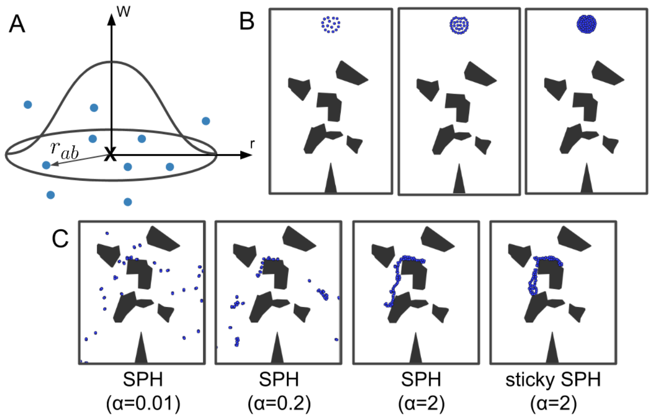

Figure 2: Our intuitive fluids engine is based on a smoothed particle hydrodynamics (SPH) method for simulating fluids. (*A*) How SPH approximates a fluid. For any location in the fluid, marked "X" on the diagram, particles in the local neighborhood are used to approximate the fluid's, density, pressure, and dynamics at that point. The bell-shaped envelope depicts the strength of each neighbor's influence on the approximation, which falls off with distance. (*B*) SPH simulations can be allocated more resources to achieve more precise approximations. In the second and third panels, more particles are allocated than in the first, which will result in more accurate and stable simulated fluid dynamics. (*C*) The rules by which particles interact can be varied to produce different qualitative fluids and materials. The first three panels show differences in splashing behavior as a function of viscosity. The fourth panel shows a non-Newtonian fluid that sticks to rigid surfaces (like honey).

mechanical devices (Hegarty, 2004; Battaglia et al., 2013), as well as recent work that has probed how people estimate viscosity from visual cues (Kawabe, Maruya, Fleming, & Nishida, 2014). Our cognitive model is based on a commonly used method for simulating fluids via collections of interacting particles, which we chose because it can handle a wide variety of materials, includes natural methods for adjusting the computational resource demands by varying the number of particles and complexity of their interactions, and is relatively simple compared to alternative methods. In graphics, particle systems are often used to simulate rigid and non-rigid objects, as well as liquids and gases, in isolation and in interaction. Humans have similar versatility, and thus particle-based simulation may offer a unified explanation of people's reasoning across many physical domains.

Our work asks specifically: Can particle-based simulation models account for people's predictions about how liquids move in complex scenes? How do people's uncertainty and computational resource limitations influence their judgments? Can the differences in people's predictions about different types of liquids be explained by corresponding differences in the interaction rules among simulated particles? We asked human participants to predict how a liquid would flow through and around complex arrangements of obstacles (Figure 1C) and compared their judgments to a family of simulation-based cognitive models (Figure 1D), as well as several non-simulation and non-physical alternatives, and found that the probabilistic simulation models provided a better account of people's responses.

Models

Simulation

SPH Our simulation-based models use a method from computational fluid dynamics called smoothed particle hydrodynamics (SPH) (Monaghan, 2005), which is used widely in physics, engineering, and graphics for approximating the dynamics of many types of compressible and incompressible fluids. The state of the fluid is represented by a set of particles at discrete time steps. Each particle carries information about a volume of fluid in a particular locality in space, including its position, velocity, density, pressure, and mass.

On each simulation time step, the particles' densities and pressures are computed, which are then used to update the accelerations, velocities and positions. A particle's density is calculated by interpolating its neighbors' densities, weighted by their distances, $\rho_i = \sum_{j=1}^{N_i} mW(r_{ij}, h)$, where ρ_i is the density at particle i's location, m is the mass of each particle, W is the kernel function, and r_{ij} is the distance between particles i and j (Figure 2A). The weighting is determined by W, which has a cutoff radius, h, beyond which particles have no influence. After computing particle i's density, its pressure is updated, followed by its viscous damping forces. Its acceleration is a linear combination of the pressure and viscous damping, the velocity update is proportional to the acceleration, and the position update is proportional to the velocity.

