AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001439

An Approximate Representation of Objects Underlies Physical Reasoning

Yichen Li[1], YingQiao Wang[1], Tal Boger[2], Kevin A. Smith[3], Samuel J. Gershman[1], and Tomer D. Ullman[1]

[1] Department of Psychology, Harvard University

[2] Department of Psychology, Yale University

[3] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

People make fast and reasonable predictions about the physical behavior of everyday objects. To do so, people may use principled mental shortcuts, such as object simplification, similar to models developed by engineers for real-time physical simulations. We hypothesize that people use simplified object approximations for tracking and action (the *body* representation), as opposed to fine-grained forms for visual recognition (the *shape* representation). We used three classic psychophysical tasks (causality perception, time-to-collision, and change detection) in novel settings that dissociate body and shape. People's behavior across tasks indicates that they rely on coarse bodies for physical reasoning, which lies between convex hulls and fine-grained shapes. Our empirical and computational findings shed light on basic representations people use to understand everyday dynamics, and how these representations differ from those used for recognition.

Public Significance Statement

People interact with objects in the world in real-time, which requires mental shortcuts in physical reasoning. We propose that a key physical mental shortcut is the simplification of fine-grained shapes into coarser bodies. Such simplified bodies explain novel results across several psychophysical tasks, including judgments of causality, time-to-collision, and change detection.

Keywords: intuitive physics, object representation, visual tracking, resource rationality

Supplemental materials: https://doi.org/10.1037/xge0001439.supp

Color, shape, and texture help us tell apples from oranges. But when trying to reason about an apple hurled toward your face, you may not care that it is green, or shiny, or even that it is an apple. All that matters is how fast, heavy, and where the apple is. For all reasonable purposes, it might as well be an orange.

We suggest that people use at least two representations of objects: *shape* and *body*. The shape encodes features relevant to visual recognition, including fine-grain form and subtle textures. The body encodes properties relevant for tracking, collisions, and physical prediction. These properties include weight, position, and coarse form. The existence of something like a shape representation is not under dispute, though its exact nature has been greatly debated (Biederman, 1987; Marr, 1982; S. Ullman, 1989). The existence of a body representation is a less explored hypothesis by comparison, although across fields there are theories that people represent objects with limited fidelity.

In a parallel line of research in vision and attention, frameworks and empirical findings in multi-object tracking (MOT) show that when observing multiple moving objects, people have a hard time tracking perceptual features such as color, compared to indexing the location of objects (e.g., Saiki, 2002; Saiki & Holcombe, 2012; Suchow & Alvarez, 2011). For example, in Saiki (2002), participants had a hard time detecting the color-switch of partially occluded objects in the middle of a regular rotation of a pattern, suggesting a failure in color–shape conjunction during dynamic motion. Our current hypothesis builds on MOT

Yichen Li https://orcid.org/0000-0003-0435-5522

A preprint of this work has been made available on PsyArXiv (https://psyarxiv.com/vebu5/). Data, stimuli, analysis code, and preregistrations for all experiments are publicly available on Open Science Framework (https://osf.io/z9dpu/).

We have no conflicts of interest to disclose.

Yichen Li served as lead for data curation, formal analysis, investigation, methodology, visualization, and writing–original draft, review, and editing. YingQiao Wang served in a supporting role for writing–review and editing. Kevin A. Smith served in a supporting role for methodology and writing–original draft, review, and editing. Samuel J. Gershman served in a supporting role for methodology and writing–original draft, review, and editing. Tomer D. Ullman served as lead for funding acquisition and supervision and contributed equally to writing–review and editing. Yichen Li, Kevin A. Smith, Samuel J. Gershman, and Tomer D. Ullman contributed equally to conceptualization. YingQiao Wang and Tal Boger contributed equally to data curation, formal analysis, and visualization. YingQiao Wang, Tal Boger, and Tomer D. Ullman contributed equally to investigation, methodology, and writing–original draft. Kevin A. Smith and Samuel J. Gershman contributed equally to funding acquisition and supervision.

Correspondence concerning this article should be addressed to Yichen Li, Department of Psychology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, United States. Email: yichenli@fas.harvard.edu

work and emphasizes a dissociation between the representation for recognition versus for physical tracking. Our framework suggests why and how form representations should be limited in reasoning about the physical behavior of objects, and speaks to the ongoing debate in MOT about which features are useful in tracking (Li et al., 2019).

The distinction between shape and body is motivated by engineering principles, and by converging evidence from cognition, developmental studies, and neuroscience. We next detail the relevance and convergence of these lines of research.

Engineers who design real-time physical simulators and game engines (Gregory, 2018) often use principled approximations for greater speed and efficiency. Pressures of speed and efficiency may have led cognitive architectures to develop and adopt approximations similar to those used in such real-time simulators (T. D. Ullman et al., 2017). A central approximation used by real-time simulators is to approximate bodies for physical interactions such as collision detection, separate from the fine-grain forms used for rendering objects (Figure 1). Body approximations can be refined meshes, but those are more computationally expensive, and approximations such as bounding boxes or convex hulls often produce reasonable results while reducing computational costs.

Previous work has proposed that noisy mental game engines underlie much of human intuitive physical reasoning (Battaglia et al., 2013; Hamrick et al., 2016; Sanborn et al., 2013; Smith & Vul, 2013; T. D. Ullman et al., 2018). This proposal has been challenged, with some researchers taking the mental game engine proposal to mean that intuitive physical reasoning should be a veridical simulation of reality. And, since physical reasoning deviates from reality, mental game engines cannot explain human behavior (Ludwin-Peery et al., 2020; Marcus & Davis, 2013). However, it is likely that mental physical simulations (if they exist) use approximations in a resource-rational way, in line with resource-rational cognition (Lieder & Griffiths, 2020; Smith et al., 2018).

Studies in cognitive development show that in many cases infants below 12 months do not use fine-grained form information to track objects (Xu, 2005; Xu & Carey, 1996), with follow-up work showing that such effects also exist in 18 months old under memory load (Zosh & Feigenson, 2012). These findings are often taken to suggest that young infants do not use "kind" information to track objects, though infants are certainly able to individuate objects based on shape, pattern, and color (Wilcox et al., 2010). We interpret the previous work as showing that young infants may often be relying on rough approximations for tracking. Such rough approximations are also central to recent artificial intelligence models that pass benchmarks designed to test models of core infant physics (Smith et al., 2019). Other developmental work on object individuation has also led to proposed modules for dealing with bodies (Leslie , 1994), and more recently to a distinction between features and objects in infant visual memory (Kibbe, 2015; Kibbe & Leslie, 2019), which may map onto our body-shape distinction. If such a distinction exists early in development, it likely persists into adulthood.

In neuroscience, a traditional split divides cortical visual processing in primates into ventral ("what") and dorsal ("where" or "how") streams (Goodale & Milner, 1992; Schneider, 1969). While the dorsal stream is often taken to encode spatial information about objects, more recent studies have refined this account (Kravitz et al., 2011), suggesting that the dorsal stream also encodes information that guides action. Research with nonhuman primates further suggests that the dorsal stream encodes action-relevant details of the form, orientation, and size of objects (Murata et al., 2000; Sereno & Maunsell, 1998). Moreover, recent functional magnetic resonance imaging evidence shows that a network of dorsal regions engage in intuitive physical inference tasks (Fischer et al., 2016) and represent physical variables of objects (Schwettmann et al., 2019). Such an action-relevant form found in the dorsal stream may map onto a body approximation, encoding variables relevant to intuitive physics.

