Probabilistic Simulation Predicts Human Performance on Viscous Fluid-Pouring Problem

James Kubricht\textsuperscript{1}  
kubricht@ucla.edu

Song-Chun Zhu\textsuperscript{2,3}  
sczhu@stat.ucla.edu

Chenfanfu Jiang\textsuperscript{2}  
cffjiang@cs.ucla.edu

Demetri Terzopoulos\textsuperscript{3}  
dt@cs.ucla.edu

Yixin Zhu\textsuperscript{3}  
yixin.zhu@ucla.edu

Hongjing Lu\textsuperscript{1,2}  
hongjing@ucla.edu

\textsuperscript{1} Department of Psychology \quad \textsuperscript{2} Department of Statistics \quad \textsuperscript{3} Department of Computer Science

\textsuperscript{*} Equal Contributors University of California, Los Angeles

Abstract

The physical behavior of moving fluids is highly complex, yet people are able to interact with them in their everyday lives with relative ease. To investigate how humans achieve this remarkable ability, the present study extended the classical water-pouring problem (Schwartz & Black, 1999) to examine how humans take into consideration physical properties of fluids (e.g., viscosity) and perceptual variables (e.g., volume) in a reasoning task. We found that humans do not rely on simple qualitative heuristics to reason about fluid dynamics. Instead, they rely on the perceived viscosity and fluid volume to make quantitative judgments. Computational results from a probabilistic simulation model can account for human sensitivity to hidden attributes, such as viscosity, and their performance on the water-pouring task. In contrast, non-simulation models based on statistical learning fail to fit human performance. The results in the present paper provide converging evidence supporting mental simulation in physical reasoning, in addition to developing a set of experimental conditions that rectify the dissociation between explicit prediction and tacit judgment through the use of mental simulation strategies.

Keywords: Intuitive physics; mental simulation; animation; reasoning

Introduction

Imagine that you are preparing to pour pancake batter onto a griddle. To pour the correct amount, you must decide where to hold the container, at what angle, and for how long. We encounter similar situations frequently in our daily lives when interacting with viscous fluids ranging from water to honey, and with different volumes, contained in receptacles of various shapes and sizes.

In the majority of these situations, we are able to reason about fluid-related physical processes so as to implicitly predict how far a filled container can be tilted before the fluid inside begins to spill over its rim. However, people perform significantly worse when asked to make explicit reasoning judgments in similar situations (McAfee & Proffitt, 1991; Schwartz & Black, 1999). In the well-known Piagetian water level task (WLT; Rebelsky, 1964), participants receive instructions to draw the water level at indicated locations on the inside of tilted containers. Surprisingly, about 40\% of adults predict water levels that deviate from the horizontal by 5 degrees or more (e.g., McAfee & Proffitt, 1991). Schwartz and Black (1999) modified the WLT to include two containers, one wider than the other. The investigators asked participants to judge which container would need to be tilted farther before the water inside begins to pour out. Only 34\% of the participants correctly reported that the thinner container would need to be tilted farther than the wider one. However, when instructed to complete the task by closing their eyes and imagining the same situation, nearly all (95\% of) the participants rotated a thinner, imaginary container (or a real, empty one) farther. These findings suggest that people are able to reason successfully about relative pour angles by mentally simulating the tilting event. An apparent contrast in human performance between an explicit reasoning task and a simulated-doing task has also been found in people’s inferences about the trajectories of falling objects (Krist, Fieberg, & Wilkening, 1993; K. Smith, Battaglia, & Vul, 2013). Thus, empirical findings in the literature of physical reasoning suggest that people employ both explicit knowledge about physical rules and mental simulation when making inferences (Hegarty, 2004).