The precision of the liquid simulation can be adjusted by how many particles are used: with more particles, the simulated liquid's movement is more closely matched to that of a real liquid (Figure 2B). But increasing the number of particles also increases the computational cost of the simulation, thus effecting a trade-off between efficiency and accuracy.

Intuitive fluids engine Our cognitive model, the "intuitive fluids engine" (IFE), is analogous to Battaglia et al. (2013)'s intuitive physics engine, but is capable of reasoning about a fluid's dynamics. It posits that when the brain observes the initial conditions of a physical scene that contains liquids, it instantiates a corresponding particle-based simulation (which we capture using SPH) to predict future states of the scene. The model's inputs are the configuration of the scene, including the solid elements and the liquid's spatial state and material attributes, such as viscosity and stickiness. The number of particles that are instantiated can be varied, to adjust the computational resources demanded by the simulation.

Fluid dynamics can be complex, and we assume that the brain understands that a fluid's behavior can be influenced by many uncertain factors, such as imprecise and incomplete knowledge about the positions, shapes, and volumes of the solid and liquid elements of the scene, their underlying physical attributes, and the specific rules that govern how the elements transmit and respond to forces. We express the brain's uncertainty in our IFE model by consolidating these potential sources into a single, catch-all quantity, termed *physical uncertainty*. The model's physical uncertainty was implemented by adding a random offset (in a randomly chosen 2D direction), whose magnitude was sampled from an isotropic Gaussian distribution with mean 0 and standard deviation σ, to all particles' initial positions. The models ran a set of K independent simulations, with different random offsets, and aggregated the results to form a probability distribution over its prediction about what would happen. The model's judgments, J, took values between 0 and 1, where 0 represented all liquid flowing into the left basin, and 1 represented all liquid flowing into the right. That is, $J = n_{right}/N$, where N is the total number of particles, and n_{right} is the number that flowed into the right basin. We also computed predictions from a "ground truth" model which did not include uncertainty, and which used a single, deterministic simulation with a high number of particles and the correct viscosity, to predict the fluid behavior as accurately as possible.

We created different liquids whose viscosities correspond to water and oil, respectively, and also a higher viscosity liquid, with a viscosity typical of honey. This "Newtonian honey" behaves differently than real honey, because it does not stick to surfaces (since SPH particles collide with obstacle surfaces as inelastic spheres). Figure 2C reports viscosity values, α, for each fluid, which correspond to the viscosity constant in our simulations. The parameter α is not directly equal to either dynamic or kinematic viscosity, but serves to show the relative difference in viscosity between the different liquids. In order to model non-Newtonian, sticky liquids, we created a different kind of SPH liquid that creates the effect of stickiness by damping the normal and parallel components of velocity for particles that are in collision with obstacles (last panel of Figure 2C). Due to computational limitations, we only considered one set of parameter values, but the space of possible sticky SPH liquids could be further explored by varying the viscosity in combination with the parameters controlling stickiness.

Non-simulation

We contrasted the simulation-based models with two non-simulation alternatives: one that used shallow geometric heuristics, comparable to Gardin and Meltzer (1989), and another that used a convolutional neural network (Krizhevsky, Sutskever, & Hinton, 2012; Jia et al., 2014) trained on thousands of examples similar to our experimental test conditions. These alternative models are theoretically unappealing because they are highly specialized to the task, and thus should not be expected to work in even slightly different conditions, regardless of whether they depend on the same underlying physical laws. However, they represent popular competing perspectives on the general mechanisms of human perceptual cognition (Gigerenzer, Hertwig, & Pachur, 2011; McClelland, 2013), and offer other advantages, such as simplicity of computation and clear statements about the learning process.

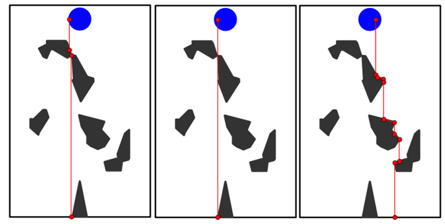

Figure 3: Heuristic model. Each panel depicts the path of a different particle.

Heuristic model The geometric heuristic model used deterministic rules that did not directly involve physics. It instantiates "particles" along the midline of the liquid's starting position and generates a path straight downward. When an obstacle is encountered, the path continues along the obstacle surface until it can go straight down again. The judgments were calculated by counting the number of "particles" that end up on each side of the divider, as in the simulation models (Figure 3).

