Flexible and Efficient Use of Visual Motion Features in the Perception of Physical Object Properties

Vivian C. Paulun[1,2,3*], Florian S. Bayer[3], Joshua B. Tenenbaum[1], and Roland W. Fleming[3,4]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States

[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, United States

[3]Department of Experimental Psychology, Justus Liebig University Giessen, Germany

[4]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen

[*]To whom correspondence should be addressed. Email: vpaulun@mit.edu

Abstract

To interact with the physical world, intelligent agents must infer object properties like elasticity by sight. While this challenges artificial systems, humans do so effortlessly. We showed observers physics-based simulations of chaotically bouncing cubes, where small differences in initial conditions produced starkly different trajectories. Yet, observers predicted physical elasticity accurately. Here, we propose a resource-rational model based on the statistics of how objects typically behave. By analyzing the simulated trajectories of 100k bouncing cubes, we identified 23 motion features that could be used heuristically to estimate elasticity. Carefully synthesized stimuli allowed us to disentangle these highly correlated hypotheses. We find that humans can use different motion features to judge the elasticity of bouncing objects but rely on only one feature at a time in any given stimulus context. Depending on the information available, observers switch flexibly between different computationally efficient yet highly informative heuristics. Because these heuristics can be derived from natural motion variation across situations, they could plausibly be learned in an unsupervised fashion from everyday experience.

Keywords: Visual Perception, Intuitive Physics, Computational Rationality

Author Contributions

VCP, FSB, and RWF conceived and designed the study, developed the features and computational model. VCP simulated the data set, rendered the stimuli, collected and analyzed the data, made the figures, and wrote a first draft of the manuscript. All authors discussed the results and wrote and approved the manuscript.

Introduction

To grasp, catch, stack, or avoid objects, we need to infer their physical properties such as elasticity, mass, compliance, or friction [1–7]. In most cases, we see objects before we interact with them, making vision the primary source of information to perceive and predict the physical world. Still, researchers do not yet fully understand the cues and computations the brain relies on to estimate the internal properties of objects [8–22]. Unlike an object's shape, size or identity, physical properties like mass or elasticity can only be *inferred* from the observed behavior of the object or substance [8–10,13,14,17], e.g., how a fluid flows, jelly wobbles or a ball bounces. The challenging nature of such inferences is underlined by the fact that even though AI models have matched or surpassed human performance in tasks like object recognition[23,24] or segmentation[25], they still struggle with intuitive physical reasoning[26,27], especially for non-rigid objects. What makes visual inference of physical properties so difficult?

Consider a bouncing elastic object: How it bounces depends on many factors besides its elasticity, e.g., the initial direction and force with which it was thrown. An individual object can produce an infinite variety of trajectories, i.e., spatiotemporal paths, while objects with different elasticities can trace very similar paths depending on other factors, such as the object's initial speed, height or direction of motion (**Figure 1A**). In previous work[8], we have shown that observers estimate the elasticity of bouncing cubes based on their motion trajectory. But, if there is no unique mapping between an object's elasticity and its trajectory, how does the brain estimate the former from the latter?

Warren and colleagues[28] suggested that observers use the relative height of a simulated, two-dimensional ball around a bounce (i.e. the ratio of initial and final height) to visually judge elasticity, and the duration between two bounces when the ball's height is occluded. While their study elegantly isolated different cues and demonstrated that observers are sensitive to them, it remains unclear how people judge elasticity in more natural settings with more complex trajectories and when no single cue is a perfect determinant of elasticity (such as relative bounce height was in their study).

Our work addresses these two key questions: (1) How do people visually infer elasticity in naturalistic scenes, where no single cue alone perfectly predicts elasticity? (2) How does the brain learn to visually infer elasticity without ever having access to the ground truth? Although individual trajectories vary, motion trajectories of the same elasticity will somewhat resemble each other in terms of their overall characteristic motion features, e.g., bounce height, speed of velocity decay and trajectory length. While no individual feature is perfectly diagnostic of elasticity, variations across different trajectories are also not random because they result from lawful physical constraints. By observing a number of examples, the brain could learn the dominant feature dimensions along which bouncing objects vary and represent elastic objects within the space of these features. A given heuristic (such as the bounce height ratio suggested by Warren and others) could be thought of as a special instance of this, in which the brain might identify just *one* feature along which trajectories are varying and thus elasticity judgments will rely on. However, another possibility is that the brain encodes elastic objects along multiple different features which would lead to a more robust representation in naturalistic settings. By considering different visual features, e.g., number of bounces and bounce height ratio, the brain could overcome the potential pitfalls of single heuristics.

This idea leads to several testable predictions, which we evaluate here. First, motion features can be used to disentangle physical elasticity from other confounding factors (such as initial speed). Second, the relation between physical elasticity and motion features can be learned through observation alone. Third, either a single motion feature (i.e., a heuristic) or a robust combination of several features can explain the pattern of successes and failures in human perception. To test these assumptions, we employed a data-driven approach. For this purpose, we simulated 100,000 short (4 sec) trajectories of a bouncing cube in a room (**Figure 1A**). The cube's elasticity (coefficient of restitution) varied from 0.0 (not elastic) to 0.9 (very elastic) in ten steps. Importantly, we also varied the initial position, orientation, and velocity of the cubes to gain 10,000 different trajectories for each level of elasticity. Although computer simulations are only approximations of the real world, we validated that they reproduce several crucial physical behaviors of bouncing objects[8]. Only through simulation can we generate sufficient number and diversity of trajectories to identify and evaluate statistical regularities. We chose nonrigid, i.e., deformable, cubes as stimuli, because they result in chaotic and highly variable trajectories while being feasible in terms of the parameters to create and analyze them and are, thus, the ideal case example to study. Next, we identified 28 candidate 3D motion features (**Figure 2A-D, Table 1**) based on the physics of bouncing objects, and previously proposed cues[28,29]. We then determined how they statistically relate to physical elasticity in our dataset und used PCA to determine the optimal feature combination to predict elasticity. Our analysis revealed several competing hypotheses of how humans visually judge elasticity using motion features, all of which could be learned in an unsupervised fashion from observation alone. In a series of carefully designed experiments, we selected stimuli that systematically decouple these highly correlated alternative hypotheses and find the one that best predicts human perception on a stimulus-by-stimulus basis. To begin, we first established human accuracy and consistency in elasticity perception in a random subset of our dataset as a benchmark to test out models of perception against.

Results

Observers accurately infer elasticity, but make systematic errors

Fifteen observers rated the apparent elasticity of bouncing cubes in 150 simulated animations—fifteen different trajectories for each of the ten elasticities (see **Methods** and **Figure 1A** and **Movie S1**). Although the initial speed, position, and orientation of the cubes varied randomly, yielding widely variable trajectories, observers were very accurate at estimating the cube's relative elasticity (**Figure 1B**). Average ratings increased systematically with physical elasticity (linear regression: $R^2 = .84$, $F(1, 148) = 748.73$, $p < .001$). However, if observers had perfect elasticity constancy, they would give videos showing the same elasticity the same ratings. This is not what we found: Cubes with identical physical elasticity were perceived to have different elasticities (average SD per elasticity level was 0.09 and significantly different from zero: $t(9) = 16.40$, $p < .001$). Importantly, the pattern of errors was not random but highly consistent between different observers ($r = .91 \pm .04$; M ± SD) as well as within repeated ratings of the same individual (r = .90 ± .04). In fact, there was no significant difference between intra- and inter-observer variability ($t(14) = 2.08$, $p = 0.056$). What causes this systematic pattern of errors? If humans represent elastic objects in terms of their characteristic motion features, perceptual errors should occur whenever a trajectory falls onto an "atypical location" in that feature space. In the following, we test this prediction.