The noisy Newton framework for physical reasoning hypothesizes that inferences about dynamical systems can be generated by combining noisy perceptual inputs with the principles of classical (i.e., Newtonian) mechanics, given prior beliefs about represented variables (Bates, Yildirim, Tenenbaum, & Battaglia, 2015; Battaglia, Hamrick, & Tenenbaum, 2013; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2015; Sanborn, 2014; Sanborn, Mansinghka, & Griffiths, 2013; K. A. Smith & Vul, 2013). In this framework, the locations, motions and physical attributes of objects are sampled from noisy distributions and propagated forward in time using an intuitive physics engine. The resulting predictions are queried and averaged across simulations to determine the probability of the associated human judgment. Bates et al. (2015) extended the framework from physical scene understanding (e.g., Battaglia et al., 2013) to fluid dynamics using an intuitive fluid engine (IFE), where future fluid states are approximated by probabilistic simulation via a Smoothed Particle Hydrodynamics (SPH) method (Monaghan, 1992). The particle-based IFE model matched human judgments about final fluid states and provided a better quantitative fit than alternative models that did not employ simulation or account for physical uncertainty.

The present study aims to determine whether a particle-based IFE model coupled with noisy input variables can account for human judgments about the relative pour angle of two containers filled with fluids differing in their volume and...
viscosity. The experiment reported here was inspired by previous empirical findings in water-pouring tasks; e.g., participants tilt containers filled with imagined molasses farther than those with an equal volume of water, suggesting that people are able to take physical attributes such as viscosity into account when making fluid-related judgments (Schwartz & Black, 1999). Bates et al. (2015) also found that their participants’ judgments were sensitive to latent attributes of the fluid (e.g., stickiness and viscosity).

To quantify the extent to which humans employ their perceived viscosity of fluids in subsequent reasoning tasks, we utilized a recent development in graphical fluid simulation (Bridson, 2008; Jiang, Schroeder, Selle, Teran, & Stomakhin, 2015) to simulate the dynamic behavior of fluids in vivid animations. Previous work has shown that realistic animations can facilitate representation of dynamic physical situations (Tversky, Morrison, & Betrancourt, 2002). Furthermore, recent research on human visual recognition indicates that latent attributes of fluids (e.g., viscosity) are primarily perceived from visual motion cues (Kawabe, Maruya, Fleming, & Nishida, 2015), so displaying realistic fluid movement is needed to provide the input of key physical properties that enable mental fluid simulations. The present study, which uses a modification of Schwartz and Black’s (1999) water-pouring problem coupled with advanced techniques in computer graphics, aims to test the hypothesis that animated demonstrations of flow behavior facilitate inference of latent fluid attributes, which inform mental simulations and enhance performance in subsequent reasoning tasks.

Experiment

Participants A total of 152 participants were recruited from the Department of Psychology subject pool at the University of California, Los Angeles, and were compensated with course credit.

Materials and Procedure Prior to the reasoning task, participants viewed animated demonstrations of the movement of a moderately viscous fluid in two situations. The fluid used in the demonstrations was colored orange and was not observed in the judgment task. In the first demonstration, the fluid pours over two torus-shaped obstructions in a video looped three times and lasting for 11.5 seconds. The flow demonstration videos were presented to provide visual motion cues to inform participants’ perceived viscosity. Following the flow demonstration, participants viewed a video of a cylindrical container filled with the same orange fluid tilting at a constant angular rate ($\omega = 22\text{ deg s}^{-1}$; see Fig. 1) from the upright orientation of the container and moving towards the horizontal. The video was looped three times for a duration of 14.7 seconds.

Following the demonstration videos, two new fluids were introduced, one with low viscosity ($LV$; similar to water) and one with high viscosity ($HV$; similar to molasses). The $LV$ and $HV$ fluids were colored either red or green (counterbalanced across subjects). As shown in the top panel of Fig. 1, participants viewed a flow video of both the $HV$ and $LV$ fluids (looped three times) for a duration of 11.5 seconds before each judgment trial\(^1\). The two flow videos were presented side by side for comparison, and the relative position of each fluid was counterbalanced across subjects. The $LV$ and $HV$ fluids were selected to readily distinguish the two fluids based on their perceived viscosities, which were inferred from visual motion cues in the flow videos (Kawabe et al., 2015).

In the subsequent reasoning task, participants viewed a static image of two containers side by side filled with the $LV$ and $HV$ fluids (see bottom panel of Fig. 1). Participants were instructed to assume that each container was tilted simultaneously in the same way as observed earlier for the orange fluid in the tilting demonstration. They were informed that both containers were tilted at the same rate, and were provided with the quantity of fluid in each container. Participants were then asked to report “which container will need to be tilted with a larger angle before the fluid inside begins to pour out” and received no feedback following completion of each trial. The experiment manipulated the volume of the $LV$ and $HV$ fluids ($V_{LV}$ and $V_{HV}$, respectively) in each container across the values 20%, 40%, 60%, and 80%, representing the proportion of the container filled. Hence, the experiment consisted of 16 trials presented in a randomized order, including all possible volume pairs between the $LV$ and $HV$ fluids. The experiment lasted approximately 10 minutes.