Taken together, findings from cognitive science, cognitive development, and neuroscience align with engineering principles to suggest that body approximations may be cognitively useful in physical reasoning and may be separate from fine-grain forms for visual recognition. In order to examine the existence of this body-shape distinction in people, we created three distinct psychophysical tasks based on classic experiments (Figure 3A): perception of causality in launching (Experiment 1), time-to-collision (TTC) prediction (Experiment 2), and change detection (Experiment 3). While different in their design, these experiments used similar stimuli, and shared an underlying logic—body and shape are dissociated by having concave and convex conditions (see an illustration of concave versus convex in Figure 2).

In Experiment 1, participants rated perceptions of causality when seeing an agent (the first-moving object) colliding with a

Figure 1

Game Engines Use Different Representations for Rendering Versus Physical Interactions

Note. (A) "Shape" is used for rendering an object onto the screen. (B)"Body" is an approximation used to determine collisions, apply forces, and track objects. Example approximations are shown in increasing coarseness from left to right: mesh collider, convex hull, cylinder collider, and bounding box. See the online article for the color version of this figure.

Figure 2

Illustration of Concave Collision (Left) and Convex Collision (Right)

Note. Concavity is an inward curvature, while convexity curves outwards. The collision point on the left is in the concavity, and the collision point on the right is in the convexity. We created concave and convex collision conditions in our experiments to dissociate the body from shape. See the online article for the color version of this figure.

patient (the second object to move, and see Figure 3A, left). We expected that coarse bodies will result in a smaller perceived collision distance than the ground truth, but only for concave collisions (Figure 3B, left). In Experiment 2, participants were asked to press a spacebar to indicate when an agent and a (disappearing) patient collided (Figure 3A, middle). We expected that coarse bodies will result in smaller TTC in concave collisions than in convex collisions (Figure 3B, middle). In Experiment 3, an object either changed or remained the same when passing behind an occluder, with the changes happening within or outside a coarse body approximation (Figure 3A, right). We expected that concave changes within the body are more difficult to detect than changes outside the filled concavity, and changes outside the convex hull (Figure 3B, right).

To summarize our overall hypothesis: we propose that people use rough body approximations for physical tasks, and derive from this proposal distinct differences between concave and convex stimuli. Our more specific hypotheses are: (a) in Experiment 1, people will perceive concave-sided collisions as more causal than convex-sided collisions, (b) in Experiment 2, people will expect concave collisions to happen earlier than convex collisions, and (c) in Experiment 3, people will be less likely to detect a change within the body approximation than an equally sized change outside of the approximation. If people use fine-grain shapes for physical tracking, there should be no observable difference between concave and convex conditions across these experiments.

In addition to our broad empirical predictions, we considered a list of approximation models that allowed us to quantitatively examine a space of possible body approximations. Among all models we tested (α-shape, Gaussian noise, buffer, Ramer–Douglas–Peucker, bounding box, convex hull, and center-of-mass), we focused on the α-shape model (Edelsbrunner et al., 1983) that has one parameter α controlling the coarseness of the approximation. By changing α, we examined different body approximations, ranging from fine forms to convex hulls (Figure 4A). We treat the α-shape model as an exploratory model that coarsely differentiates between approximations closer to fine-grain forms, convex hulls, and intermediate representations. We emphasize that the specific α-shape model used here is not a process-level account of the approximation people use, as mathematicians and engineers have come up with many ways of simplifying and compressing shape information (Gregory, 2018; Luebke et al., 2003). Instead, we use α-shape as a stand-in for a class of models that instantiate our theory that people's shape approximations are simpler than the fine-grain form, and a trend toward coarse convex representations. We discuss other possible approximations later, in the context of our findings.

Transparency and Openness

All hypotheses, analyses, and procedures for the experiments were preregistered. Data, stimuli, analysis code, and registrations for all experiments are publicly available on the project's Open Science Framework page: https://osf.io/z9dpu/.

Experiment 1: Causality

Our first test of body approximations used physical causality judgments, based on the classic Michottean launching task (Kominsky et al., 2017; Michotte, 1963) with varying collision distances, but with concave and convex shapes.

Participants observed videos of one shape (the agent) moving toward a stationary target (the patient). At the moment when the agent was adjacent to the patient, the agent stopped, and the patient started to move away from the agent. Michotte's original studies found that people's causality judgments decreased as the spatial distance at collision time increased. We adopted the launching task paradigm and modified the shapes to create a dissociation between body and shape. We expected that people's causality judgments will track the spatial distance between approximate bodies, rather than shapes (see Figure 3A and B, left).

We aimed to test (a) an overall predicted effect of concavity versus convexity, such that concave collisions would be seen as more causal than convex collisions for the same horizontal collision gap, and (b) a specific α-shape model, exploring the α parameter that best explained people's body approximation, if it exists.

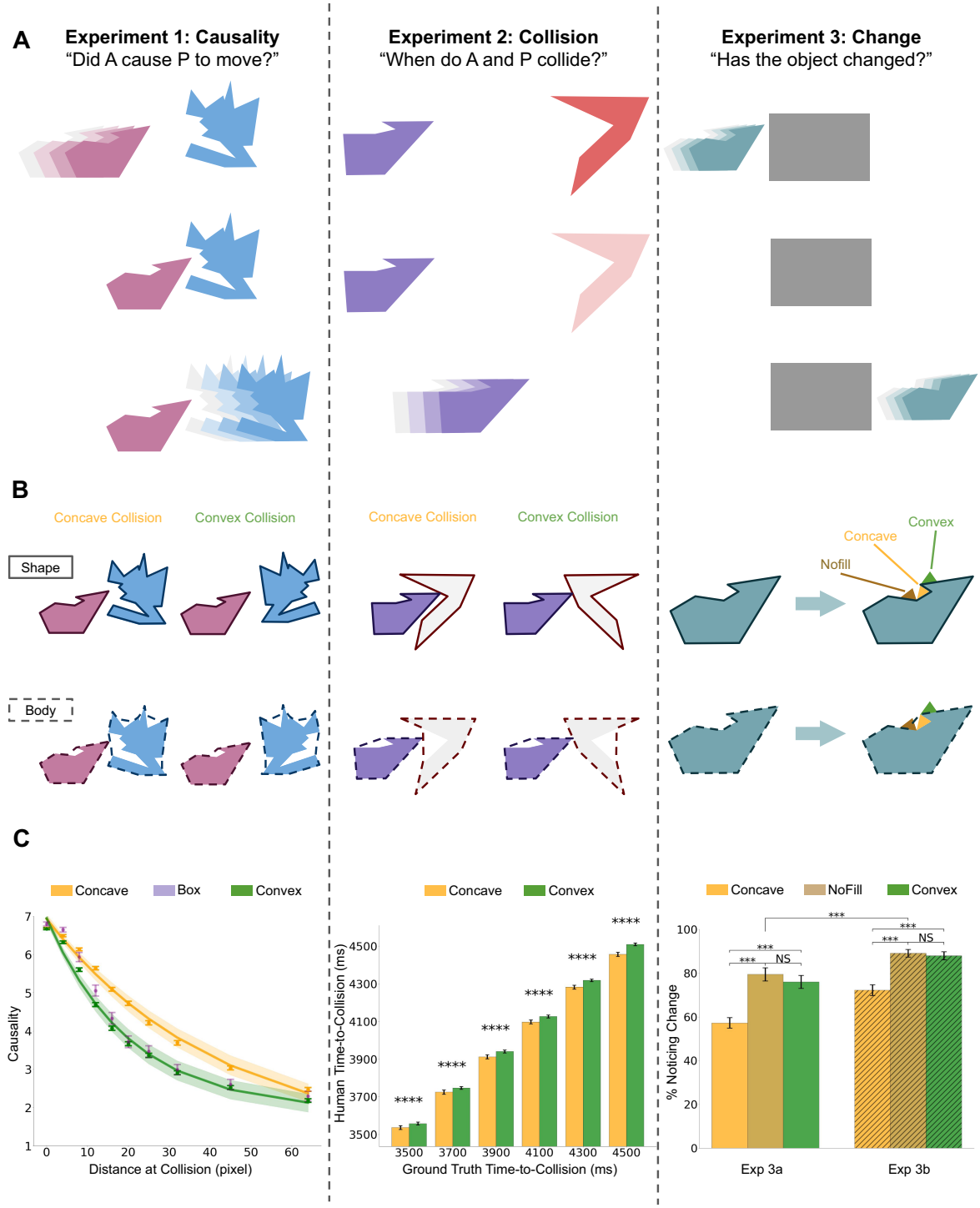
Experiment 1—including design, hypotheses, analyses, and exclusion criteria—was preregistered at https://osf.io/f3kwd. Screenshots of the stimuli and other details can be found in the online supplemental materials.

