bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Efficient Visual Heuristics 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, USA

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

[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

Vision is more than object recognition: In order to interact with the physical world, we estimate object properties such as mass, fragility, or elasticity by sight. The computational basis of this ability is poorly understood. Here, we propose a model based on the statistical appearance of objects, i.e., how they typically move, flow, or fold. We test this idea using a particularly challenging example: estimating the elasticity of bouncing objects. Their complex movements depend on many factors, e.g., elasticity, initial speed, and direction, and thus every object can produce an infinite number of different trajectories. By simulating and analyzing the trajectories of 100k bouncing cubes, we identified and evaluated 23 motion features that could individually or in combination be used to estimate elasticity. Experimentally teasing apart these competing but highly correlated hypotheses, we found that humans represent bouncing objects in terms of several different motion features but rely on just a single one when asked to estimate elasticity. Which feature this is, is determined by the stimulus itself: Humans rely on the duration of motion if the complete trajectory is visible, but on the maximal bounce height if the motion duration is artificially cut short. Our results suggest that observers take into account the computational costs when asked to judge elasticity and thus rely on a robust and efficient heuristic. Our study provides evidence for how such a heuristic can be derived—in an unsupervised manner—from observing the natural variations in many exemplars.

Keywords: Visual Perception, Intuitive Physics, Computational Rationality

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Significance Statement

How do we perceive the physical properties of objects? Our findings suggest that when tasked with reporting the elasticity of bouncing cubes, observers rely on simple heuristics. Although there are many potential visual cues, surprisingly, humans tend to switch between just a handful of them depending on the characteristics of the stimulus. The heuristics predict not only the broad successes of human elasticity perception but also the striking pattern of errors observers make when we decouple the cues from ground truth. Using a big data approach, we show how the brain could derive such heuristics by observation alone. The findings are likely an example of 'computational rationality', in which the brain trades off task demands and relative computational costs.

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 understand the cues and computations the brain relies on to estimate the internal properties of objects(8–22). Inferring the mass or elasticity of an object by observation means going beyond simple pattern recognition. Recognizing a familiar object or material activates prior knowledge about its typical properties and behavior. Thus, we would be surprised to pick up a piece of Styrofoam as heavy as a brick (23, 24) or throw a hacky sack that bounces back like a tennis ball.

However, the human mind is not limited to learned world knowledge about specific attributes of specific objects. We can also estimate the physical properties of objects and materials by observing how they move and interact (8, 9, 12–14, 17). This is an integral ability of an intelligent and robust visual system and one of the major challenges for state-of-the-art artificial vision systems (25–29). In previous work (8), we have shown, that observers estimate the elasticity of unfamiliar bouncing objects based on their motion trajectory. Computationally, this ability is not as trivial as it may introspectively appear: A single bouncing object can produce an infinite number of trajectories, 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**). If there is no direct mapping between an object's elasticity and its trajectory, how does the brain estimate the former from the latter?

While an object's exact motion trajectory is determined by multiple factors (only one of which is elasticity), variations across different trajectories are not random. For example, objects of higher elasticity tend to bounce higher and for a longer time before they come to rest. The brain could exploit such statistical regularities in the visual features of bouncing objects that arise from the underlying physical constraints. Identifying salient visual features that vary in lawful ways between different objects could result in a mental model of the typical appearance of elastic objects. Such '*statistical appearance model'* (30) could be learned in a completely data-driven fashion by observing the variation of salient features across sufficient examples.

Here, we test this hypothesis using a 'big data' approach. We simulated 100,000 short (4 sec) trajectories of a bouncing cube in a room (**Figure 1A**). The cube's elasticity (coefficient of

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 diversity of trajectories to identify and evaluate statistical regularities.

We used this dataset to measure human relative elasticity judgments and to systematically explore the statistical regularities in the behavior of elastic objects. Specifically, we first identified 28 candidate 3D motion features (**Figure 2A-D, Table 1**) based on the physics of bouncing objects, and previously proposed cues (15, 16). We then determined how they statistically relate to elasticity in our dataset. While no feature is a perfect estimator of the physics, some are remarkably good (explained variance > 80 %). We found that the best of the features we considered—movement duration—predicts human elasticity judgments on a stimulus-by-stimulus basis better than ground truth elasticity does (**Figure 2E**). In a series of carefully designed experiments, we leverage the big-data approach to select stimuli that systematically decouple several competing and highly correlated alternative hypotheses. We conclude that when asked to judge the elasticity of bouncing objects, observers rely on just a single motion feature at a time, whereby the feature they use is determined by the properties of the stimulus—a robust and computationally efficient strategy.