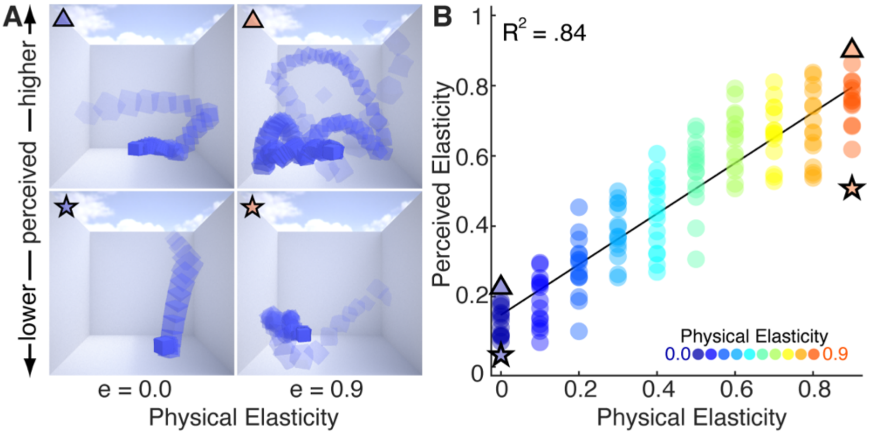

Figure 1. *Stimuli and results of Experiment 1.* ***A)*** *Example stimuli of lowest (e = 0.0) and highest elasticity (e = 0.9), frames of the animations were overlaid for illustration purposes. Even though both images in each row show the same cube (i.e., the same physical properties), the trajectories are different because we randomly varied the initial speed, position, and orientation.* ***B)*** *Average elasticity ratings of Experiment 1 together with a linear fit. Dots of the same color show simulations of the same elasticity but varying initial parameters.*

Motion features disentangle physical elasticity from other latent factors

We propose that the brain represents trajectories of bouncing objects using one or more spatiotemporal features and infers elasticity from their systematic variation. To test this hypothesis, we explored a set of motion features derived from the 3D trajectories of the object. We started with 28 potential features that between them capture many aspects of bounce trajectories (**Table 1**; see **Table S1** for additional details). The features were selected by: (a) consideration of the physics of ideal bouncing objects, (b) proposals from previous literature[28,29], and (c) subjective observations of the simulations. Some features describe characteristics of individual bounces (e.g., average bounce height, rebound velocity) or measure the coefficient of restitution in simplified, idealized settings (e.g., bounce height ratio). Others capture summary statistics that integrate over time and might be useful in realistic scenes that deviate from ideal conditions (e.g., number of bounces, movement duration; **Figure 2A-B**). Such statistics provide several different ways of measuring how quickly the object dissipates kinetic energy as it bounces around. All motion features are stimulus computable from observable quantities, i.e., positions and changes of positions over time, and derived from first principles. We computed the motion features from the trajectories of the cube's center of mass (CoM) and eight corners for all 100,000 simulations (see **Methods** and **Supplementary Methods**). Although object rotation and deformation are important for a complete physical representation of the object's motion, we do not consider them here, as our previous findings show that they have a negligible effect on the perceived elasticity in these stimuli[8]. With this exception we aimed to achieve a comprehensive characterization of the trajectories by defining a diverse set of features to follow a data-driven approach and constrain our hypothesis space based on the data rather than a priori assumptions.

Table 1. *Motion features with % variance in physical elasticity explained. Grey features excluded from further analysis*

%	Feature (acronym; unit)
82.29	Movement duration until the cube stopped moving. (movDur; sec)
78.93	Number of bounces from the floor, the ceiling and the walls. (nBounce)
78.92	Duration until the cube landed after the last bounce from any wall. (bounceDur; sec)
77.87	Number of bounces from the floor. (nBounceFloor)
67.27	Cumulative length of the motion trajectory. (trajLen; m)
52.76	Mean ratio of energy before and after a bounce. (mEnerRatio)
51.91	Mean acceleration over time. (mAccel; m/s^2)
50.80	Conserved energy over time. (consEner)
45.85	Maximum ratio of energy before and after a bounce. (maxEnerRatio)
44.86	Maximum length of bounce arcs from floor (maxArcLenFloor; m)
41.80	Mean ratio of incident to rebound velocity of all bounces. (mVelRatio)
39.42	Maximal ratio of incident to rebound velocity of all bounces. (maxVelRatio)
36.51	Maximal ratio of durations of consecutive bounces from the floor. (maxBounceDurRatio)
35.46	Maximal duration of individual bounces from the floor. (maxBounceDur; sec)
35.21	Maximal rebound velocity of bounces from every wall. (maxReboundVel; m/s)
35.13	Maximal ratio of bounce heights of two consecutive bounces from the floor. (maxBounceHtRatio)
35.10	Maximal height of bounces from the floor. (maxBounceHt; m)
30.84	Mean ratio of bounce durations of consecutive bounces from the floor. (mBounceDurRatio)
23.42	Mean ratio of bounce heights of two consecutive bounces from the floor. (mBounceHtRatio)
16.25	Maximal length of bounce arcs, i.e., trajectory between consecutive bounces. (maxArcLen; m)
6.24	Mean height of bounces from the floor. (mBounceHt; m)
5.49	Mean velocity over time. (mVel; m/s)
5.25	Mean length of bounce arcs, i.e., trajectory, between consecutive bounces. (mArcLen; m)
4.06	Mean length of bounce arcs from floor. (mArcLenFloor; m)
1.86	Difference between movement and bounce duration. (otherMotionDur; sec)
0.77	Mean duration of individual bounces from the floor. (mBounceDur; sec)
0.15	Mean height of the object over time. (mHeight; m)
0.01	Mean rebound velocity of bounces from all walls. (mReboundVel; m/s)

First, we evaluated how well each of the individual features captured the variance across different elasticities. We found that many features varied systematically with physical elasticity (**Figure 2C-D & 3B, Table 1**). The most diagnostic features (which share the most variation with physical elasticity) were those that integrate information over time, such as movement duration or the number of bounces. Interestingly, we found that heuristics that were previously identified for idealized settings, e.g., related to the height and duration of bounces, were not among the best features in our complex scenario. We narrowed our hypothesis space by excluding features that explained < 5 % of the variance from further analysis (greyed items in Table 1). We found that the remaining 23 features were significantly correlated with one another across the set of 100,000 trajectories (mean absolute correlation, M = 0.48; see **Figure S1**). To identify independent dimensions of variation, we applied principal component analysis (PCA) to the normalized and equalized motion features of all trajectories. Representing the trajectories in the space of the first two PCs reveals that physical elasticity varies largely along the first dimension (**Figure 2E**). Indeed, we found that ground truth elasticity and the first PC share 82.83% of their variance. In other words, physical elasticity emerges as the latent variable driving most variance in the feature representation of all trajectories. Although adding further PCs necessarily increases the explained variance of the dataset (**Figure S2A**), adding more PCs to a multiple linear regression model fitted to physical elasticity does not increase the shared variance by much (with all PCs: 86.25%). Moreover, while PC1 robustly predicts physical elasticity, it is mostly independent of the other latent parameters we used to initialize our simulations (e.g., velocity; all < 1.0%, **Figure S2B**). Thus, this linear combination of motion features (see **Figure S3** for feature loadings) successfully disentangles physical elasticity from other scene factors that contribute to the raw physical trajectory of bouncing objects. Notably, this feature weighting is not the result of a fitting process but emerges naturally and without supervision from the statistics across many examples. This underlines the potential of motion features to form a statistical appearance model of bouncing objects in a completely data-driven fashion. In the following we test whether PC1 can explain the perceptual patterns found in Experiment 1 ('multi-feature model'). Importantly, applying a PCA to the raw motion trajectories (**Figure 2F**) does not yield comparable elasticity estimates—highlighting the crucial role of appearance features.