Participants

In this experiment as well as all other experiments, sample sizes were determined by power analysis (99% power, significance level = 0.05) based on pilot data, with the exception that Experiment 3b used the same number of participants to match Experiment 3a. Across all experiments, a total of 670 participants were recruited online, through Amazon Mechanical Turk (Experiment 1; Crump et al., 2013) and Prolific (Experiments 2 and 3; Peer et al., 2017). All participants were US-based, and all experiments were approved by the Harvard University Area Institutional Review Board (protocol no. 19-1861).

In Experiment 1, 330 participants were recruited, with a link directing to a survey page on Qualtrics. Participants were compensated 4.5 USD for their time, at a rate of about 10 USD per hour. In the optional demographic questionnaire, we provided a free-text response box to collect participants' gender (female = 127, male = 199, other = 4). The median participant age was 37. The median completion time of the study was 26.2 min. We excluded from analysis participants who did not pass one or more catch/comprehension questions ("What is the color of the sky?", "What was your task in this study?", "Which entity was the agent in this study?", and "What

Figure 3
Experimental Design and Results From Experiments 1–3

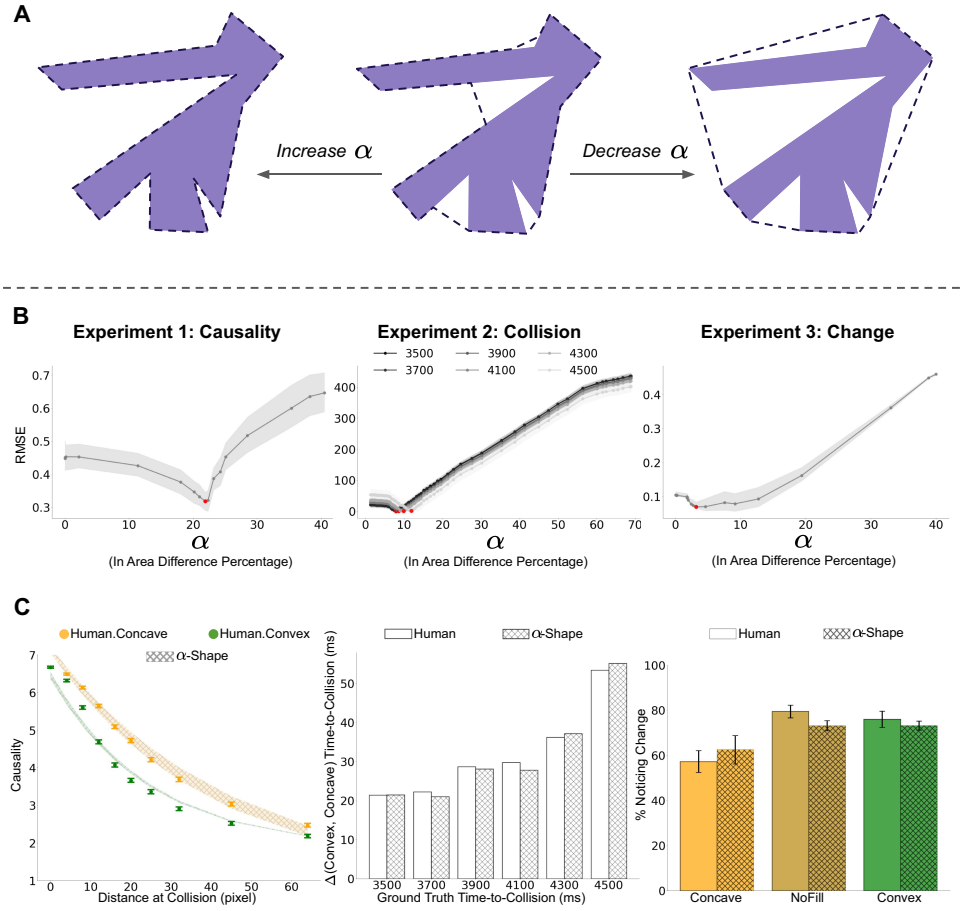


Note. (A) Diagrams of stimuli. (B) Differences in interactions between shapes and bodies. (C) Average participant responses (with standard error of measurements [SEMs] and confidence intervals of curve fit). See the online article for the color version of this figure.

was the number corresponding to the highest level of agreement in this study? (1–7)"), as well as participants who gave constant ratings across the experiment. This left 147 participants for analysis. The exclusion rate was typical of Amazon Mechanical Turk experiments at the time of running these studies. For a discussion of the reliability of this subject pool and others, see Peer et al. (2017).

Figure 4

α-Shape Model Overview, and Modeling Results for Experiments 1–3



Note. (A) The α-shape model (dotted lines) fits different approximations to a given shape (solid purple). (B) Best-fit α-shape approximations for the different experiments. As differences in the α are nonlinear in a cognitively irrelevant way, the *x*-axis shows the α parameter measured in average area-difference percentage between the original shape and the approximated body. The *y*-axis shows root mean squared error (RMSE) and 95% confidence intervals for predicting participant responses, with the best α-shape models indicated by a red dot. (C) Prediction from the single best-performing α-shape model in each experiment, compared with participant data. See the online article for the color version of this figure.

Design and Procedure

In all experiments, the stimuli were based on a set of eight irregular shapes, taken from a classic study on mental rotation (Cooper, 1975) for having a low verbal association.

As in the standard Michottean launching task, participants saw an object (the agent) moving toward a stationary object (the patient) at a constant speed. At a predetermined point, the agent stopped moving, and the patient began moving away from the agent, at the same speed and direction as the agent did. Participants saw 180 such videos, each 5 s long (see below for a breakdown of the different videos). At the end of each video, participants used a 7-point Likert scale to report their agreement with the statement "The agent caused the patient to move" (cf. Kominsky et al., 2017). The level of agreement was the dependent variable.

As illustrated in Figure 3B (left), the agent could collide with the concave or convex side of the patient. Each patient always had both a concave and convex side. On the concave side was a divot that would contain the point of contact should the two shapes collide. On the convex side, the collision point would be on the convex hull of the shape. We used mirror images of the eight irregular shapes, to create a concave–convex pair for each irregular patient shape (including a slight rotation to align with the agent's point of contact). Concave trials had the agent moving toward the concave side of the patient, and the point of contact was within the concave divot of the patient. Convex trials had the agent moving toward the convex side of the patient, and the point of contact was on the convexity of the patient.

We defined the distance at collision as the horizontal distance between the two objects at collision time, that is, the distance along the width of the screen between the hypothetical contact points on the two objects if there was no spatial gap. The distance at collision (i.e., the horizontal distance at the time of collision) between the agent and the patient was one of the following

values: 0, 4, 8, 12, 16, 20, 25, 32, 45, and 64 pixels. The longest distance, 64 pixels, corresponded to about half the length of the agent.