Convolutional network We tested the possibility that people use purely visual cues by implementing a deep learning model. We replaced the top layer of a widely used convolutional network (Krizhevsky et al., 2012), pre-trained on a very large collection of images, with a sigmoid output layer, and performed backpropagation (using Jia et al., 2014) to learn a regression from images to labels in a supervised fashion. The dataset consisted of 10,000 randomly generated scenes, with 101 evenly spaced label values in the range $[0,1]$ that corresponded to the proportion of water that went to the right bin (as determined by a deterministic ground truth simulation with $N = 100$). The network was not shown any intermediate fluid positions. Judgments were calculated as the output of the model for each scene.

Methods

Participants All participants (N=65) were recruited from MIT Brain and Cognitive Sciences' human participants database (whose participant population is composed of roughly half MIT students and employees, and half local community members). All gave informed consent, were treated according to protocol approved by MIT's IRB, and were compensated $10/h for participation. All experimental sessions were one hour long, and each participant ran in one session in one experiment. All had normal or corrected-to-normal vision. Stimuli were presented on a liquid-crystal display, which participants free-viewed from a distance of 0.5-0.75 m. They indicated their responses by depressing a key on the keyboard or by adjusting the computer mouse and clicking to lock in their choice.

Stimuli and procedure In order to test people's ability to predict the behavior of a liquid, participants were presented with 120 virtual 3D scenes, 1.0 m x 1.50 m in size, that depicted a cylindrical volume of liquid positioned above a randomly generated obstacle course composed of fixed, solid objects and asked to predict what fraction of the liquid would flow into each basin below the obstacles under gravity. Participants either saw a low-viscosity, water-like liquid (Experiment 1) or high-viscosity, honey-like liquid (Experiment 2). The automatically generated, 2D scenes (see Random scene generation) were converted to 3D Blender scenes, by adding a small amount of 3D depth. The liquid was simulated using Blender's Lattice-Boltzmann liquid simulator, and video frames were rendered with Blender's Cycles ray-tracer, and composed into a video with a framerate of 30 Hz. All participants saw the same scene order, which was randomly shuffled. The 120 trials were divided into three blocks of 40, with a short break in between.

Both experiments included a practice and test phase. During the practice phase, participants received visual feedback on all trials (i.e., a video of the liquid simulation, also rendered with Blender Cycles) in order to familiarize them with the characteristics of the liquid and the response keys. On each trial (in both practice and test phases), participants viewed an image of the scene, with the liquid in its starting position. They were instructed to predict the proportion of liquid that would end up on each side of the divider (see the dark wedge at the bottom of the stimuli in Figure 4C), moving a virtual slider with the mouse left or right to indicate their response, and pressing ENTER to submit.

Random scene generation The obstacles in the test scenes were generated automatically by first dividing a plane into polygonal cells using 2D Voronoi tessellation, then selecting a random subset as solid obstacles. Coarse SPH simulations were run to filter out those scenes in which liquid particles remained trapped in obstacle concavities or had little interaction with the obstacles.

Intuitive fluids engine parameters The IFE models varied in their physical attributes of viscosity, α, and whether or not they used "stickiness", as well as the magnitude of their physical uncertainty, σ, and the number of particles, N.

In both the non-sticky and sticky IFE models, the number of particles used to represent the liquid was varied from 1 to 100. The stochastic noise, which implemented physical uncertainty, perturbed the initial position of each particle position in the simulated liquid by a random, additive 2D offset, sampled from an isotropic Gaussian distribution with mean 0 and standard deviation σ. The value of σ varied from 0 to R, where R was the radius of the initial disk of water. For each combination of viscosity, number of particles, and noise level, the model ran a set of 16 independent simulations, with different random offsets. The average prediction over the 16 samples was taken, such that each unique combination of model parameters made a single prediction.

In addition to our IFE models, we also compare participant data to two ground truth models, generated by setting $\sigma = 0$ and $N = 200$ (which was the highest value tested), and behaved as similarly as possible to the stimulus liquids in each experiment.