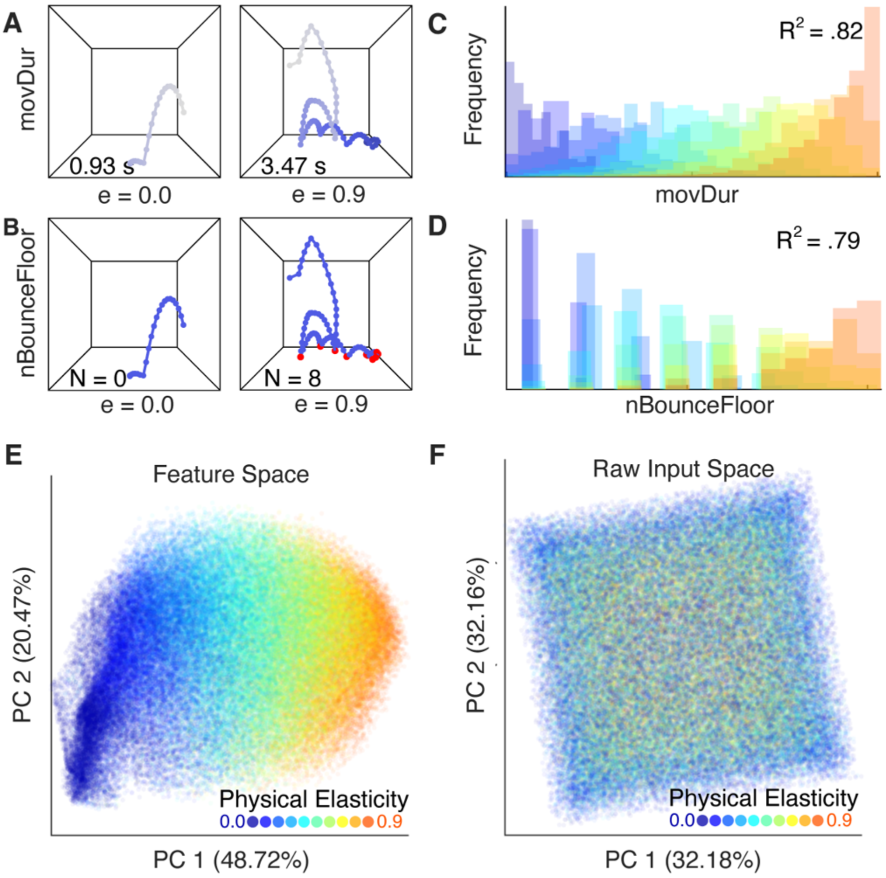

Figure 2. *Spatiotemporal motion features of bouncing objects.* ***A)*** *Example trajectory of a low (e = 0.0) and high (e = 0.9) elastic cube, each dot represents one frame; color gradient represents movement duration.* ***B)*** *The same two trajectories, red dots represent bounces off*

the floor. ***C)*** *Distribution of movement durations in the set of 100,000 trajectories, true elasticity is color-coded.* ***D)*** *Distribution of "number of bounces off the floor" in the set of 100,000 stimuli.* ***E)*** *All 100,000 simulations in the space of the first two PCs resulting from a PCA on the motion features ("feature space"). Physical elasticity (color-coded) seems to vary mainly along the first PC, which explains most of the variance.* ***F)*** *100,000 trajectories in the space of the first two PCs resulting from a PCA on the raw trajectories. Rather than physical elasticity (color-coded), the PCs seem to be related to the position of the cube in 3D space. Note that although the initial position of the cube is uniformly sampled, its 3D position over time is biased due to gravity. This results in a tilted square in the 2D representation of the PCs.*

Optimal motion features predict elasticity perception

Having established that motion features are highly diagnostic of physical elasticity and that their relation to elasticity can be learned without supervision from observation alone, our analysis revealed several strong hypotheses for how the brain could visually infer elasticity. Next, we sought to answer the key question whether the human brain relies on a single motion feature (i.e., a heuristic) when estimating elasticity or instead combines different visual features to a potentially more robust estimate, similar to PC1.

Interestingly, we found that motion features that turned out to be good, i.e., diagnostic, heuristics of *physical* elasticity, were also the best to predict perceived elasticity in experiment 1 (**Figure 3B**). Strikingly, movement duration, the best feature for predicting *physical* elasticity, was also the best to predict *perceived* elasticity ($R^2 = .91$, $F(1, 148) = 1515.1$, $p < .001$, **Figure 3A-B**). On a stimulus-by-stimulus basis, movement duration was a better predictor of human ratings than physical elasticity (evidence ratio: $w_{\mathrm{movDur}}/w_{\mathrm{Physics}} = 1.51e{+}20$). Can a combination of features outperform this? We find that a multi-feature model, i.e., PC1, is a very good predictor of perceived elasticity in Experiment 1 (linear regression: $R^2 = .89$, $F(1, 148) = 1210.5$, $p < .001$, see **Figure 3B-C**). This is impressive given that the feature weighting was derived from observing the covariation of features in a large data set rather than a fitting procedure to the perceptual (or any) data. PC1 predicts perception better than the ground truth does (evidence ratio: $w_{\mathrm{FeatureModel}}/w_{\mathrm{Physics}} = 3.38e{+}13$), but worse than movement duration (evidence ratio: $w_{\mathrm{movDur}}/w_{\mathrm{FeatureModel}} = 4.46e{+}06$; $w_{\mathrm{movDur}} \approx 1$). However, the predictions of both models are strongly correlated ($r = .95$, $p < .001$, in the complete data set). In Experiment 2 we therefore systematically decouple their predictions.

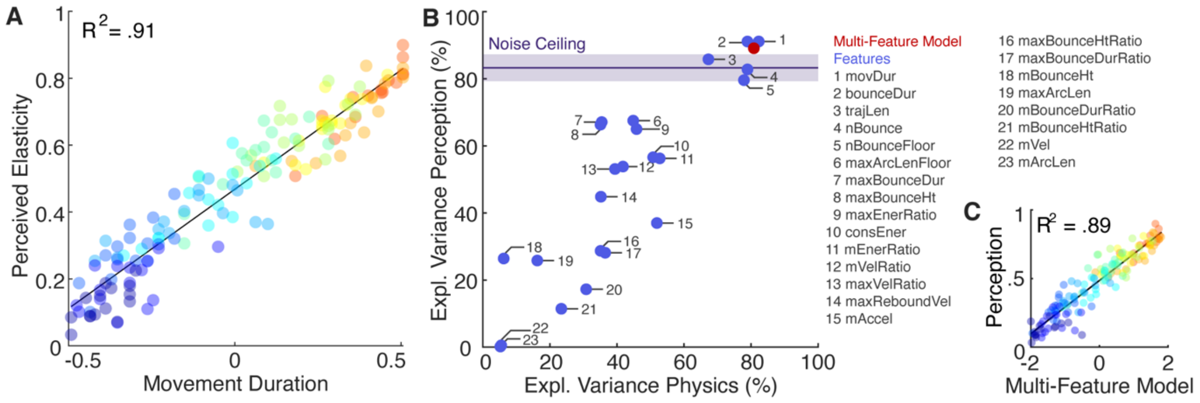

Figure 3. Prediction of perceived elasticity by different compteting models. ***A)*** *Perceived elasticity (in Experiment 1) as a function of the prediction made by the statistically optimal feature: movement duration.* ***B)*** *Explained variance in terms of perceived elasticity (in Experiment 1) as a function of explained variance of physical elasticity (in the data set of 100,000) for individual features (blue) and the multi-feature model (PC1, red). The noise ceiling shows the average explained variance between individual subjects and the average subject (± 95%-CI).* ***C)*** *Rated elasticity from Experiment 1 as a function of the prediction made by the feature combination from PC1, i.e., the multi-feature model.*