Results

Experiment 1: Relative elasticity judgments

To characterize human perception of relative elasticity, fifteen observers rated the apparent elasticity of a bouncing cube 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, 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). That is, constancy was not perfect. Importantly, the pattern of errors was not random but highly consistent between different observers (*r* = .91 ± .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).

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

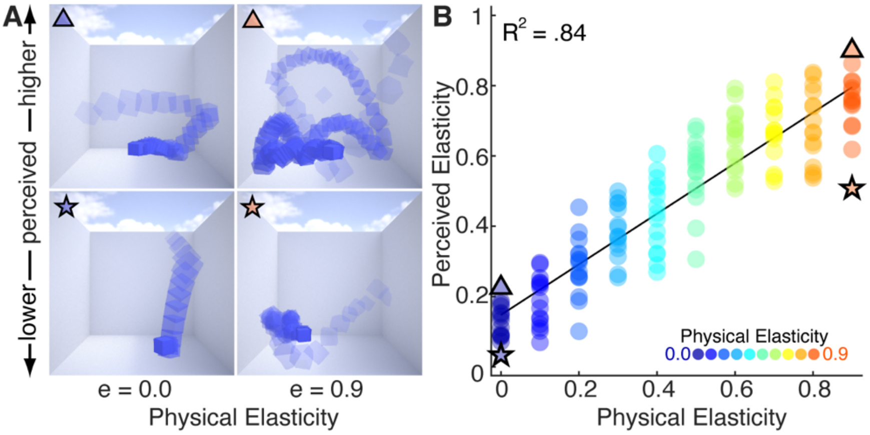

Figure 1. Stimuli and results of Experiment 1. ***A)*** *Example stimuli of lowest (e = 0.0) and highest elasticity (e = 0.9), frames of the animations were overlaid for illustration purposes. Even though both images in each row show the same cube (i.e., the same physical properties), the trajectories 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.*

Visual features of elasticity

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 4D 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 (31, 32), 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 simple, ideal settings (e.g., bounce height ratio). Others capture summary statistics that integrate over time and might be useful in imperfect but more realistic scenes (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. By defining a large number of features, we aimed to achieve a comprehensive characterization of the trajectories and constrain our hypothesis space based on the data rather than a priori assumptions. 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). We computed all motion features from the trajectory of the cube's center of mass (CoM) and eight corners for all 100,000 simulations (see **Methods**). Thus, all features are computed from observable quantities, i.e., positions and changes of positions over time, but vary in their computational complexity.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 whole 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 every wall. (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 & G, Table 1**). The best features (i.e., those that share the most variation with physical elasticity) are those that integrate over time, such as movement duration or the number of bounces. Physically based measures of the coefficient of restitution in idealized settings (related to the ratio of bounce heights, bounce durations, velocities, or energy) are not among the best features, underlining the complexity of bounce trajectories of cubical objects. We narrowed our hypothesis space by excluding features that explained < 5 % of the variance from further analysis (greyed items in Table 1).

Strikingly, we found that the best feature for predicting *physical* elasticity—movement duration—was also the best to predict *perceived* elasticity (R^2 = .91, F(1, 148) = 1515.1, p < .001, **Figure 2E&G**). On a stimulus-by-stimulus basis, movement duration was a better predictor of human ratings than physical elasticity (evidence ratio: $w_{\text{movDur}}/w_{\text{Physics}}$ = 1.51e+20). This suggests that when asked to judge elasticity, human observers actually rely on the main observable covariate of elasticity and judge how long objects keep moving.

However, our results also show that other features explained the behavioral data similarly well (**Figure 2G**) and that the 23 features were highly correlated with one another across the set of 100,000 trajectories (mean absolute correlation, M = 0.48; see **Figure S1**). This has two important implications. First, it means that our results could in principle be explained by

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

features other than movement duration or by a combination of several or all features. Second, because of the high multicollinearity of features, it is challenging to isolate and test their effect on perception individually.

To overcome these limitations and 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 a space of the first two PCs reveals that elasticity varies largely along the first dimension (**Figure 2F**). We found that true elasticity and the first PC share 82.83% of their variance. Thus, 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 elasticity, it is mostly independent of the other latent parameters we used to initialize our simulations (e.g., velocity; all < 1.0%, **Figure S2B**). In other words, this linear combination of motion features (see **Figure S4** for feature loadings) successfully disentangles 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 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. Importantly, applying a PCA to the raw motion trajectories (**Figure S3**) does not give rise to elasticity estimates in a comparable manner—highlighting the crucial role of appearance features in the process.