For trials with irregular shapes, the agent was always the same irregular shape, while the patient shape varied across trials. In addition, we created 10 warm-up trials (one trial for each distance at collision) and 10 control trials (one trial for each distance at collision), in which both the agent and the patient were regular boxes of the same size. The purpose of the control trials was to replicate the base launching effect, as well as validate participants' interpretation of the stimuli. We randomized the direction of motion (left-to-right or right-to-left) and the agent/patient colors in every video. This allowed us to compare participants' causality judgments between concave and convex side-of-hit, fixing a spatial distance and the overall visual complexity. In total, there were 180 videos (10 regular box warm-up trials + 10 distances × 8 irregular patient shapes × 2 side-of-hit conditions + 10 regular box control trials). Participants first saw a block of 10 warm-up collisions in randomized order to establish baselines and exclusion criteria, followed by a block of randomized presentations of the other 170 videos (including control trials).

Results

We predicted that under the same distance at collision, causality judgments in the concave condition would be higher than in the convex condition. The reasoning is as follows: body approximation is coarse and fills in parts or all of a concavity. If people used an approximate body to track the agent and the patient, their perceived distance at a collision in the concave condition should be smaller than the actual distance at collision. In the convex condition, the subjective distance at collision should be close to the ground-truth distance at collision, as the body approximation does not change the convex side of the patient compared to the original shape.

To quantitatively test this, we fit two separate exponential decay curves to participant causality judgment ratings, one for concave trials and one for convex trials. Denoting participant causality ratings as (C) and the horizontal distance at collision (D), the formulation was:

$$C = a \cdot e^{-D/b} + c,$$

where a, b, and c were parameters that control the displacement, curvature, and intercept of the curves. We were interested specifically in the curvature for concave and convex collisions, meaning the difference between b_{concave} and b_{convex}.

We used 1,000 bootstraps of participant responses. In every bootstrap, we sampled the full sample size with replacement, averaged responses across sampled participants for every ground-truth distance, and fit curves over the averaged data. In total, we obtained 1,000 bootstrapped parameter estimates for a curve fit. A paired t test was used to compare 1,000 bootstrapped b_{concave} estimates and 1,000 bootstrapped b_{convex} estimates. Then, we repeated this comparison between b_{concave} estimates and b_{convex} estimates 10,000 times. In every repeat, we sampled 1,000 b_{concave} estimates and b_{convex} estimates with replacement and tested their distribution by a paired t test ($\alpha = 0.05$). We calculated the percentage of repeats with significant t test results.

As shown in Figure 3C (left), causality ratings for all collision types decreased exponentially with the collision distance, replicating Michotte's finding that perceived causality was a decreasing function of distance at collision. Importantly, for the same ground-truth distance at collision, concave collisions were perceived as more causal than convex collisions (the concave decay curve is above the convex curve except at the endpoints). Causality ratings in the control trials (regular box colliding with regular box) mostly overlapped with ratings in convex trials.

The best parameters fit to the exponential decay curves were (with 95% confidence interval [CI] in parentheses): $a_{\text{concave}} = 5.8$ [5.3, 6.0], $a_{\text{convex}} = 5.04$ [4.7, 5.3]; $b_{\text{concave}} = 41.6$ [35.6, 47.1], $b_{\text{convex}} = 20.6$ [18.3, 22.9]; $c_{\text{concave}} = 1.1$ [1.0, 1.5], and $c_{\text{convex}} = 1.9$ [1.7, 2.1]. For the curvature parameter of interest, $b_{\text{concave}} > b_{\text{convex}}$, by a paired t test on the bootstrapped b_{concave} and b_{convex}, $T(999) = 3.1 \times 10^2$, $p < .001$; 100% of 10,000 bootstrapped comparisons showed that bootstrapped $b_{\text{concave}} > b_{\text{convex}}$.

Taken together, the results indicate that (a) participants perceived the distance at collision, and accordingly reported decreased causality ratings when distance at collision increased; (b) participants' interpretation of the distance at collision was similar to Michotte's study; (c) participants perceived the distance at collision for irregular convex patients in a similar way to that for regular boxes, and (iv) under the same ground-truth distance at collision, the effective distance at collision that participants *perceived* for the concave collision was shorter than the perceived distance at collision for the convex collision. In addition, we ruled out several other measures of distance at collision, such as Euclidean distance and center-of-mass distance. Details on these other measures can be found in the online supplemental materials.

α-Shape Model Analysis

The previous results showing that participants gave higher causality ratings for concave collisions than convex collisions were predicted by the hypothesis of approximate bodies, but were not based on a specific body approximation model. In all of our experiments, we considered an α-shape approximation algorithm (Edelsbrunner et al., 1983) to examine in more detail the approximate body representation people may be using. We also considered several alternative models, including Gaussian noise, buffer, Ramer–Douglas–Peucker, bounding box, convex hull, and center-of-mass (see the online supplemental materials for an analysis and description of these models). We decided to use the α-shape model for the rest of the analyses, both for principled reasons (this model is solving the same challenge we take the cognitive system to be solving, simplifying shapes given limited resources), and because it was the best model among the ones we considered in terms of fitting the human data. We stress that we do not take this model to be a process-level account of the simplification algorithms people use, and that further work is needed to establish the simplification algorithms (assuming they exist), especially for real-world three-dimensional objects. However, this model is useful for quantitatively assessing questions like "Are people using a simplification?" and "Is this simplification somewhere between a convex hull and a fine-grain shape?"

The α-shape algorithm produces an approximate polygon of a given shape, with one parameter α controlling the coarseness of the resulting approximation, ranging from convex hulls to fine-grain forms (Figure 4A). By varying the α value, we obtained a range of possible approximate representations for both the agent and the

patient. So, each setting of the α value produced different predictions of the effective distance at collision between the agent and the patient. Using the effective distance at collision as input, we fit a single exponential decay curve (same formulation as above, but replacing *D* with the effective distance at collision) with least squares to predict participant causality ratings for both concave *and* convex trials. That is, the *a*, *b*, and *c* parameters were constrained to be the same between concave and convex conditions.

The expectation was that a reasonable α-shape model should produce body approximations that can account for causality ratings in both concave and convex trials, using only effective distance at collision. The performance of the α-shape model fit was measured by root mean squared error (RMSE) between the participant's response and the model prediction.

In Figure 4B, the *x*-axis shows the α parameter using the average area-difference percentage between the original shape and the approximate body, with the left-most point (0) corresponding to body = shape, and the right-most point corresponding to body = convex hull. The *y*-axis shows RMSE and 95% CIs when using different α values to predict participant responses (lower values suggest a closer fit). It is important to note that in general, the raw α value is not linear, and is less informative because the absolute magnitude of the α value range may vary across tasks, depending on the scale of raw images. For example, a slight increase in α may cause an entire section of an object to be much more coarsely approximated, but any additional increase in α hardly changes this approximation. We only cared about the cognitively relevant and interpretable outcome of changing α, which is the relative difference in area between the original shape and the approximate body with respect to the original shape, and this is what is shown in Figure 4.

We found that the best-performing α value (indicated by the larger red dot, Figure 4B, left) corresponds to an average area-difference percentage between the approximate body and the original shape of 21.9%. This best α-shape model accurately explained participant causality ratings for both concave and convex conditions (Figure 4C, left), and this parameter setting aligned with a body approximation that is between a convex hull and a fine-grained shape. These results further support the hypothesis that people's coarse body approximation is different from the fine-form shape.

Experiment 2: TTC

Experiment 2 tested predicted collision times between two objects, varying concave and convex collision types, and the ground-truth collision times. The experiment was based on classic TTC tasks (Gray & Thornton, 2001; Rosenbaum, 1975; Tresilian, 1995), but with varying bodies. Similar to the logic behind Experiment 1, we expected that in Experiment 2, participants' response time profile between concave and convex conditions will reflect the use of body approximations to track objects. Specifically, because approximate bodies fill in concavities, we expected people to predict concave collisions will occur sooner than convex collisions, for the same ground-truth collision time.