When observing complete motion trajectories people use movement duration as a heuristic to elasticity

The aim of Experiment 2 was threefold: First, we systematically decoupled the predictions of the multi-feature model from those of the movement duration heuristic to bring both models into conflict. Second, in order to test whether any of the other features are—individually—a better predictor of perceived elasticity, we systematically decoupled all other features from the multi-feature model. Since it is impossible to isolate each of the 23 features from all other features one by one, we decoupled each feature from the weighted combination of all features to test its causal contribution to elasticity perception. In doing so, we are able to overcome the purely correlational analysis reported so far and experimentally tests 24 competing hypotheses at once, thereby going beyond previous studies [9–15]. Third, because any good model of elasticity perception should be able to predict the pattern of errors on a stimulus-by-stimulus basis, all stimuli in this experiment had the same physical elasticity, i.e., all perceptual differences are illusory. This provides an even more stringent test of our 24 competing models.

For this purpose, we simulated another 90,000 motion trajectories of the cube with medium elasticity (e = 0.5). From the total of 100,000 simulations of medium elasticity, we selected 23 sets of stimuli (one for each of the candidate motion features) for which individual feature and multi-feature model predictions were essentially uncorrelated (|*r*| < .05; see **Methods** and **Figure S4** for more details). A new group of 30 participants judged the elasticity of these stimuli. Note that this rigorous stimulus selection process risks diminishing the very effects we seek to find: We first narrow the range of features by keeping elasticity constant and then select stimuli that, by definition, include outliers with a low correlation between a given feature and PC1.

Although these careful steps may have limited our statistical power, Experiment 2 provided clear results. For each feature, **Figure 4A** shows the correlation of perceived elasticity in the specific stimulus set (chosen for that feature) with the feature prediction (x-axis) and the multi-feature model prediction (y-axis). Seventeen features show a significantly lower correlation with perception than the multi-feature model ($p < .0022$, Bonferroni corrected). Only for one feature—movement duration—does the correlation with perception ($r = .45$) significantly exceed the multi-feature model ($r = .07$, $p < .0022$). In other words, when brought directly into conflict, movement duration can explain perceived elasticity better than a weighted feature combination. Thus, the high correlation between the multi-feature model and perception in Experiment 1 is presumably mediated by the contribution of movement duration (which has the third highest loading of all features to PC1). Is movement duration also driving the high correlations between the multi-feature model and perception in the other stimulus sets of Experiment 2? **Figure S5B** shows the partial correlations between perception and single features vs. perception and multi-feature model prediction when controlling for the effect of movement duration. The correlations between perception and multi-feature model ($r = .56 \pm .14$ (M ± SD)) decrease significantly when controlling for movement duration ($r = .12 \pm .11$; $t(21) = 12.97$, $p < .001$), indicating that movement duration is indeed the driving factor. Across all stimuli, movement duration was—again—the best predictor of perceived elasticity ($R^2 = .78$, $F(1, 223) = 787.61$, $p < .001$, see also **Figure 4B and S5A**). Thus, the longer an object moved in the scene, the more elastic it appeared. Experiment 2 showed that this relation holds true even if physical elasticity is constant, leading to a powerful perceptual illusion. **Supplementary Movie S2** demonstrates these large, systematic, and robust illusory differences in apparent elasticity.

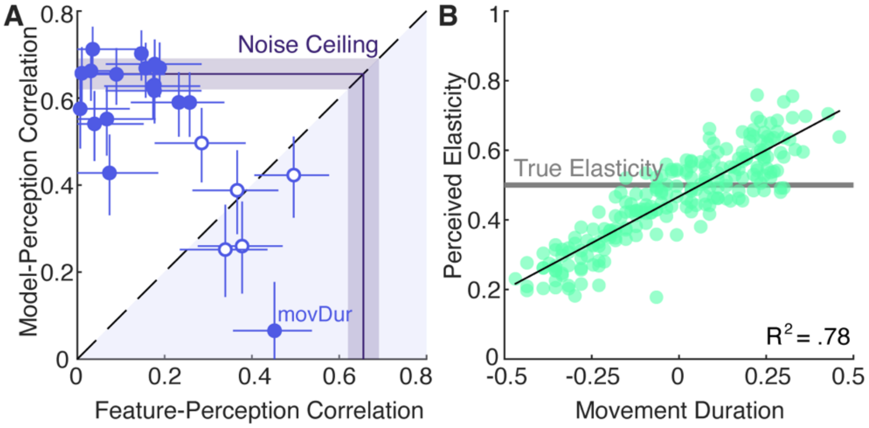

***Figure 4**. Results of the decorrelation experiment. **A)** Correlation of the pooled perceptual ratings with the multi-feature model (y-axis) and the individual features (x-axis). Each dot represents the correlations for one set of stimuli that were specifically selected to decouple the prediction of one feature from the model. Features that fall below the diagonal (light blue shaded area) exceed the model, i.e., their predictions correlate more strongly with perception than the model does. Filled dots indicate a significant difference between the two correlation*

coefficients. Error bars show 95% confidence intervals. Please note, that the correlation coefficients are lower than in Experiment 1 because the data is pooled across participants (instead of averaged) to get a more reliable estimate from the small number of stimuli in each set. For the noise ceiling, we calculated for each stimulus set how much the pooled responses correlate with the average response. The noise ceiling shows the mean (± 95 % - CI) across features. ***B)*** *Average elasticity ratings for all stimuli of Experiment 2 as a function of movement duration together with a linear fit. Elasticity ratings clearly increase with an increase in movement duration. All stimuli had the same physical elasticity of 0.5 (grey line). Thus, all perceived differences in elasticity between stimuli are illusory.*

Observers flexibly switch to another heuristic when movement duration is unobservable

Our everyday experience suggests that we are able to judge an object's elasticity even without observing for how long the object moves, e.g., if someone catches it before it comes to rest. To study systematically whether and how well people can estimate elasticity when this one cue is not available, we truncated a subset of the videos from Experiment 1 to exactly 1 second and presented these to a new group of 15 observers in **Experiment 3**, see **Movie S3**. In these videos it was not possible to observe movement duration. Yet, we found that the average elasticity ratings increased systematically with physical elasticity (linear regression: $R^2 = .73$, $F(1, 78) = 215.41$, $p < .001$, see **Figure S6A**) and showed a near-perfect correlation ($r = .97$, $p < .001$) with ratings for the full movies (Exp. 1; see **Figure 5A**), although the consistency between observers was moderately lower here ($r = .80 \pm .16$, M ± SD) than in Experiment 1 ($r = .91 \pm .04$; $t(28) = -2.63$, $p = .05$). How do observers infer elasticity when movement duration cannot be observed? Do they rely on a different heuristic?