Results

Our analysis aimed to address three main questions: (1) Are people's judgments better accounted for by simulation, rather than non-simulation models? If so, are people's simulations appropriately sensitive to (2) physical attributes of the liquids, and (3) physical uncertainty about the scenes? To answer these questions, we calculated the means across participants' judgments for each scene, and estimated Pearson correlations between those mean judgments and each model's predictions as a measure of how well the model fit the human data. Figure 4A summarizes these correlations for the best-fitting instances of each model.

All simulation models (at all values of σ and N) outperformed the heuristic model, and most outperformed the ConvNet model, in both experiments. The heuristic model had correlations of $r = 0.25[0.22, 0.27]$ (the interval in brackets is a 95% CI, estimated by a bootstrap analysis with $10,000$ resamples (Efron & Tibshirani, 1994)) for Experiment 1 and $r = 0.22[0.18, 0.26]$ for Experiment 2 (see Figure 4A). The convolutional neural net was trained on the Experiment 1 ground truth, and its predictions were compared to subject responses from both experiments. Correlations with Experiment 1 and Experiment 2 subjects were $r = 0.53[0.50, 0.55]$ and $r = 0.30[0.26, 0.34]$, respectively. In both experiments, all IFE models show peak performance at $N \leq 100$. This could potentially reflect cognitive resource constraints in participants, which should be addressed more directly in future work.

In both experiments, participants were at least partially sensitive to the physical attributes of viscosity and stickiness of each liquid: the ground truth model whose viscosity and stickiness corresponded to Experiment 1 was a better fit to Experiment 1's participants' responses than the model whose viscosity and stickiness corresponded to Experiment 2, and Experiment 2's participants' responses were better fit by the ground truth model with Experiment 2's physical attributes (see Figure 4A). In Experiment 1, the correlations between the mean participant responses and the non-sticky and sticky ground truth models were $r = 0.76[0.74, 0.77]$ and $r = 0.43[0.40, 0.45]$, respectively. Experiment 2 showed the reverse effect: $r = 0.49[0.45, 0.52]$ for non-sticky and $r = 0.61[0.58, 0.64]$ for sticky.

In both experiments, including uncertainty in the simulation models also improved fits with people's judgments. In Experiment 1, best fitting values of N and σ had correlations of $r = 0.84[0.83, 0.86]$ and $r = 0.89[0.88, 0.90]$ for non-sticky and sticky IFE, respectively. Experiment 2 had best fit correlations of $r = 0.60[0.56, 0.63]$ and $r = 0.83[0.80, 0.85]$ for non-sticky and sticky, respectively. Figure 4B shows the correlations as a function of physical attributes and N.

An interesting and unexpected interaction between uncertainty and physical attributes emerged across our experi-