Rather than relying on the single most salient feature, the brain might represent bouncing objects in a multidimensional feature space. Albeit computationally more complex, a weighted feature combination can be expected to be more robust across scene variations. Since PC1 is the statistically optimal data-driven weighted combination of features, we interpret PC1 as the 'predicted elasticity' of a multi-feature model, i.e., a competing hypothesis to *movement duration*. Indeed, we find that 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 2G-H**). It predicts perception better than the ground truth does (evidence ratio: $w_{\text{FeatureModel}}/w_{\text{Physics}} = 3.38\text{e}+13$), but worse than movement duration (evidence ratio: $w_{\text{movDur}}/w_{\text{FeatureModel}} = 4.46\text{e}+06$; $w_{\text{movDur}} \approx 1$). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

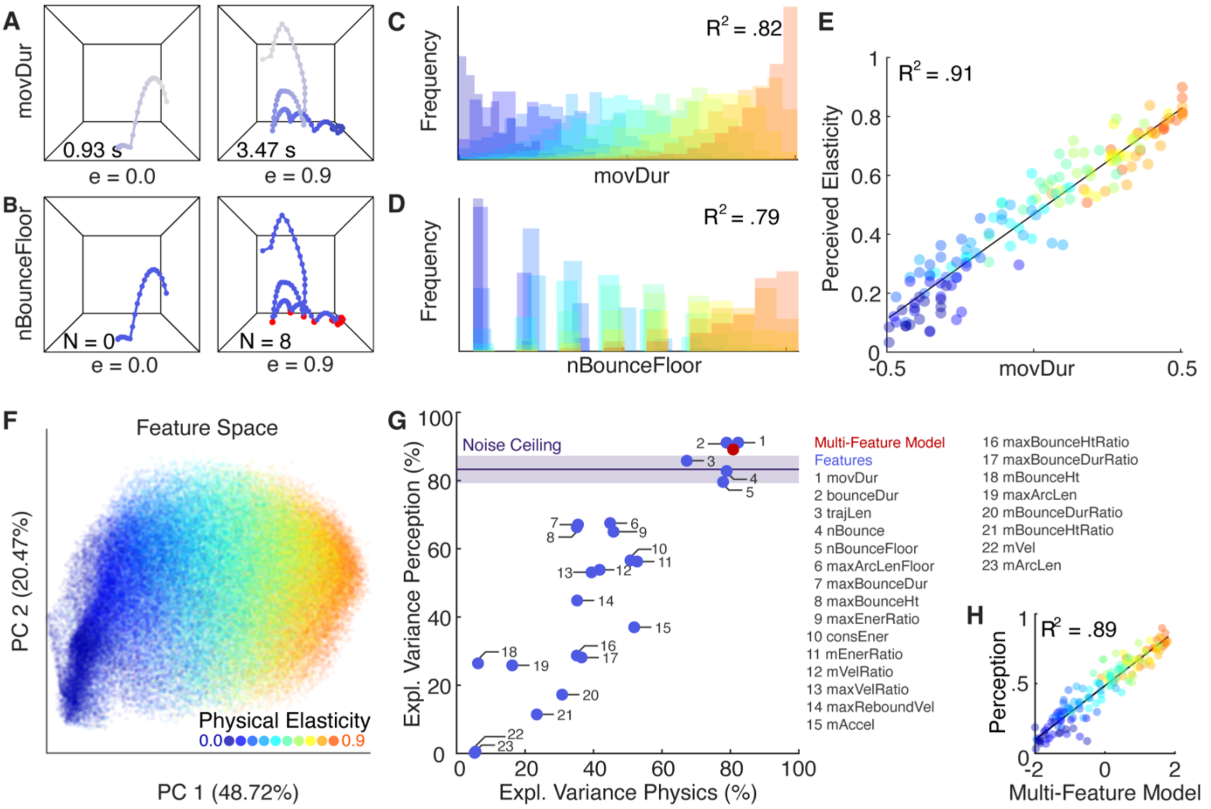

Figure 2. ***A)*** *Example trajectory of a low (e = 0.0) and high (e = 0.9) elastic cube, each dot represents one frame; color gradient represents movement duration.* ***B)*** *The same two trajectories, red dots represent bounces off the floor.* ***C)*** *Distribution of movement durations in the set 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)*** *Perceived elasticity from Experiment 1 as a function of the prediction made by the statistically optimal feature: movement duration.* ***F)*** *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.* ***G)*** *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).* ***H)*** *Rated elasticity from Experiment 1 as a function of the prediction made by the feature combination from PC1, i.e., the multi-feature model.*