Experiment 2—including design, hypotheses, analyses, and exclusion criteria—was preregistered at https://osf.io/unfzd.

Participants

Experiment 2 recruited 226 participants through Prolific (female = 118, male = 105, prefer not to say = 1, unknown/expired = 2). The median completion time of the study was 17.8 min. We excluded 48 participants for failing to answer at least three of four catch/comprehension questions ("What is your task in this experiment?" "Which Entity was the agent in this study?" "Which entity might vanish in this study?" "What is the color of the sky?"), or for having half of their responses on the control trials being outliers. We excluded participants' TTC data in a control or test trial if it scored as an outlier (i.e., 3 standard deviations away from the mean). After applying the exclusion criteria, 178 participants were left for analysis.

Design and Procedure

Participants saw 4–5 s videos of two objects, an agent and a patient. At first, the agent and the patient were stationary. After 1.6 s, the patient faded away (i.e., gradually became invisible), and the agent began moving at a constant speed towards the now-invisible patient (Figure 3A, middle). Participants were asked to press the spacebar at the moment when they predicted the agent and the patient collided.

As in Experiment 1, we created concave and convex conditions, by varying the collision side of the patient. Based on Experiment 1, we slightly simplified the irregular shapes (e.g., making the concave divots larger), and slowed the object moving speed, to obtain more accurate motor responses. The same irregular agent object was paired with varying irregular patient objects.

Our dependent variable was the TTC, meaning the time difference between the agent initiating motion and the participant pressing the spacebar. We used six different initial horizontal distances between the agent and the patient, ranging from 1.6 to 2.5 times the length of the agent. These corresponded to six ground-truth TTCs: 3,500 ms, 3,700 ms, 3,900 ms, 4,100 ms, 4,300 ms, and 4,500 ms.

In addition to the test videos showing collisions between an irregular agent and a vanishing irregular patient, we used several control trials. These control trials showed videos with either regular boxes colliding, or collisions in which the patient did not vanish, but rather remained visible throughout the video. These box-collision and nonfading control trials were used to establish baselines, ceiling performance, and exclusion criteria.

In total, we had 120 videos (96 test videos: 6 ground-truth TTC conditions × 8 irregular patient shapes × 2 side-of-collision conditions; 24 control videos: 20 with irregular shapes but no vanishing + 2 with boxes and vanishing + 2 with boxes and no vanishing). We randomized the moving direction of the agent (left-to-right or right-to-left) and the colors of objects in each video to control for visual complexity.

Results

We predicted that people's TTC should be based on the approximate bodies of the agent and the patient, in which case people's TTC for concave collisions should be smaller than convex collisions for the same ground-truth TTC. This is because a coarse body representation partially fills in shape concavities, making the perceived distance that the agent must travel to contact the patient shorter in concave collisions.

We first preprocessed all TTC data by correcting them using nonvanishing control data. We used these control trials to estimate baseline motor reaction error. The corrected TTC was calculated by

subtracting a participant's raw TTC from their mean error in the control trials. All following analyses used the adjusted TTC data.

We compared the distributions of participant TTC data in concave and convex collisions across all ground-truth TTC conditions using kernel density estimation, with a Gaussian kernel to estimate the ΔTTC distribution for concave and convex trials separately. The *ΔTTC* was defined as the difference between participant TTC and the ground-truth TTC. We then compared the mean and 95% CI between the estimated $\Delta TTC_{concave}$ distribution and the ΔTTC_{convex} distribution to see if they were significantly different.

The average difference (and the 95% CI) between participant concave TTC and ground-truth TTC (i.e., $\Delta TTC_{concave}$) was 2.4 ms [−1.4, 6.2]; the average difference between participant convex TTC and ground-truth TTC (i.e., ΔTTC_{convex}) was 34.3 ms [31.8, 36.7]. As expected then, people overall behaved as if concave collisions happened earlier than convex collisions, given the same ground-truth TTCs.

We next considered each ground-truth TTC separately and tested if the difference between concave and convex TTC still held. We used paired *t* tests ($\alpha = 0.05$) and found that participant concave TTC and convex TTC were significantly different in every ground-truth TTC condition (Figure 3C, middle): $T(1161) = -4.4$, $p = 1.3 \times 10^{-5}$ in 3,500 ms ground-truth TTC condition; $T(1,181) = -4.4$, $p = 1.2 \times 10^{-5}$ in 3,700 ms ground-truth TTC condition; $T(1,182) = -5.7$, $p = 1.2 \times 10^{-8}$ in 3,900 ms ground-truth TTC condition; $T(1,179) = -5.9$, $p = 3.9 \times 10^{-9}$ in 4,100 ms ground-truth TTC condition; $T(1,133) = -7.2$, $p = 8.6 \times 10^{-13}$ in 4,300 ms ground-truth TTC condition; and $T(1,145) = -10.3$, $p = 2.2 \times 10^{-16}$ in 4,500 ms ground-truth TTC condition. This indicates that participants always predicted that concave collisions happened earlier than convex collisions regardless of the ground-truth TTC. The results again align with our hypothesis.

An exploratory analysis further found that as the ground-truth TTC increased, the difference between convex TTC and concave TTC also increased. This is in line with a memory effect on the coarseness of the object approximation, such that the body approximation grows coarser in working memory over time.

Taken together, the results indicate that (a) there is a difference in TTC between concave collisions and convex collisions, (b) this difference is predicted by a coarse body representation that partially fills in object concavity, and (c) the difference increases over time, which is in line with increasing coarseness in the body approximation over time. However, it is possible to explain this last result in different ways, which we consider in the discussion.

α-Shape Model Analysis

As in Experiment 1, we considered an α-shape model to more finely test our approximation hypothesis. We again examined a range of approximation parameter values, ranging from convex hulls to fine forms. Every α value produced approximated representations for the agent and the patient, which dictated the effective TTC. We used a hierarchical linear model to predict participant TTC responses, using the effective TTC as input and taking into account individual participant differences in TTC. The α-shape model performance was measured in RMSE.

We supposed that body approximations can change across different scenarios (e.g., task context, memory load, incentive, etc.), meaning that we do not assume that the α parameter should stay the same across all contexts. We intended to test whether the object representation people used in this task is different from a shape representation. In this experiment, we considered both a static α (i.e. α being consistent across all ground-truth TTC conditions) and a time-varying α (i.e., α varying across ground-truth TTC conditions). The static α corresponds to the notion that people's body approximations are fixed. The time-varying α corresponds to the notion that people's body approximations may grow coarser over time, for example, due to a memory effect. Both the static and the time-varying versions indicated that people use coarse approximations that are different from convex hulls and fine forms. See the online supplemental materials for more details on this analysis.

As shown in Figure 4B (middle), the best time-varying α-shape in each ground-truth TTC condition (3,500–4,500 ms) filled in the concave divot for an average of 7.9%, 7.9%, 8.6%, 8.6%, 9.9%, and 11.9% in size with respect to the original shape. The best time-varying α-shape parameters reproduced the memory effect in participant data (Figure 4C, middle), such that the difference between convex and concave TTC increased over time. These results again support the claim that participants were using a body approximation that is different from a fine-grain shape representation.

Experiment 3: Change

Experiment 3 was based on classic change detection tasks (Brady et al., 2009; Simons & Rensink, 2005). We adopted the infant change detection paradigm from Xu and Carey (1996), in which an object (e.g., a duck) moves behind an occluder, and another object (e.g., a truck) emerges. This type of paradigm is used to explore object individuation and identity tracking in infants (Kibbe & Leslie, 2019; Rivera & Zawaydeh, 2007; Spelke et al., 1995).

In this experiment, we predicted that if people use approximate bodies for physical tracking, then participants would notice changes at a higher rate when the added area caused larger changes to the underlying body representation of the object. In addition, we hypothesized that in an experiment in which no direct physical tracking of motion was involved, changes that happened within the body representation would become easier to detect, because the absence of physical movement would result in less dependence on a body representation.