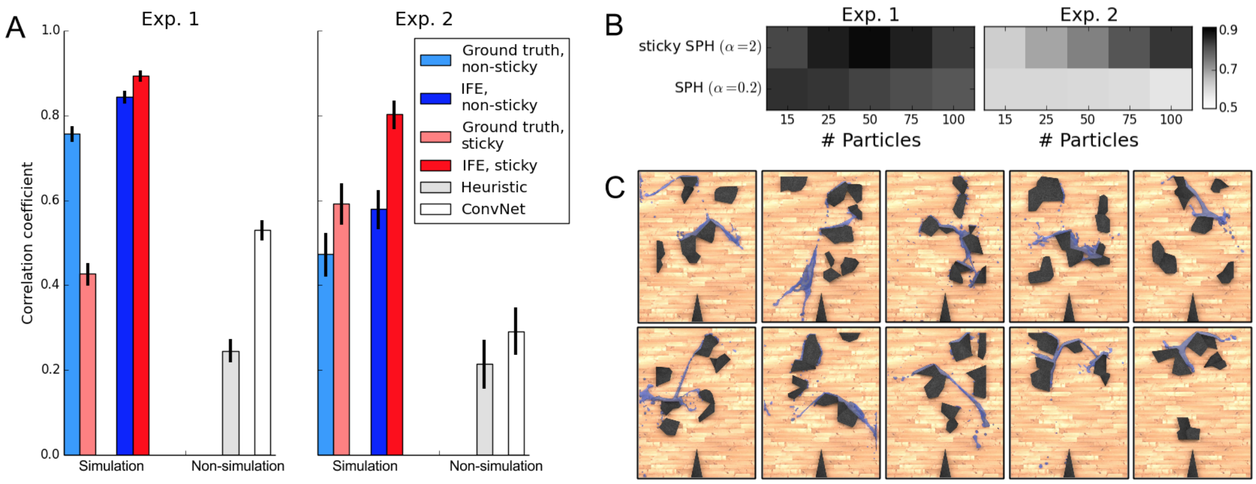

Figure 4: Stimuli and experimental results. (*A*) Mean participant data for both experiments is correlated with IFE and ground truth models (which could be sticky or non-sticky), and non-simulation models. The maximum correlation across N, σ, and α within each of the uncertain IFE models is plotted. (*B*) Heat maps of mean participant correlations with a selection of values across the uncertain IFE parameters (N, σ, α) in both experiments. Only the maximum correlation across all noise magnitudes, σ, is plotted for each combination of N and α. (*C*) Selected examples of experimental stimuli, shortly after gravity is turned on.

ments. In Experiment 2, the IFE model that appropriately accounted for both the liquid's physical attributes and physical uncertainty fit significantly better than any other model, but in Experiment 1, participants were slightly better fit by the sticky IFE model than the non-sticky IFE. The uncertainty added by the IFE models made their predictions less sensitive to physical attributes of the liquid than ground truth, making it more difficult to distinguish between sticky and non-sticky. This is evidenced by high agreement between sticky and non-sticky IFE models in both experiments 1 and 2 ($r = 0.91$ and $r = 0.85$, respectively) for best-fitting σ and N, as compared to correlations between the ground truth models ($r = 0.29$). The ground truth model results more strongly suggest that people capture the correct physical attributes in their judgments, but this issue should be explored further in future research.

In Experiment 1, split-half correlations reveal participants were highly consistent with each other ($r = 0.96[0.94, 0.97]$). Experiment 2 participants were less consistent with each other ($r = 0.80[0.72, 0.85]$) than Experiment 1, which might be attributable to less familiarity with the liquid, since Experiment 1 participants saw a liquid that behaved similarly to real water, but the liquid in Experiment 2 was not as similar to any liquids found in the real world.

Discussion

We found that models based on approximate probabilistic simulation provided a better quantitative account of people's judgments than alternative models which did not include simulation or did not take into account physical uncertainty. We also found that people were sensitive to the physical attributes of each liquid (stickiness and viscosity), and that this sensitivity can be captured, at least in part, by our models.

More generally, the particle-based simulations underlying our IFE models can potentially explain how people so flexibly and richly reason about the wide variety of liquids they encounter in everyday life. While there has been some skepticism (Marcus & Davis, 2013) toward Battaglia et al. (2013)'s hypothesis that many of people's everyday physical intuitions can be explained by approximate probabilistic simulation, our results suggest that this approach may extend beyond the simplest domains of rigid bodies, to more complex domains such as fluids.

Prominent AI researchers studying physical reasoning have often pursued qualitative or logical approaches (Forbus, 2011; Davis, 2008), arguing that solving problems similar to our experimental tasks "to a high degree of accuracy involves computational fluid dynamics" and "it is quite unlikely that we are capable of performing such a prodigious feat mentally" (Forbus & Gentner, 1997). Our empirical data suggest, however, that people may have a richer ability to make quantitative predictions about the behavior of liquids than previously assumed, and that these abilities may be explained by probabilistic simulations that approximate the true fluid dynamics only very coarsely, but effectively enough for everyday human purposes. Our work here should be seen as only a starting point for exploring this idea: We compare a first concrete simulation-based model, which makes testable predictions for people's quantitative judgments about some aspects of liquid dynamics, with concrete (but not necessarily optimal) implementations of feature-based and heuristic alternative approaches. We hope future research on simulation-based approaches, as well as the alternatives, will yield deeper understanding.