Experiment 2: Decoupling Single- vs. Multi-Feature Models

The aim of Experiment 2 was threefold: First, we systematically decoupled the predictions of the multi-feature model from those of movement duration 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, decoupling each feature from the weighted combination of all features is the only way to test the causal contribution of each individual feature to elasticity perception. In doing so, we are able to overcome the purely correlational analysis reported so

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

far and experimentally test 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 features) for which feature and multi-feature model predictions were essentially uncorrelated ($|r| < .05$; see **Methods** and **Figure S5** for more details). A new group of 30 participants judged the elasticity of these stimuli. Note that this stimulus selection processes 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.

Despite these potential drawbacks, Experiment 2 provided clear results. For each feature, **Figure 3A** 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 S6B** shows the partial correlations between perception and feature vs. perception and 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 3B and S6A**). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

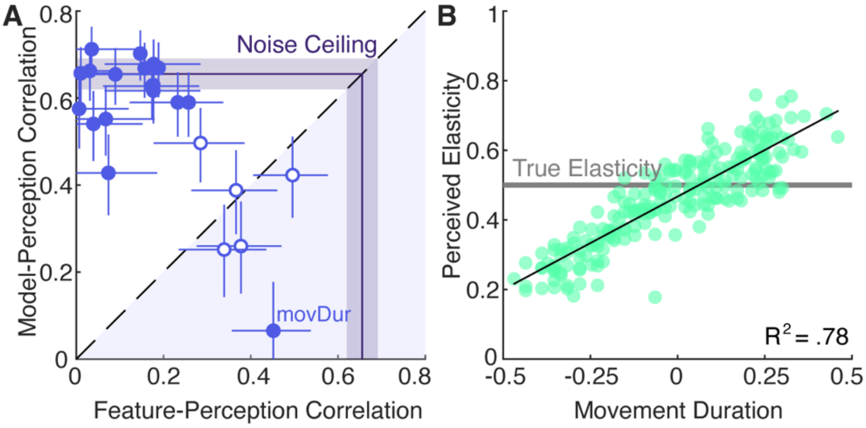

Figure 3. Results of the decorrelation experiment. ***A)*** *Correlation of the pooled perceptual ratings with the multi-feature model (y-axis) and the individual features (x-axis). Each dot represents the correlations for one set of stimuli that were specifically selected to decouple the prediction of one feature from the model. Features that fall below the diagonal (light blue shaded area) exceed the model, i.e., they correlate more strongly with the predictions of an individual feature and vice versa. 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.*

Experiments 3 & 4: Estimating elasticity in truncated videos

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 confirm this observation, 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.** Thus, the observable movement duration was the same for all stimuli. 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 S7A**) and showed a near-perfect correlation ($r = .97$, $p < .001$) with ratings for the full movies (Exp. 1; see **Figure 4A**).

Experiment 4 was conducted to systematically test which features the brain relies on to achieve this performance. For this purpose, we followed 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 was uncorrelated with the prediction of the multi-feature model. A new group of 30 observers estimated elasticity in these 1-sec stimuli. Across all 100,000 simulations (of varying elasticity), none of the 1-sec

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

features perfectly predicts movement duration in the full videos (see **Figure S8A**). Yet, we found that one of the best features—maximum bounce height—showed a significantly higher correlation with perception than the multi-feature model (r = .54 > r = .25, p < .0023, **see Figure 4C**) 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 4D** and **S9A**). 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 S9C**). Therefore, bounce duration was only a diagnostic feature when it was notably shorter than one second. For most (i.e., 12) features, we found that the multi-feature model predicts 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 S9B**). More precisely, the correlations between perception and multi-feature model (r = .49 ± .16 (M ± SD)) decreased significantly when controlling for maximum bounce height (r = .23 ± .09; t(21) = 6.65, p < .001).