Human Results

Fig. 2 depicts human performance for all 16 fluid volume combinations. To assess the relationship between $HV$ fluid

---

\(^1\)Tilting and pouring videos can be viewed online at the following URLs: [http://hydra.math.ucla.edu/~cfjiang/cogsci/][...]

Tilting videos: [...] = tilt_movies/
Flow videos: [...] = pour_fluid_to_box_varying_viscosity/.
volume and human judgments, we performed a logistic regression analysis and compared estimated coefficients for each participant across LV fluid volume conditions. We found that coefficients for each participant varied significantly between $V_{HV}$ conditions ($F(3, 149) = 113.89, p < .001$). In the highest LV fluid volume condition ($V_{LV} = 80\%$), $V_{HV}$ had a minor impact on participants’ responses ($\beta_{80} = .04, \sigma_{80} = .25$) relative to the lowest LV fluid volume condition ($V_{HV} = 20\%$: $\beta_{20} = .57, \sigma_{20} = .43$). These results demonstrate that participants’ responses were increasingly sensitive to HV fluid volume for the greater $V_{LV}$ conditions.

Next, we examined whether humans rely on heuristic-based reasoning to make their judgments. One candidate heuristic is that given two containers filled with different volumes of each fluid, the container with lesser fluid volume requires a greater rotation before beginning to pour. While participants adhered to this rule for trials where $V_{HV} < V_{LV}$, their judgments for each of the $V_{LV} < V_{HV}$ trials did not accord to the same heuristic. For example, in trials where $V_{HV} = 40\%, 60\%$, and $80\%$ and $V_{LV} = 20\%, 40\%, and 60\%$, respectively, the lesser-volume heuristic predicts LV fluid responses. However, HV response proportions for those trials were significantly greater than zero ($t(151) = 9.92, 8.86, 8.10, p < .001$). A second heuristic is to always choose the HV fluid as requiring a greater rotation since it moves slower than the LV fluid. The above three cases also disagreed with this heuristic since the rule would predict unity HV response proportions. In summary, response proportions in the specified trials reveals that participants adhered to latent fluid attributes (e.g., viscosity) and volume difference when making their tilt angle judgments (see Fig. 3).

## Models

### Fluid Simulation with Physical Dynamics

The simulation of incompressible flows through numerical evaluation of physical equations has become one of the most significant topics in computer graphics and mechanical engineering. The velocity field of simulated fluids is determined according to the constraints specified in the Navier-Stokes equations:

\[
\begin{align*}
\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + \frac{1}{\rho} \nabla p &= \mathbf{g} + \mu \nabla \cdot \nabla \mathbf{v}, \\
\n\n\end{align*}
\]

where $\mathbf{v}$ is the velocity, $\rho$ is the density, $p$ is the pressure, $\mathbf{g}$ is the gravitational acceleration (approximately $9.8 \text{ m s}^{-2}$), and $\mu$ is the viscosity. Equation (1) is called the momentum equation—it is simply Newton’s second law (i.e., $F = ma$) applied to fluid dynamics. Equation (2) is the incompressibility constraint on fluid velocity, where the null divergence of the velocity field corresponds to constant density within volumetric regions. Most liquids need to satisfy this constraint in order to maintain constant volume while moving.