Future work should explore further how people predict sticky versus non-sticky liquids, and how the precision, temporal duration, and other structural and parametric features of the simulation are implemented. For example, how does the precision vary as a function of how far into the future one must mentally simulate? How closely do the attributes represented by the simulated particles correspond to the actual physical characteristics of a liquid? Physically accurate simulation models can account for people's judgments, but are there simpler particle-based models that can as well? We explored an alternative model that replaced the SPH rules with simpler rigid-body (e.g., marble-like) interactions among particles, and found preliminary evidence that it could fit people's judgments well in many of the situations we study here, although it clearly fails in others. Can our general approach extend to capture other classes of physical intuitions that go beyond rigid-body dynamics? Particle-based models can provide reasonable simulations for liquids and gases, as well as collections of solid elements (e.g., piles of sand) and composite materials (e.g., mashed potatoes or Play-Doh) whose dynamics share similarities with liquids, and we plan to test probabilistic versions of these models as accounts of people's predictions in these domains. What are the limits of simulation-based models – what kinds of non-rigid dynamics can people make coherent predictions about, and which might be better explained by alternative approaches such as qualitative reasoning? Recent work has presented evidence that internal forward models in the motor system are involved in predicting dynamics that cannot be reenacted by the body, such as the motion of "rolling ocean waves" (Schubotz, 2007). Perhaps establishing a connection between this work and our probabilistic simulation framework could lead to a deeper understanding of people's capabilities and limitations when reasoning about physics.

Another important question is: How are intuitive physics engines represented in the brain? A plausible candidate might be a recurrent neural network with dynamic, parallel, and distributed structure (Michalski, Memisevic, & Konda, 2014), but to date these networks have only been able to capture the very simplest kinds of rigid-body dynamics. A more general question for development, and computational cognitive psychologists is: Where does people's knowledge of liquids come from? Five month-olds can distinguish between solids and liquids in novel contexts after observing their distinct patterns of movement (Hespos, Ferry, & Rips, 2009), which suggests either a very data-efficient experience-based learning process or innate biases. Perhaps our simulation-based model's core components can be measured in young children, or its more complex features can be observed as they emerge and mature.

In summary, particle-based simulation is a powerful framework that may explain how people understand a wide variety of complex physical processes they encounter in everyday life. This work offers the first computational model of how people reason about liquids, and provides evidence that a particle-based simulation model can account for reasoning about different types of liquids. Our results provide further evidence for a probabilistic, simulation-based cognitive mechanism of physical reasoning, whose particle approximations can support solids and liquids, and perhaps other materials as well.

Acknowledgments

We thank the reviewers for their helpful comments. Research support was provided by the Center for Brains Minds and Machines (CBMM), funded by NSF STC award CCF-1231216 and by an ONR grant N00014-13-1-0333.

References

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

Davis, E. (2008). Pouring liquids: A study in commonsense physical reasoning. *Artificial Intelligence*, *172*, 1540–1578.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Forbus, K. D. (2011). Qualitative modeling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 374–391.

Forbus, K. D., & Gentner, D. (1997). Qualitative mental models: Simulations or memories. In *Proceedings of the eleventh international workshop on qualitative reasoning* (pp. 3–6).

Gardin, F., & Meltzer, B. (1989). Analogical representations of naive physics. *Artificial Intelligence*, *38*(2), 139–159.

Gerstenberg, T., Goodman, N., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy newtons: Unifying process and dependency accounts of causal attribution. In *In proceedings of the 34th.*

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford University Press, Inc.

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

Hespos, S. J., Ferry, A. L., & Rips, L. J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychological Science*, *20*(5), 603–611.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Kawabe, T., Maruya, K., Fleming, R. W., & Nishida, S. (2014). Seeing liquids from visual motion. *Vision research*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

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

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Frontiers in psychology*, *4*.

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.

Michalski, V., Memisevic, R., & Konda, K. (2014). Modeling deep temporal dependencies with recurrent grammar cells. In *Advances in neural information processing systems* (pp. 1925–1933).

Monaghan, J. J. (2005). Smoothed particle hydrodynamics. *Reports on progress in physics*, *68*(8), 1703.

Sanborn, A. N. (2014). Testing bayesian and heuristic predictions of mass judgments of colliding objects. *Frontiers in psychology*, *5*.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological review*, *120*(2), 411.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in cognitive sciences*, *11*(5), 211–218.

Smith, K., Battaglia, P., & Vul, E. (2013). Consistent physics underlying ballistic motion prediction. In *Proceedings of the 35th conference of the cognitive science society* (pp. 3426–3431).

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