Taken together, 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 S8B**). Taken together, this suggests that early in a bouncing object's trajectory we first form an impression of its elasticity based on the largest of the bounces that it makes, and if we get to see the object continue to move until it comes to a standstill, we instead rely on the movement duration.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

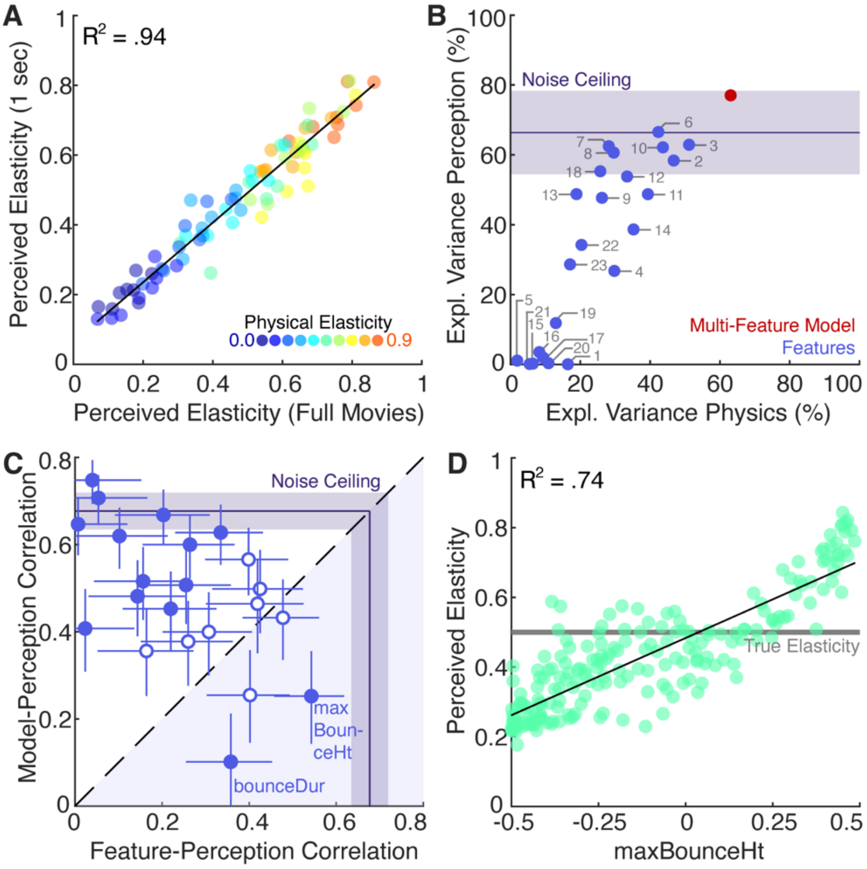

***Figure 4**. Results of Experiments 3 and 4 with truncated movies. **A**) Average perceived elasticity in 1-sec movie clips (Exp. 3) as a function of the movement duration of the apparent elasticity in full movies of the same stimuli. Physical elasticity is color-coded. **B**) Explained variance in 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., they correlate more strongly with the predictions of an individual feature and vice versa. 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).*

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Here we propose that humans represent the physical properties of objects and materials 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, observers instead rely on the maximal bounce height to judge elasticity. Taken together, 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, 33), it overcomes three critical limitations of previous work. First, we assess the statistical relations between a number 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 hypothesized that by observing the outside world and its inherent statistical relations (30, 34), 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 us to test the contribution of a large, yet testable number of different (interpretable) motion features and their combination. It would be interesting to test whether similar features emerge within an (unsupervised) deep learning framework. 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 duration of motion or the maximal bounce height in case the duration is cut short. 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 (35) 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 ones we have tested. Duration and spatial distance are quantities the visual system can estimate reliably and accurately (36–39).

Even though we found strong evidence that humans base their elasticity estimates on mainly 2-3 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 (40), fluid flow (13, 14, 41)). 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 4D 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 (42, 43), 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, more elastic objects would still tend to move longer and bounce higher. Participants presumably had little experience with bouncing 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, 29) or cheaper but less accurate heuristics (9, 12, 44, 45). 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. 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* (46, 47). 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, 48, 49). Under these conditions, the additional costs associated with internal simulation may pay off. 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 have shown how representing motion trajectories in terms of their characteristic spatiotemporal features—such as the maximum bounce height or movement duration—gives rise to elasticity estimates that are robust to the influence of external determinants. A series of experiments suggest that the brain flexibly estimates elasticity by switching between a small number of individual features based on the characteristics of the stimuli. Our model explains both the broad successes and the specific failures of human elasticity perception and correctly predicts a novel illusion for which appearance features and ground truth maximally deviate. Observers can draw on multiple cues and computations, but they select their strategies taking into consideration the relative costs of different approaches, i.e., computationally rational.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 bouncing in a cubical room. We chose a cube as the target object because it creates a larger variety of trajectories than for example spheres 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 a scene like that (8). We varied the elasticity of the cube in ten equal steps from 0.0 to 0.9. This number refers to the amount of energy that the cube retains when it collides (coefficient of restitution). We created 10,000 simulations for each level of elasticity by randomly varying its initial velocity, orientation, and position. All other parameters were 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 in all 100,000 simulations of varying elasticity. We first determined the end of the cube's motion using a velocity threshold of 0.003 m/s (because velocity never truly reaches zero). All other features were calculated 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 S10-27**. 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 that were chosen randomly 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 the prediction of the feature model as well as from ground truth elasticity. More specifically, for each of the 23 features we chose ten stimuli from the medium elasticity dataset for which the prediction of the individual feature and the multi-feature model are uncorrelated ($|r| < 0.05$) but span a