To numerically solve these partial differential equations, we adopt the Fluid Implicit Particle/Affine Particle in Cell (FLIP/APIC) method (Zhu & Bridson, 2005; Jiang et al., 2015; Jiang, 2015), which has become standard in physics-based simulation calculations due to its accuracy, stability, and efficiency. Unlike Smoothed Particle Hydrodynamics (SPH), which purely relies on particles to discretize the computational domain, FLIP/APIC uses both particles and a background Eulerian grid. The Navier-Stokes equations are solved on the grid, allowing for accurate derivative calculations, well-defined free surface and solid boundary conditions, and accurate first-order approximation of physical reality. The FLIP/APIC method also circumvents common artifacts of SPH; e.g., underestimated density near free surfaces and weakly compressible artifacts. In fact, the requirement for incompressibility is crucial in the fluid-pouring problem studied in this paper. We choose not to use SPH because it does not guarantee a divergence-free velocity field unless additional computational components are included. FLIP/APIC, however, maintains the benefits of particle-based methods due to its hybrid particle/grid nature. The presence of particles in the current model serves to facilitate visualization and the tracking of material properties. Besides modeling fluid, the state-of-the-art physics-based simulation methods have provided realistic cues for modeling complex tool and tool-uses (Zhu, Zhao, & Zhu, 2015), generic containers (Liang, Zhao, Zhu, & Zhu, 2015) and soft human body (Zhu, Jiang, Zhao, Terzopoulos, & Zhu, 2016).

Realistic visualization of simulated fluids is particularly important for the flow demonstration animations displayed before each trial in our viscous fluid-pouring task. To provide vivid impressions of the motion of the LV and HV fluids, we adopt OpenVDB (Museth et al., 2013) and utilize the latest particle fluid surface methods developed in the field of computer graphics to reconstruct a smooth level set surface from the simulated fluid particles based on their locations. The favorable efficiency and precision of the FLIP/APIC method allows for effortless generation of ground truth re-
by the number of samples, the IFE outputs the distribution
of the prediction from all samples was generated. The trained
IFE model is directly applied to the test case to pre-
ordinate model with mean 0 and variance $\sigma_v$, whereas the noisy viscosity is generated by adding a fixed amount of Gaussian noise on a
logarithmic scale (Sanborn et al., 2013); $V_i = f^{-1}(f(V_T) + \epsilon)$, where $V_T$ is the ground-truth value, $f(V_T) = \log(\omega \cdot V_T + k)$, $f^{-1}$ is the inverse of $f$, and $\epsilon \sim$ Gaussian$(0, \sigma_v)$. The results reported herein chooses $\sigma_v = 0.1$, $\sigma_v = 0.1$, $k = 1.5$, $\omega = 1$. Each simulation required approximately 10 minutes to run on a modern single-core CPU. In order to run a large number of simulations during the sampling process, we discretized the viscosity and volume spaces into finite sets. Specifically, the viscosities are discretized into 8 different cases (0.1, 1, 10, 100, 200, 500, 1000, 2000) and the volumes into 21 different cases (0% to 100% with a step-size 5%). We pre-computed the simulation results for each discretized case, and stored the results in a database. During the sampling process, the sampled inputs are numerically rounded to the pre-computed discretized cases, where the results can be immedi-
ately retrieved from the database without re-computing.

Non-Simulation Models
To examine whether fluid simulation is necessary to account for how humans reason about fluid behavior, we compare the simulation model with two statistical learning methods—
the generalized linear model (GLM) (McCullagh, 1984) and the support vector machine (SVM) (Cortes & Vapnik, 1995). These models are purely data-driven and do not involve any explicit knowledge of physical laws or physical simulation. The selected features for these models include (i) the volumes of fluids in both containers, and (ii) the viscosity ratio between the LV and HV fluids.

To predict human judgment for the $i$'th trial $J_i$, both non-
simulation models were tested with the $i$'th trial, and trained with the remaining 15 trials $\{J_i; i = 1, 2, \cdots, 16, i \neq i^*\}$. The trained GLM model is directly applied to the test case to pre-
dict which container will need to be tilted to a larger angle before the fluid inside begins to pour out. Since the SVM is a discriminative classification method which can only predict discretized labels (i.e., +1 indicating selection of the left container and −1 indicating selection of the right container), we introduced perceptual noise (the same method for the IFE) to each test trial to model physical uncertainty. For each test trial, a set with 10,000 samples was generated. The trained
SVM model is then applied to predict the labels (+1 or −1) in each sample, which are then aggregated to form the probability distribution for each test trial.