Truncating the videos altered most feature values, not just movement duration. **Figure 5B** shows how well the multi-feature model and the individual features can predict physical as well as perceived elasticity in 1-sec movies. The multi-feature model was the best at explaining both physics and perception and again better explains perception than ground truth physics ($R^2 = .77$, $F(1, 78) = 260.27$, $p < .001$; evidence ratio: $w_{\text{FeatureModel}}/w_{\text{Physics}} = 296.05$; see also **Figure S6C**). Several individual features, particularly those measuring the presence of large bounces in the trajectory (such as maxArcLenFloor or maxBounceHt), also capture a large proportion of the variance in perceived elasticity. To disentangle these competing, but correlated hypotheses, we conducted **Experiment 4** following the same logic as in Experiment 2: From the dataset of 100,000 cubes of medium elasticity, we first identified the simulations that had a movement duration of at least 1 sec. For this subset, we calculated the motion features for the first second and then selected 22 sets of stimuli (one set for each feature except movement duration) in which the prediction of that feature individually was uncorrelated with the prediction of the multi-feature model. A new group of 30 observers estimated elasticity in these 1-sec stimuli.

Again, we found that one of the most diagnostic features—maximum bounce height—showed a significantly higher correlation with perception than the multi-feature model ($r = .54 > r = .25$, $p < .0023$, **see Figure 5C**) when brought directly into conflict, and that was the best predictor of perceived elasticity across all stimuli in Experiment 4 ($R^2 = .74$, F(1, 197) = 565.56, $p < .001$, see **Figure 5D** and **S7A**). Thus, the higher the largest bounce was, the more elastic the cube appeared even if the true elasticity was equal (see **Movie S4**). There was only one other feature—bounce duration—for which the correlation between feature and perception was larger than the correlation between multi-feature model and perception ($r = .36 > r = .10$, $p < .0023$). However, bounce duration did not vary much in the stimulus set, because in most simulations the cube would have bounced for longer than 1 second had the movie not been truncated (see **Figure S7C**). Therefore, bounce duration was only a diagnostic feature when it was notably shorter than one second. For most (12/22) features, we found that the multi-feature model predicted the data better than the individual features ($p < .0023$, Bonferroni corrected). Akin to the results of Experiment 2, these high correlations seemed to be driven by the best single feature, maximum bounce height (see **Figure S7B**). More precisely, the correlations between perception and multi-feature model ($r = .49 \pm .16$ (M ± SD)) decreased significantly when controlling for maximum bounce height ($r = .23 \pm .09$; $t(21) = 6.65$, $p < .001$).

In sum, Experiment 4 showed that observers reported robust perceptual differences between truncated stimuli even though all had the same physical elasticity. Perceived elasticity was best explained by one of the most predictive features, maximum bounce height. Intuitively this makes sense, as the maximal bounce height is easy to compute (i.e., requires only one position) and it occurs within the first second in most trajectories (94.1%, see **Figure S8**). Taken together, this suggests that if unable to fully observe an object's movement until it comes to a standstill, we instead form an impression of its elasticity based on the highest of the bounces that it makes.

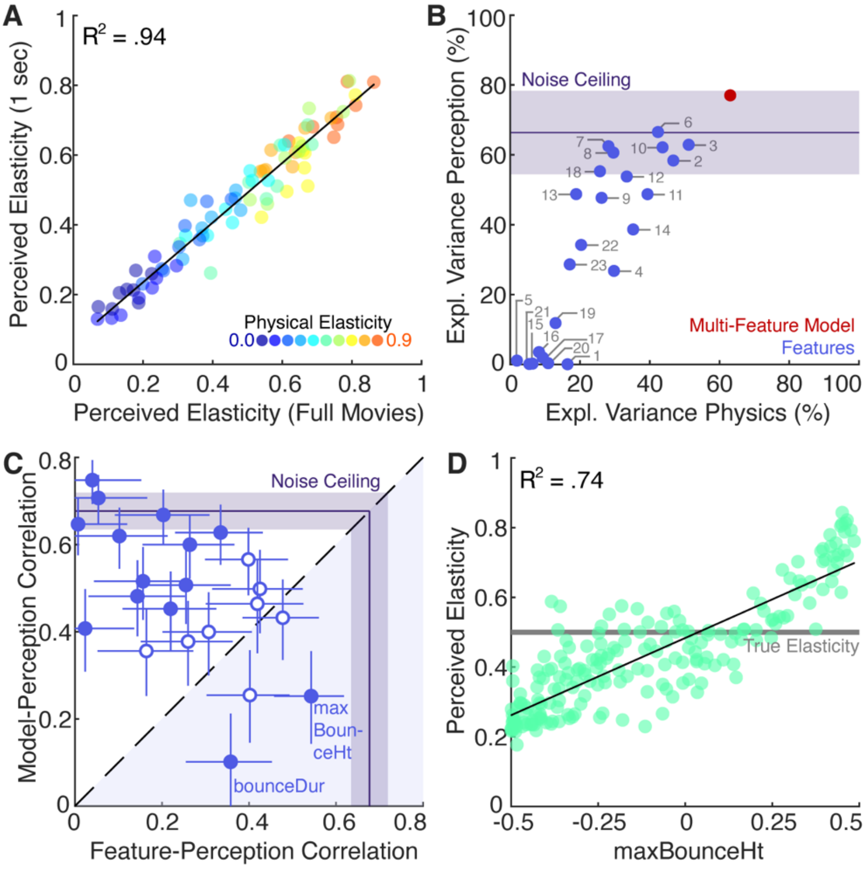

***Figure 5**. Results of Experiments 3 and 4 with truncated movies. **A**) Average perceived elasticity in 1-sec movie clips (Exp. 3) as a function of the movement duration of the apparent elasticity in full movies of the same stimuli. Physical elasticity is color-coded. **B**) Explained variance in terms of perceived elasticity (in Experiment 3) as a function of explained variance of physical elasticity in 1-sec movies (in the data set of 100,000) for individual features (blue), the multi-feature model (red). For a legend of individual features see Figure 2G. The noise ceiling shows the average explained variance between individual subjects and the average subject (± 95%-CI). **C)** Correlation of the pooled perceptual ratings with the multi-feature model (y-axis) and the individual features (x-axis). Each dot represents the correlations for one set of stimuli that were specifically selected to decouple the prediction of one feature from the model. Features that fall below the diagonal (light blue shaded area) exceed the model, i.e., their predictions correlate more strongly with perception than the model does. Filled dots indicate a significant difference between the two correlation coefficients. Error bars show 95% confidence intervals. For the noise ceiling, we calculated for each stimulus set how much the pooled responses correlate with the average response. The noise ceiling shows the mean (± 95 % - CI) across features. **D)** Average elasticity ratings of Experiment 4 as a function of the maximum bounce height together with a linear fit. Elasticity ratings clearly increase with an increase in maximum bounce height. All stimuli had the same physical elasticity of 0.5 (grey line).*

Discussion

Here we propose that when visually judging the physical properties of objects and materials, people often represent them in terms of their typical appearance—i.e., in terms of their typical mid-level spatiotemporal features. Specifically, our results suggest that when asked to judge the elasticity of a bouncing object, observers judge how long the object moves. If the motion duration is cut short, i.e., it cannot be observed, observers instead rely on the maximal bounce height to judge elasticity. This implies a flexible and computationally efficient strategy.