Experiment 3—including design, hypotheses, analyses, and exclusion criteria—was preregistered at https://osf.io/krzq2 (Experiment 3a) and https://osf.io/nre7s (Experiment 3b).

Participants

We recruited 60 participants each for Experiments 3a and 3b through Prolific. The mean completion time was 20 min. In Experiment 3a, we excluded participants who did not submit full data, failed attention checks, or answered catch questions incorrectly. The catch/comprehension questions were "Which key to press if objects are the same, P or Q?" and to describe the task. (There was one additional catch question of "what is 1 + 1," but participants were unable to submit data before correctly answering that). Additionally, we excluded participants who had < 50% accuracy on the task in total or below 75% accuracy in our "catch shape" trials. These accuracy criteria were quite lenient, so it is possible participants who fell below those thresholds were not paying attention to

the task. After applying the exclusion criteria, we were left with 56 participants. In Experiment 3b, 50 participants remained after applying the same exclusion criteria. For both Experiments 3a and 3b, the demographic questionnaire was optional. We asked for participants' age and gender (choosing from male, female, other, or prefer not to answer). The majority of participants chose not to provide demographic information.

Design and Procedure

Experiment 3a tested change detection for a shape moving behind an occluder, where a change could happen within or outside a potential body approximation. Experiment 3b controlled for the physical motion in Experiment 3a.

In Experiment 3a (Figure 3A, right), participants saw 4 s videos in which an object moved horizontally behind a centrally placed occluder. The object was briefly out of sight when it moved behind the occluder, and then either the same object or a modified object emerged out of the occluder and continued moving horizontally until out of the screen. After each video, participants were asked to report whether or not they detected a change to the object, and this binary measure was our dependent variable. We also recorded participants' confidence ratings for each of their binary responses, on a discrete scale from 1 to 9.

The base objects were the same eight irregular shapes used in Experiment 1. Modified objects had areas added to them, in three locations (concave, nofill, and convex) and two sizes (small and large). As illustrated in Figure 3B (right), the *concave* condition filled an innermost concavity of an object, the *nofill* condition had the added area still within a big concavity, but not necessarily filling in the innermost position, and the *convex* condition had the added area at a convex edge of the object. The pixel-area change and form of the added area were the same across locations, within a size and shape setting.

If the body representation is the same as the shape, then changes in the concave, nofill, and convex conditions should be noticed at the same rate, because they all violate the same amount of shape. If the body approximation is a convex hull, then nofill and concave changes should be equally harder to detect than convex changes, because concave and nofill changes both happened within a convex hull body. If the body approximation is somewhere between a convex hull and a fine-grained shape (as suggested by Experiments 1 and 2), then concave changes should be harder to detect than the nofill and convex conditions, because only concave changes would fall within the body.

We randomly inserted 12 catch trials as attention checks, in which a simple square stayed as a square (6 trials), or changed into a triangle (6 trials). We balanced the number of videos showing change and no change and randomized the horizontal motion of the object (either from left to right, or from right to left) as well as its color. In total, there were 96 test trials (8 shapes $\times$ 3 change types $\times$ 2 change sizes $\times$ 2 change/no-change conditions).

In Experiment 3b, we replicated the design and measures of Experiment 3a, except that we used static images instead of videos of a moving object as stimuli. Participants watched a stationary object at the center of the screen for 1 s, after which the object disappeared for 2 s (this period matched the approximated time that the object was hidden behind the occluder in Experiment 3a). The same object or a modified object then appeared for another 1 s. As in Experiment 3a, we used three location conditions (concave, nofill, and convex), and two sizes for the change.

In addition to these studies, we implemented a simple alternative (Experiment 3c) that encouraged people to use fine-grain shape representations in the change detection task. That is, while we are arguing that people use body representations in tasks that do not involve recognition, it is useful to show tasks in which the shape *is* used, for example, for comparison.[1] In Experiment 3c, participants saw images containing two objects side by side. The task was simple: participants had to determine whether the two objects in an image are the same or different, without time constraints. We used four irregular shapes and the catch shape from Experiment 3b, balanced the number of change and no-change trials, and randomized the object color.

Results

In both Experiment 3a and 3b, we calculated the average percentage of noticing a change (and the standard error of measurement [SEM]) in all change types (i.e., concave, nofill, and convex), using only the data from change trials, and aggregating across two change sizes. We compared the percentage of noticing change among change types by paired t tests ($\alpha = 0.05$).

As shown in Figure 3C (right), we found that indeed the odds that participants noticed a change were the lowest for the concave trials, significantly lower than either nofill trials (sample mean difference = 22.2%, $t(55) = 9.49$, $p < .001$; $d = 1.27$; 95% CIs = [17.5%, 26.9%]) or than convex trials (sample mean difference = 18.7%, $t(55) = 7.29$, $p < .001$; $d = 0.97$; 95% CIs = ~[13.5%, 23.9%]). However, the difference between the nofill and convex trials was not significant (sample mean difference = 3.4%, $t(55) = 1.76$, $p = .084$; $d = 0.24$; 95% CIs = [− 0.4%, 7.3%]). This main effect of accuracy pattern across change types still held after taking into account shape complexity (see analysis details in the online supplemental materials). This suggests that participants' body approximations are different from the fine-form, and further that the boundaries of the approximation are in between the fine-form and convex hull, in line with findings from Experiments 1 and 2.

Next, to compare results from Experiments 3a and 3b, we performed a generalized linear regression with a logistic link function (i.e., the binomial family) on participant data from Experiment 3a and 3b. The parameters included the main effect of change type (concave, nofill, and convex), the main effect of experiment version (Experiments 3a and 3b), and their interaction. We report the deviance and significance level of the analysis of variance (ANOVA) χ^2 tests on the regression model.

The findings from Experiment 3b replicated the overall pattern of Experiment 3a, with change detection being easier across the board (Figure 3C, right). The two-way logistic ANOVA showed that the interaction between the change type and experiment version was not significant, and both main effects of change type and experiment version were significant, interaction: $\chi^2(2) = 1.08$, $p = .58$; change type main effect: $\chi^2(2) = 211.60$, $p < .001$; experiment version main effect: $\chi^2(1) = 112.45$, $p < .001$. This suggests that the visual task in Experiment 3b was easier than the physics-tracking task in Experiment 3a, but without a differential effect on detecting concave changes. It is possible that having the before- and after-image

[1] We thank a reviewer for this point.

presented sequentially but with a temporal gap, caused people to maintain a coarse body-like representation in memory in order to perform a visual comparison.

As for the simple comparison, in Experiment 3c, in which people had the opportunity to compare two shapes without time constraints: participants were now at near-ceiling performance for all conditions (see Figure S26 in the online supplemental materials). There was no significant difference between participant accuracy on the concave and convex trials ($p = .16$), and between convex and nofill accuracy ($p = .32$), though participants did notice changes at a lower rate in concave change trials compared to nofill change trials ($p = .031$). Importantly, even this difference became nonsignificant when taking into account the number of vertices that changed. That is, a further generalized linear model analysis indicates that the concave–convex and concave–nofill gap becomes nonsignificant (chi-squared likelihood ratio test result between the full model and the restricted model without the change type variable: $\chi^2 = 4.84$, $df = 2$, $p = .089$). This supports the intuition that people can use fine-grained shape representation to detect small differences between two shapes, and that the minor accuracy difference between concave and nofill, in this case, can be explained by the number of vertices changed (nofill changes usually create more vertices than other change types, and could be easier to notice). The online supplemental materials contains the full details and analysis for this experiment. Future work is still needed to compare our results with a version of Experiment 3 that heavily involves shape representation while maintaining the same level of task difficulty.