Model Results

We first compared how well different computational models account for human performance for the 16 trials. Fig. 4 depicts results from the IFE, GLM, and SVM models with perceptual noise. Human judgments and model predictions were highly correlated ($r(14) = 0.9954, 0.9488$, and 0.9251, respectively). RMSD (root-mean-squared deviation) between human judgments and the models’ predictions are 0.0824, 0.1269, and 0.1418, respectively. Compared to the purely data-driven models (i.e., the GLM and SVM models), the simulation-based IFE model encodes material properties (e.g., viscosity) and perceptual features (e.g., volume) and provides better approximations to human judgments in the viscous fluid-pouring task. These results again support the role of simulation as a potential mental model that supports human inference in physical reasoning tasks.

Next, we examined whether the IFE model captures human performance on the three trials where the heuristic rule outlined earlier provides incorrect predictions; i.e., those trials where $V_{HV} = 40\%$, 60\%, and 80\% and $V_{LV} = 20\%$, 40\%, and 60\%, respectively. Here, the ground-truth model predicts that the HV fluid will require a greater angle of rotation before beginning to pour, while the heuristic rule discussed earlier suggests the opposite. HV response proportions for these trials were .39, .34, and .30, respectively, and the IFE model returned consistent predictions of .43, .37, and .23. Alternatively, the GLM model predicted response proportions of .49, .47, and .49, and the SVM model predicted response proportions of .49, .49, and .51. Thus, our IFE model captured human performance on the specified trials while competing data-driven methods returned predictions biased toward the ground-truth model and away from the lesser-volume heuristic.

Our results from the viscous fluid-pouring task agree with the findings of Bates et al. (2015) in that our probabilistic, simulation-based IFE model outperformed two non-simulation models (SVM and GLM). Our behavioral experiment also indicates that people naturally attend to latent attributes (e.g., viscosity) when reasoning about fluid states following observation of realistic flow demonstration animations. By extending our probabilistic IFE method to a reasoning task, we demonstrated that the noisy Newton framework can account for human performance in a fluid-related judgment task that traditionally precludes mental simulation strategies.

While simulation has been demonstrated as the default strategy in other mechanical reasoning tasks (Hegarty, 2004; Clement, 1994), the participants in Schwartz and Black’s (1999) experiments failed to spontaneously represent and simulate physical properties relevant to the water-pouring problem when making their judgments. It is important to note, however, that the present task differs from the traditional water-pouring task in several ways: (i) fluid viscosity and volume (rather than container diameter) varied across trials, (ii) a cup-tilting demonstration was displayed to visualize the rate of simulated rotation, and (iii) motion cues from flow demonstrations informed the perception of latent fluid attributes (e.g., viscosity). Comparison of our study to previous water-pouring studies suggests that the dissociation between explicit physical prediction and implicit judgment reported in the intuitive physics literature could be resolved in some situations by modifying task characteristics and instructions in ways that motivate simulated representation. While our viscous water-pouring problem indicates a set of simulation-inducing task characteristics, further research should aim to determine specific experimental factors that trigger simulation strategies. Specifically, can the conditions employed in the present task extend to classical rigid-body and fluid mechanics problems to resolve the discrepancy between people’s
explicit predictions and tacit judgments, and if so, what additional task characteristics serve to facilitate mental simulation?

Classical research in artificial intelligence has traditionally dismissed robust mental simulation as a strategy for physical reasoning due to its inherent complexity, often proposing simplified qualitative models instead (De Kleer & Brown, 1984). While the computational fluid simulations employed in the present study require extensive numerical evaluation to make predictions about future fluid states, humans appear to do so with precision and accuracy in comparatively small amounts of time. Furthermore, their performance in our reasoning task suggests representation of physical quantities that extend beyond qualitative process theory. While human results are generally consistent with physics-based simulation models coupled with noisy input variables, there remain discrepancies between model predictions and human judgments. Hence, future research should aim to address whether humans simulate fluid movements using mental models that accord to physical laws or emulate fluid dynamics by drawing on their everyday interactions with liquids across diverse physical situations (Grzeszczuk, Terzopoulos, & Hinton, 1998).

Acknowledgments

Support for the present study was provided by an NSF Graduate Research Fellowship, DoD CASIT grant W81XWH-15-1-0147, DARPA SIMPLEX grant N66001-15-C-4035 and ONR MURI grant N00014-16-1-2007.

References