While this study is not the first to suggest a role of mid-level features in the estimation of physical properties[9–15,17,30], it overcomes three critical limitations of previous work. First, we assess the statistical relations between a diverse set of potential visual features and physical elasticity in a large dataset and thereby show how—in principle—observing the variations of motion features in many examples spontaneously reveal elasticity and establish which features (or their combination) are best at doing so. Second, to the best of our knowledge, no study has yet *manipulated* the proposed visual cues to physical properties in naturalistic stimuli. Here, we achieved such manipulation by using a large dataset to identify stimuli that decouple the inherently correlated predictions of different models. Third, we identified *illusory* stimuli that decouple feature predictions from ground truth physics. Thus, we not only predict the good overall performance of observers in elasticity estimation but, critically, also their specific perceptual errors on a stimulus-by-stimulus basis. Our findings have implications on both theoretical and methodological levels.

Learning. We have previously hypothesized that by observing the outside world and its inherent statistical relations[31,32], the brain can learn—in an unsupervised manner—many dimensions along which objects in our environment vary. The statistical appearance model proposed here is not intended as a model of this learning process, but rather a proof of principle about the learnability of the cues and the impact that such unsupervised statistical observation approaches have on perception. We found that by observing various motion features of bouncing cubes, elasticity emerges spontaneously as the main dimension of variation. The motion features themselves were not the result of learning from the stimulus set but instead were explicit operationalizations of our hypotheses. This approach allowed testing the contribution of a large, yet testable number of interpretable motion features and their combination. Would similar features emerge from applying unsupervised or self-supervised learning algorithms? It would be interesting to investigate this question within different frameworks, from deep learning to program learning or simulation-based learning. For example, might the same heuristics be derived within a mental physics simulation model? How would the latent feature space of an (unsupervised) deep learning model compare to the motion features identified here? However, it would be practically impossible to test the individual contribution of the thousands of features in the trained network to perceived elasticity. Yet, here, it is precisely this decoupling of competing hypotheses that ultimately enabled us to predict human perception on a stimulus-by-stimulus basis.

Mid-level features. One of our key findings is that when asked to estimate the elasticity of bouncing objects, observers judge the movement duration or the maximal bounce height in case the duration is visibly cut short. Crucially, this implies that the brain does represent *multiple* features of bouncing objects at a time but does not combine them in the sense of classic cue combination[33] to estimate the latent parameter (elasticity). If the brain represents bouncing objects in terms of their visual motion features, as our results suggest, 'estimating elasticity' means determining the relative position of the observed object on the feature manifold. Across four experiments, we found that observers base their elasticity estimates on only 2-3 visual features. Why would the brain rely on these and not on other features? Presumably, the most effective features are both salient and inexpensive to compute. Movement duration and maximum bounce height both capture important events in the observed motion, i.e., the largest bounce and the end of the motion. It is not trivial to determine the computational costs of different features. Yet, at a minimal level, it seems plausible to assume that single measures, e.g., height or duration, will be computationally cheaper than their derivatives or ratios. In that sense, movement duration and maximum bounce height, are among the computationally simplest features we tested. Duration and spatial distance are quantities the visual system can estimate reliably and accurately[34–37].

Even though we found strong evidence that humans base their elasticity estimates mainly on two motion features, some other features may play an important role in identifying the stimulus as a bouncing object in the first place. A key assumption of our model is that the observed motion is due to a semi-elastic object bouncing in an environment, as opposed to some other cause (e.g., animate motion[38], fluid flow[13,14,39]). If applied to other trajectories the resulting 'elasticity estimate' would be meaningless, e.g., for a feather gliding in the wind or a driving car. An important line of future research is to investigate the cues underlying the recognition process through which we identify the stimulus as a bouncing object in the first place.

The motion features we tested here are stimulus-computable, yet they assume a perfect representation of the object's trajectory. As such, they oversimplify the input available to elasticity-estimating processes in the biological brain. For example, humans have a more accurate representation of image-plane motions than motion in depth[40,41], and may not be equally sensitive to all velocities in these displays. Thus, to transform the heuristic model into a truly image-computable one, future work will also need to incorporate aspects of low-level vision, including object segmentation. Yet we reasoned that important insights into the estimation of material properties can still be gained even without fully modeling all preceding processing stages.

Generalization. Deformable cubical objects produce diverse and complex trajectories. We have shown that visual motion features generalize across large variations caused by several independent factors. Movement duration and maximum bounce height are likely to generalize to some extent across other scenes and objects. For example, if the object had a different shape or if it interacted with other objects in a different space, higher elasticity objects would still tend to move longer and bounce higher. Participants presumably had little experience with bouncing non-rigid cubes prior to our experiments. Yet, they were broadly able to judge elasticity reliably, suggesting they could generalize from previous experience with other scenes and objects. In an experimental setting, it would be possible to break the relation between motion features and elasticity. For example, if the floor was completely inelastic, like sand, no object would rebound. It is, however, unlikely that human observers would be able to estimate the objects' elasticity in these cases. Thus, although motion features would not capture physical elasticity, they might still be reliable predictors of perceived elasticity. Because our model is stimulus computable (based on the true or estimated 3D position), such hypotheses can be easily tested in future research.

Simulation vs. heuristics. A current topic of active discussion is the extent to which physical perception and reasoning proceed through sophisticated but computationally costly internal simulations[19–22,42] or cheaper but potentially less accurate heuristics[9,10,43,44]. How do our results fit into this theoretical spectrum? Representing objects and materials in terms of their appearance features entails an understanding of the observable consequences of natural variations between objects, e.g., the ways in which elastic objects bounce. Yet, the resulting estimation strategy appears like a classic heuristic, i.e., a simple but sufficient rule of thumb such as "the longer it moves, the more elastic it is". In fact, our results could provide an explanation of how the brain derives such heuristics from observation alone and of how it switches from using one feature to another (i.e. when there is no variation along the first feature dimension). This does not mean that observers *cannot* simulate possible future behaviors of objects, such as how the trajectory of a bouncing cube continues, just that they may not choose to do so when simpler yet near-optimal heuristics are available. This interpretation is consistent with previous work by Battaglia and colleagues[20], who found that when a simple heuristic is a more efficient and optimal way to make a prediction (e.g., "How far will the blocks fall when the block tower falls over?", observers tend to use such heuristics (e.g., height of the tower) rather than simulation. Thus, we suggest that observers can draw on different forms of computation, but do so taking into consideration the relative costs and demands of the specific task at hand—an example of *bounded* or *computational rationality*[45,46]. For example, when asked to infer a single parameter (e.g., elasticity) from an observed trajectory, time- and energy-consuming simulations represent a poor allocation of resources when a simple read-out from the feature estimation provides high accuracy. However, visual features are likely too inaccurate when making time- or location-critical predictions about an object's future trajectory[7,47,48]. Under these conditions, the additional costs associated with internal simulation may pay off. Similarly, when no standard heuristics apply, observers may use simulation even for physical inference of material properties, such as mass, as shown by Hamrick et al[21]. Future studies should further investigate the different cognitive strategies humans use under various circumstances as well as the metacognitive process that switches between different strategies.

Conclusion

Visually estimating physical object properties is a crucial, yet computationally challenging task. The visual input is highly ambiguous because an object's behavior depends on numerous entangled factors. Estimating the elasticity of a bouncing object requires disentangling the different causal contributions of elasticity, initial speed, position, and other factors. Using a 'big data' approach, we showed that representing trajectories in terms of their characteristic spatiotemporal features—such as the maximum bounce height or movement duration—yields elasticity estimates that are inexpensive to compute and robust to external factors. Our experiments suggest that the brain estimates elasticity by flexibly switching between a few single-feature heuristics based the information available in the stimuli. Our model explains both the broad successes and the systematic failures of human elasticity perception and correctly predicts a novel illusion in which appearance features maximally diverge from ground truth. Observers can draw on multiple cues and computations, and appear to select strategies with lower computational costs, i.e., computationally rationally. Similar principles might underlie the visual perception of other physical objects properties, such as mass or softness.