α-Shape Model Analysis

As in Experiments 1 and 2, we tested different α values for an approximation model, ranging from convex hulls to the fine forms of the original shapes. Each α setting produced an approximation for the objects before and after a change. To calculate the relative amount of effective body violation, we aligned the approximations before and after the change, extracted the area that was different between the two approximations, and calculated its size ratio with respect to the size of the before-change approximation.

Independent of the approximation, we took the complexity of the original shape into account, as we found that the odds of noticing a change varied across shapes, which themselves varied in complexity. We parameterized visual complexity as the number of vertices a given shape had before entering the occluder. We used the effective area change ratio x calculated above and *complexity* to predict the percentage of noticing a change $P(\text{change})$, with the logarithm functional form (other functional forms are discussed in the online supplemental materials):

$$P(\text{change}) = [(P(\text{falseAlarm}) + a)] \times \log(e + b \times x) - a + k \times \text{complexity}.$$

Free parameters a, b, and k were estimated using least squares optimization. The constant $P(\text{falseAlarm})$ was the false alarm rate of participants reporting a change in the no-change trials containing irregular shapes. Performance was measured using the mean RMSE across averaged concave predictions, averaged nofill predictions, and averaged convex predictions. The best-performing α-shape model matched people's performance in Experiment 3a and replicated the qualitative finding that changes in the concave condition were less likely to be detected than in the nofill or convex conditions (Figure 4B and C, right). The best-performing model on average filled in 4.6% of the concavities in size of the original shape. We stress that this should not be taken to suggest that the true underlying body approximation fills in concavities to this specific amount, but simply that the approximation fills in the concavities to some degree in between a convex hull and a fine-grain form, and further work should elucidate the specific approximation people used.

Discussion

Interacting with the physical world in real-time presents a computational challenge. We proposed that a central and useful approximation for dealing with this challenge is the simplification of physical bodies. We examined whether people actually use such an approximation by constructing concave and convex conditions in several variations of classic psychophysical tasks.

Our results suggest that people do indeed use a coarse body approximation for reasoning about the behavior of objects and that this body representation can accommodate representations in between convex hulls and fine-grained forms. There are many ways to model people's uncertainty over the behavior of objects, but a general "fuzziness" does not reproduce the distinction between convex and concave shapes. We explored several possible simplification models, and the one that best accounted for people's data was the α-shape model. This model uses one simple parameter, a knob that dials the approximation between a fine-grain shape and a coarse convex hull, which allowed us to compare predictions along this continuum of hypotheses.

The use of body approximations is in line with the general proposal that people's intuitive physics is not a perfect simulation, but rather relies on principled shortcuts and workarounds (Bass et al., 2021; Battaglia et al., 2013; Smith & Vul, 2013; T. D. Ullman et al., 2017). It also supports the proposal that cognitive scientists can use the principled approximations of real-time simulations as working hypotheses for cognitive models of intuitive physics. Other approximations to explore in human cognition (T. D. Ullman et al., 2017) include the static/dynamic distinction (physics engines often treat objects that actively participate in simulations as dynamic, and others such as walls/floors as static), and the wake/sleep distinction (dynamic objects that are not moving or expected to move are marked as "sleeping," with no need to rerender or recalculate dynamics from frame to frame). While our experiments provide a first step in showing that people use body approximations for reasoning about physical events, further work is required to determine when people use approximate body representations, how they are formed, and how they might change across time and tasks.

It is likely that people's approximations are task- and context-specific in a dynamic way, which takes into account available mental resources, the importance of the task, and scene-specific variables. For example, the simple α-shape model we considered treated all parts of a given object as equally important, but people might use fewer resources to approximate areas of an object that are less relevant for a given task. For example, suppose an object is about to be hit from the left, then it is less important to spend resources on approximating the object's right-hand side. Or, consider that to catch a cup falling off of the table, it may not be necessary to represent the handle in full detail. Just a cylinder or bounding box approximation would be sufficient to initiate a catching action quickly. However, holding a cup's handle or hanging it requires representing

the handle more precisely. Still, the resulting representation does not equal a fully detailed shape representation. Parts other than the handle of the cup may still be approximated in a coarse form, such as the concavity of the cup where it contains liquid or the fine-grained color or texture of the cup.

The importance and difficulty of a task may also affect the approximation used Vul et al. (2009). For example, if it is vitally important to precisely assess the trajectory of an object, more cognitive resources may be spent on finer-grained approximations to increase accuracy. The results of Experiment 2 also suggest the approximation model is time-variant, with people's approximation growing rougher with time up to a point (the body approximation may grow closer and closer to a convex hull the longer it spends behind an occluder or in memory).

All of these examples and complications are not alternatives to the current proposal, but suggestions for refinement that build on a basic suggestion. In all of our experiments, the α-shape model suggests that the approximations people used are different from the precise shape representations. Our central claim emphasizes a distinction between the representations used for physical reasoning and those used for visual recognition under a resource rationality assumption for human cognition. The possibility that body approximation may vary can easily lend itself to further experiments and additions to the model, to answer the exact form and dynamics of the body representation.

The α-shape model we considered is useful in teasing apart several possibilities for whether and which approximation people use, but it is only one suggestion for the approximations people might use when simplifying two-dimensional shapes. It is quite likely that people do not use exactly this model. Various shape-simplification models have been put forward by mathematicians, and possibly different algorithms are used for two-dimensional versus three-dimensional approximations (Edelsbrunner & Mücke, 1994). Follow-up work can further constrain the different approximation model(s) used by people.

Body approximations may also be influenced by kind information. For example, a cylinder may be used to approximate a mug, but it is important for a prototypical mug that it has a handle. Such information is useful for recognition, but also for making physical predictions. A useful body approximation algorithm may include a library of standard shapes (cf. Smith et al., 2019) that is expanded over time, with language helping to scaffold the importance of different shapes. The failure of infants to detect a change in shape when objects move behind an occluder (Xu & Carey, 1996) may then reflect either a very rough body approximation or the lack of relevant bodies in a standard body library.

Kind information may help constrain body approximations, but this can only happen up to a point, and some insensitivity to kind information may carry through from infancy to adulthood. For example, it was recently shown Kominsky et al. (2021) that people "fill in" the perceived trajectory of objects, even when those objects change identity (from a basketball to a soccer ball). But, this effect did not exist when objects changed spatiotemporal continuity (a basketball is seen coming in from above, then from below). Our proposal predicts such behavior, since body approximations used for physical tracking do not necessarily encode information relevant to identity. Our proposal further predicts that changing the object outside of a rough body approximation will disrupt filling-in effects (e.g., changing a basketball to a much larger basketball or a basketball to a towel).

While body approximations may be useful in many tasks, they are not the only relevant representation for tracking the number of entities in a given scene. Absent other information, 10-month-old infants may fail to distinguish two objects moving behind a screen due to their similar body approximations (T. D. Ullman et al., 2017; Xu & Carey, 1996), but even young infants can use early developing markers such as function-use (Futó et al., 2010), ontological distinctions such as agency/nonagency (Kibbe & Leslie, 2019; Wilcox et al., 2010), and so on.

Returning to the dorsal–ventral distinction in visual processing in primates (Goodale & Milner, 1992; Kravitz et al., 2011; Schneider, 1969), a body approximation would be in line with information-for-action, rather than recognition. Above and beyond "where" something is, acting on something requires knowing its rough physical form. A small doughnut centered in a particular position is not the same as a large box centered in the same location. In game engines, the body representation is a carrier not just of rough form, but also of orientation, location, and physical properties, such as elasticity and weight. It is an interesting avenue for future research, to examine to what degree this analogy carries into primate visual processing, although it is unlikely to be a neat split (Zimmer, 2008).