Methods

Physical simulations

The dataset was created with the Caronte physics engine of RealFlow 2014 (V.8.1.2.0192; Next Limit Technologies, Madrid Spain), a 3D dynamic simulation software. The dataset contains 100,000 simulations of a cubical object (0.1 m^3) bouncing in a cubical room (1.0 m^3). We chose a cube as the target object because it produces a greater variety of trajectories than, for example, a sphere because the rebound direction depends not only on the direction of the object but also its orientation. We have previously shown that human observers can judge the elasticity of a bouncing cube in such a scene[8]. We varied the cube's elasticity in ten equal steps from 0.0 to 0.9. This value corresponds to the coefficient of restitution—the proportion of energy the cube retains upon collision. We created 10,000 simulations for each level of elasticity by randomly varying its initial velocity, orientation, and position, while keeping all other parameters constant. We simulated 121 frames at 30 fps of the cube moving through the room under gravity. In addition to the original dataset, we simulated another 90,000 trajectories of just one elasticity (0.5). As before, initial velocity, orientation, and position varied randomly. We used the 90,000 simulations + 10,000 simulations of the medium elasticity from the original dataset to search for stimuli in Experiments 2 and 4.

Motion features and multi-feature model

We calculated 28 motion features based on the CoM and the eight corners of the cube for all 100,000 simulations. The end of the cube's movement was defined as the point at which its velocity dropped below 0.003 m/s, since simulated velocity never reaches zero. All other features were computed only for the frames in which the cube was moving. The exact definition of all 28 motion features is described in **Table S1** and **Figures S9-26**. Next, we normalized every motion feature to a range between [0.0, 1.0] and equalized their histograms. We determined the R^2-score, the shared variance with physical elasticity, for each feature and excluded features from further analysis if they explained <5% of the variance. We performed a principal component analysis (PCA) with the remaining 23 features. The resulting scores of the first principal component (PC) were used to predict perceived elasticity. See **SI Results** for details on PCA.

Psychophysical experiments

Participants. Ninety undergraduate students (68 females) from the University of Giessen participated in the experiments (15 in Exp. 1 and 3, 30 in Exp. 2 and 4). Their average age was 24 years (SD = 3.5 years). No person participated in more than one experiment. All participants were naïve with regard to the aims of the study and they gave written informed consent before the experiment. Participants were compensated with 8€/h. The experimental procedure was in accordance with the declaration of Helsinki and approved by the local ethics committee (LEK FB06) at Giessen University.

Stimuli. Experiment 1 contained 15 stimuli per level of elasticity, randomly selected from the original dataset (i.e., 150 stimuli). For Experiment 2 we selected 225 stimuli that systematically decoupled the predictions of each individual feature from both the multi-feature model and physical elasticity. Specifically, for each of the 23 features we chose ten stimuli from the medium elasticity dataset for which the predictions of the individual feature and the multi-feature model were uncorrelated ($|r| < 0.05$), while spanning the widest possible range on both dimensions (**Figure S4**). In Experiments 1 and 2, each stimulus was presented for the full duration of the cube's movement. In Experiments 3 and 4, only the first second of each stimulus was presented (and no stimulus had a movement duration that was shorter than 1 sec). For Experiment 3, we used a random subset of eight stimuli per elasticity level from the stimuli of Experiment 1 (i.e., 80 stimuli). For Experiment 4, we selected 213 stimuli that systematically decoupled the predictions of each individual feature (except movement duration) from both the prediction of the multi-feature model and physical elasticity. The selection procedure was the same as in Experiment 2, but all stimuli were truncated to exactly one second.

The simulations selected as stimuli were rendered using RealFlow's built-in Maxwell renderer. The room was rendered with a white matte material, and the target object was rendered with a blue opaque material. The scene was illuminated brightly using an HDR map through the transparent ceiling.

Set up. All experiments were conducted using the same setup. Stimuli were presented on an Eizo LCD monitor (ColorEdge CG277; resolution: 2560 × 1440 pixels; refresh rate: 60 Hz). Participants used a chin rest to maintain a constant viewing distance of 54 cm. At this distance, the stimuli had a visual angle of 19.6 x 19.6 degrees.

Procedure. All experiments followed the same basic procedure. Participants were instructed to watch a short movie of an object and rate its elasticity. Elasticity was defined to them as the property that distinguishes for example a bouncy ball from a hacky sack. On each trial, one stimulus was presented in a loop until a response was given. Below the movie, a horizontal rating bar was displayed, ranging from 'not elastic' to 'very elastic'. Participants adjusted a slider along the bar to indicate their rating. Each stimulus was repeated three times over the course of the experiment, and all stimuli were presented in random order. Before the main experiment, participants completed ten practice trials, one for each level of elasticity (unknown to participants) to provide an impression of the stimulus range without biasing their response scale. The experimental code was written in Matlab 2018a using Psychtoolbox 3[49–51].

Analysis. In all experiments, we averaged across repetitions to obtain one rating per stimulus from every participant. For Experiments 1 and 3, we calculated the average across participants, as well as inter- and intra-observer variability (standard deviation). We fitted linear regression models to the average elasticity ratings using either physical elasticity, the multi-feature model, or each of the individual feature as predictors. Models were compared using AIC values, specifically their Akaike weights and evidence ratios[52]. Akaike weights represent the probability that a given model is the best among those tested, while evidence ratios indicate the relative likelihood of two competing models given the data. For Experiments 2 and 4, we pooled the data across participants. For each feature's stimulus set, we computed correlations between pooled ratings and both the corresponding feature prediction and the multi-feature model prediction. Pooling (rather than averaging) allowed more reliable estimation of the correlation coefficients based on full trial counts rather than just 10 averages. For each feature, the resulting correlation coefficients were compared using a two-tailed significance test for dependent groups with one overlapping variable[53]. Additionally, we computed the explained variance in perceived elasticity (averaged across participants) for each feature and the model across *all* stimuli, independent of the stimulus set.

Acknowledgments

The authors thank Britta Fritz, Saskia Honnefeller, Judith Kanehl, and Jasmin Kleis for their help with data collection. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; Project 222641018–SFB/TRR 135 TP C1 and Project PA 3723/1-1) and by the European Research Council (ERC) Advanced Grant "STUFF" (Project 101098225).

References

1. Paulun, V. C., Gegenfurtner, K. R., Goodale, M. A. & Fleming, R. W. Effects of material properties and object orientation on precision grip kinematics. *Exp Brain Res* **234**, 2253–2265 (2016).

2. Klein, L. K., Maiello, G., Paulun, V. C. & Fleming, R. W. Predicting precision grip grasp locations on three-dimensional objects. *PLOS Computational Biology* **16**, e1008081 (2020).

3. Weir, P. L., MacKenzie ,Christine L., Marteniuk ,Ronald G., Cargoe ,Sherri L. & and Frazer, M. B. The Effects of Object Weight on the Kinematics of Prehension. *Journal of Motor Behavior* **23**, 192–204 (1991).

4. Weir, P. L., MacKenzie, C. L., Marteniuk, R. G. & Cargoe, S. L. Is Object Texture a Constraint on Human Prehension!: Kinematic Evidence. *Journal of Motor Behavior* **23**, 205–210 (1991).

5. Glowania, C., van Dam, L. C. J., Brenner, E. & Plaisier, M. A. Smooth at one end and rough at the other: influence of object texture on grasping behaviour. *Exp Brain Res* **235**, 2821–2827 (2017).

6. Fikes, T. G., Klatzky, R. L. & Lederman, S. J. Effects of Object Texture on Precontact Movement Time in Human Prehension. *Journal of Motor Behavior* **26**, 325–332 (1994).
7. Diaz, G., Cooper, J., Rothkopf, C. & Hayhoe, M. Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision* **13**, 20 (2013).
8. Paulun, V. C. & Fleming, R. W. Visually inferring elasticity from the motion trajectory of bouncing cubes. *Journal of Vision* **20**, 6 (2020).
9. Paulun, V. C., Kawabe, T., Nishida, S. & Fleming, R. W. Seeing liquids from static snapshots. *Vision Research* **115**, 163–174 (2015).
10. Paulun, V. C., Schmidt, F., van Assen, J. J. R. & Fleming, R. W. Shape, motion, and optical cues to stiffness of elastic objects. *Journal of Vision* **17**, 20 (2017).
11. Schmidt, F., Paulun, V. C., Van Assen, J. J. R. & Fleming, R. W. Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues. *Journal of Vision* **17**, 18 (2017).
12. Schmid, A. C. & Doerschner, K. Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *Journal of Vision* **18**, 14 (2018).
13. Van Assen, J. J. R., Barla, P. & Fleming, R. W. Visual Features in the Perception of Liquids. *Current Biology* **28**, 452-458.e4 (2018).
14. Kawabe, T., Maruya, K., Fleming, R. W. & Nishida, S. Seeing liquids from visual motion. *Vision Research* **109**, 125–138 (2015).
15. Bi, W., Jin, P., Nienborg, H. & Xiao, B. Manipulating patterns of dynamic deformation elicits the impression of cloth with varying stiffness. *Journal of Vision* **19**, 18 (2019).
16. Bi, W., Shah, A. D., Wong, K. W., Scholl, B. & Yildirim, I. Perception of soft materials relies on physics-based object representations: Behavioral and computational evidence. *bioRxiv* 2021–05 (2021).
17. Bi, W. & Xiao, B. Perceptual constancy of mechanical properties of cloth under variation of external forces. in *Proceedings of the ACM Symposium on Applied Perception* 19–23 (ACM, Anaheim California, 2016). doi:10.1145/2931002.2931016.

18. Aliaga, C., O'Sullivan, C., Gutierrez, D. & Tamstorf, R. Sackcloth or silk?: the impact of appearance vs dynamics on the perception of animated cloth. in *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception* 41–46 (ACM, Tübingen Germany, 2015). doi:10.1145/2804408.2804412.

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

20. Battaglia, P. W., Hamrick, J. B. & Tenenbaum, J. B. Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18327–18332 (2013).

21. Hamrick, J. B., Battaglia, P. W., Griffiths, T. L. & Tenenbaum, J. B. Inferring mass in complex scenes by mental simulation. *Cognition* **157**, 61–76 (2016).

22. Yildirim, I., Smith, K. A., Belledonne, M. E., Wu, J. & Tenenbaum, J. B. Neurocomputational modeling of human physical scene understanding. (2018).

23. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems* vol. 25 (Curran Associates, Inc., 2012).

24. Szegedy, C. *et al.* Going deeper with convolutions. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1–9 (IEEE, Boston, MA, USA, 2015). doi:10.1109/cvpr.2015.7298594.

25. Kirillov, A. *et al.* Segment Anything. Preprint at https://doi.org/10.48550/arXiv.2304.02643 (2023).

26. Tung, H.-Y. *et al.* Physion++: Evaluating Physical Scene Understanding that Requires Online Inference of Different Physical Properties.

27. Motamed, S., Culp, L., Swersky, K., Jaini, P. & Geirhos, R. Do generative video models understand physical principles? Preprint at https://doi.org/10.48550/arXiv.2501.09038 (2025).

28. Warren, W. H., Kim, E. E. & Husney, R. The Way the Ball Bounces: Visual and Auditory Perception of Elasticity and Control of the Bounce Pass. *Perception* **16**, 309–336 (1987).

29. Nusseck, M., Lagarde, J., Bardy, B., Fleming, R. & Bülthoff, H. H. Perception and prediction of simple object interactions. in *Proceedings of the 4th symposium on Applied perception in graphics and visualization* 27–34 (ACM, Tubingen Germany, 2007). doi:10.1145/1272582.1272587.

30. Kawabe, T. & Nishida, S. Seeing jelly: judging elasticity of a transparent object. in *Proceedings of the ACM Symposium on Applied Perception* 121–128 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2931002.2931008.

31. Fleming, R. W. Visual perception of materials and their properties. *Vision Research* **94**, 62–75 (2014).

32. Fleming, R. W. & Storrs, K. R. Learning to see stuff. *Curr Opin Behav Sci* **30**, 100–108 (2019).

33. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).

34. *Perception of Space and Motion*. (Elsevier, 1995). doi:10.1016/b978-0-12-240530-3.x5000-7.

35. Eagleman, D. M. *et al.* Time and the Brain: How Subjective Time Relates to Neural Time. *J. Neurosci.* **25**, 10369–10371 (2005).

36. Eagleman, D. M. Human time perception and its illusions. *Curr Opin Neurobiol* **18**, 131–136 (2008).

37. Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* **6**, 755–765 (2005).

38. Scholl, B. J. & Tremoulet, P. D. Perceptual causality and animacy. *Trends Cogn Sci* **4**, 299–309 (2000).

39. Morgenstern, Y. & Kersten, D. J. The perceptual dimensions of natural dynamic flow. *Journal of Vision* **17**, 7 (2017).

40. Murdison, T. S., Leclercq, G., Lefèvre, P. & Blohm, G. Misperception of motion in depth originates from an incomplete transformation of retinal signals. *Journal of Vision* **19**, 21 (2019).

41. Welchman, A. E., Lam, J. M. & Bülthoff, H. H. Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12087–12092 (2008).

42. Wu, J., Yildirim, I., Lim, J. J., Freeman, B. & Tenenbaum, J. Galileo: Perceiving physical object properties by integrating a physics engine with deep learning. *Advances in neural information processing systems* **28**, (2015).

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

44. Kubricht, J. R., Holyoak, K. J. & Lu, H. Intuitive Physics: Current Research and Controversies. *Trends in Cognitive Sciences* **21**, 749–759 (2017).

45. Gigerenzer, G. & Todd, P. M. *Simple Heuristics That Make Us Smart*. (Oxford University Press, New York, 2001).

46. Gershman, S. J., Horvitz, E. J. & Tenenbaum, J. B. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science* **349**, 273–278 (2015).

47. Mann, D. L., Nakamoto, H., Logt, N., Sikkink, L. & Brenner, E. Predictive eye movements when hitting a bouncing ball. *Journal of Vision* **19**, 28 (2019).

48. Mrowca, D. *et al.* Flexible neural representation for physics prediction. in *Advances in Neural Information Processing Systems* vol. 31 (Curran Associates, Inc., 2018).

49. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* **10**, 437–442 (1997).

50. Brainard, D. H. The Psychophysics Toolbox. *Spat Vis* **10**, 433–436 (1997).

51. Kleiner, M. *et al.* What's new in psychtoolbox-3. *Perception* **36**, 1–16 (2007).

52. Burnham, K. P. & Anderson, D. R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* **33**, 261–304 (2004).

53. Olkin, I. Correlations revisited. in *Improving Experimental Design and Statistical Analysis* (ed. Stanley, J. C.) 102–128 (Rand McNally, 1967).