In sum, our findings suggest that human perception and reasoning respect the body-shape distinction. We used a contrast between concave and convex trials in three psychophysical tasks to create a dissociation between body and shape. We observed in all three experiments that human behavior in concave trials was significantly different from convex trials, as predicted by a distinction between body and shape representations. We used the α-shape algorithm to produce a specific realization of body representations and found that reasonable α values quantitatively and qualitatively predict human behavior. While our model is unlikely to be a perfect match for people's representations, our finding suggests that they are a decent approximation and provides indirect evidence to a coarse body approximation.

Constraints on Generality

Across our experiments, we recruited participants from online platforms (Amazon Mechanical Turk and Prolific), restricting the participant pool to US-based adults. Although both platforms contain representative samples in the US, such a population can still limit the generality of our findings. For example, it is not guaranteed that the results we report here can be generalized to people from other countries. As we are using variations on basic psychophysical tasks which have been examined cross-culturally, we conjecture that our results will generalize more broadly, but this remains to be shown empirically.

References

Bass, I., Smith, K., Bonawitz, E., & Ullman, T. (2021). Partial mental simulation explains fallacies in physical reasoning. *Cognitive Neuropsychology*, *38*(7–8), 413–424. https://doi.org/10.1080/02643294.2022.2083950

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 of the United States of America*, *110*(45), 18327–18332. https://doi.org/10.1073/pnas.1306572110

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative and Integrative Biology*, *2*(1), 1–3. https://doi.org/10.4161/cib.2.1.7297

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43. https://doi.org/10.1016/0010-0285(75)90003-1

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3), Article e57410. https://doi.org/10.1371/journal.pone.0057410

Edelsbrunner, H., Kirkpatrick, D., & Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, *29*(4), 551–559. https://doi.org/10.1109/TIT.1983.1056714

Edelsbrunner, H., & Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics (TOG)*, *13*(1), 43–72. https://doi.org/10.1145/174462.156635

Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(34), E5072–E5081. https://doi.org/10.1073/pnas.1610344113

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, *117*(1), 1–8. https://doi.org/10.1016/j.cognition.2010.06.003

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Gray, R., & Thornton, I. M. (2001). Exploring the link between time to collision and representational momentum. *Perception*, *30*(8), 1007–1022. https://doi.org/10.1068/p3220

Gregory, J. (2018). *Game engine architecture*. CRC Press.

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition*, *157*, 61–76. https://doi.org/10.1016/j.cognition.2016.08.012

Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science*, *24*(6), 433–439. https://doi.org/10.1177/0963721415605831

Kibbe, M. M., & Leslie, A. M. (2019). Conceptually rich, perceptually sparse: Object representations in 6-month-old infants' working memory. *Psychological Science*, *30*(3), 362–375. https://doi.org/10.1177/0956797618817754

Kominsky, J. F., Baker, L., Keil, F. C., & Strickland, B. (2021). Causality and continuity close the gaps in event representations. *Memory and Cognition*, *49*(3), 518–531. https://doi.org/10.3758/s13421-020-01102-9

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, *28*(11), 1649–1662. https://doi.org/10.1177/0956797617719930

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Leslie, A. 1994). A theory of ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge University Press .

Li, J., Oksama, L., & Hyönä, J. (2019). Model of multiple identity tracking (MOMIT) 2.0: Resolving the serial vs. parallel controversy in tracking. *Cognition*, *182*, 260–274. https://doi.org/10.1016/j.cognition.2018.10.016

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1–e60. https://doi.org/10.1017/S0140525X1900061X

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2020). Broken physics: A conjunction-fallacy effect in intuitive physical reasoning. *Psychological Science*, *31*(12), 1602–1611. https://doi.org/10.1177/0956797620957610

Luebke, D., Reddy, M., Cohen, J. D., Varshney, A., Watson, B., & Huebner, R. (2003). *Level of detail for 3D graphics*. Morgan Kaufmann.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360. https://doi.org/10.1177/0956797613495418

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt.

Michotte, A. (1963). *The perception of causality*. Basic Books.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, *83*(5), 2580–2601. https://doi.org/10.1152/jn.2000.83.5.2580

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Rivera, S. M., & Zawaydeh, A. N. (2007). Word comprehension facilitates object individuation in 10-and 11-month-old infants. *Brain Research*, *1146*, 146–157. https://doi.org/10.1016/j.brainres.2006.08.112

Rosenbaum, D. A. (1975). Perception and extrapolation of velocity and acceleration. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(4), 395–403. https://doi.org/10.1037//0096-1523.1.4.395

Saiki, J. (2002). Multiple-object permanence tracking: Limitation in maintenance and transformation of perceptual objects. *Progress in Brain Research*, *140*, 133–148. https://doi.org/10.1016/S0079-6123(02)40047-7

Saiki, J., & Holcombe, A. O. (2012). Blindness to a simultaneous change of all elements in a scene, unless there is a change in summary statistics. *Journal of vision*, *12*(3), Article 2. https://doi.org/10.1167/12.3.2

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), Article 411. https://doi.org/10.1037/a0031912

Schneider, G. E. (1969). Two visual systems: Brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science*, *163*(3870), 895–902. https://doi.org/10.1126/science.163.3870.895

Schwettmann, S., Tenenbaum, J. B., & Kanwisher, N. (2019). Invariant representations of mass in the human brain. *eLife*, *8*, Article e46619. https://doi.org/10.7554/eLife.46619

Sereno, A. B., & Maunsell, J. H. (1998). Shape selectivity in primate lateral intraparietal cortex. *Nature*, *395*(6701), 500–503. https://doi.org/10.1038/26752

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

Smith, K. A., Battaglia, P. W., & Vul, E. (2018). Different physical intuitions exist between tasks, not domains. *Computational Brain and Behavior*, *1*(2), 101–118. https://doi.org/10.1007/s42113-018-0007-3

Smith, K. A., Mei, L., Yao, S., Wu, J., Spelke, E., Tenenbaum, J. B., & Ullman, T. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 8983–8993). Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2019/file/e88f243bf341ded9b4ced444795c3f17-Paper.pdf

Smith, K. A., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*, *5*(1), 185–199. https://doi.org/10.1111/tops.12009

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*(2), 113–142. https://doi.org/10.1111/bjdp.1995.13.issue-2

Suchow, J. W., & Alvarez, G. A. (2011). Motion silences awareness of visual change. *Current Biology*, *21*(2), 140–143. https://doi.org/10.1016/j.cub.2010.12.019

Tresilian, J. (1995). Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Perception and Psychophysics*, *57*(2), 231–245. https://doi.org/10.3758/BF03206510

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*(3), 193–254. https://doi.org/10.1016/0010-0277(89)90036-X

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. https://doi.org/10.1016/j.tics.2017.05.012

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive Psychology*, *104*, 57–82. https://doi.org/10.1016/j.cogpsych.2017.05.006

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. A. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems*. Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2009/file/d79aac075930c83c2f1e369a511148fe-Paper.pdf

Wilcox, T., Haslup, J. A., & Boas, D. A. (2010). Dissociation of processing of featural and spatiotemporal information in the infant cortex. *NeuroImage*, *53*(4), 1256–1263. https://doi.org/10.1016/j.neuroimage.2010.06.064

Xu, F. (2005). Categories, kinds, and object individuation in infancy. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 63–89). Lawrence Erlbaum Associates.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153. https://doi.org/10.1006/cogp.1996.0005

Zimmer, H. D. (2008). Visual and spatial working memory: From boxes to networks. *Neuroscience and Biobehavioral Reviews*, *32*(8), 1373–1395. https://doi.org/10.1016/j.neubiorev.2008.05.016

Zosh, J. M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *Journal of Experimental Child Psychology*, *113*(3), 322–336. https://doi.org/10.1016/j.jecp.2012.07.005

Received August 1, 2022
Revision received April 14, 2023
Accepted April 19, 2023 ■