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

range that was as large as possible in both, see **Figure S5**. In Experiments 1 and 2, each stimulus was presented for the whole duration of the movement. In Experiments 3 and 4, only the first second of each stimulus was shown to participants (and no stimulus had a movement duration that was < 1 sec). For Experiment 3, we chose a random subset of eight stimuli per elasticity level from the stimuli used in Experiment 1 (i.e., 80 stimuli). For Experiment 4, we chose 213 stimuli that systematically decouple the predictions of each individual feature (except movDur) from the prediction of the multi-feature model as well as from ground truth elasticity. The selection procedure was the same as in Experiment 2, but all stimuli were cropped to exactly one second.

Simulations that we selected as stimuli were rendered using RealFlow's built-in Maxwell renderer. The room was rendered with a white matter material; the target was blue opaque material. The scene was illuminated brightly with an HDR map through the transparent ceiling.

Set up. All experiments were conducted with the same setup. Stimuli were presented on an Eizo LCD monitor (model ColorEdge CG277; resolution 2560 × 1440 pixels; refresh rate 60Hz). Participants placed their chin on a chin rest to keep a constant distance of 54 cm between their eyes and the monitor. The stimuli covered about 19.6 degrees of visual angle.

Procedure. All experiments followed the same basic procedure. Participants were instructed that on every trial, they would see a movie of an object and they would have to rate the object's elasticity. Elasticity was defined to them as the property that distinguishes for example a bouncy ball and a hacky sack from one another. On each trial, one stimulus was presented in a loop until a response was given. Below the movie, there was a horizontal rating bar 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. 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 and not bias their response scale. The experimental code was written in Matlab 2018a using Psychtoolbox 3 (50–52).

Analysis. In all experiments, we averaged across repetitions to gain one rating for every stimulus from every participant. For Experiments 1 and 3, we calculated the average (and SD for Exp. 1) across participants as well as the inter- and intra-observer variability. We fitted linear regression models to the average elasticity ratings using either the physical elasticity as a predictor, the normative model, or the best individual feature. We compared the models by evaluating their AIC values, more specifically their Akaike weights and evidence ratios (53). Akaike weights can be interpreted as the probability that the given model is the best model, whereas the evidence ratio of pairs of models indicates the relative likelihood of both models given the data. For Experiments 2 and 4, we pooled the data across participants. For each stimulus set (one for each feature), we calculated the correlation between pooled ratings and the prediction of that feature and between the ratings and the multi-feature model prediction. The data were pooled instead of averaged to get a more reliable estimate of the correlation coefficients (relying on more than just 10 average values). For each feature, we compared the resulting correlation coefficients with a two-tailed significance test for dependent groups with one overlapping variable (54). Furthermore, we calculated the

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

explained variance in terms of the perceived elasticity (averaged across participants) for each feature and the model across *all* stimuli (independent of the stimulus set).

Author Contributions

All authors conceived and designed the study and wrote the manuscript. VCP, FSB, and RWF 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.

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) Consolidator Award “SHAPE” (Project ERC-CoG-2015-682859).

References

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

2. L. K. Klein, G. Maiello, V. C. Paulun, R. W. Fleming, Predicting precision grip grasp locations on three-dimensional objects. *PLoS Comput Biol* **16** (2020).

3. P. L. Weir, C. L. Mac Kenzie, R. G. Marteniuk, S. L. Frazer, Is object texture a constraint on human prehension!: Kinematic evidence. *J Mot Behav* **23**, 205–210 (1991).

4. P. L. Weir, C. L. MacKenzie, R. G. Marteniuk, S. L. Cargoe, M. B. Frazer, The Effects of Object Weight on the Kinematics of Prehension. *J Mot Behav* **23**, 192–204 (1991).

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

6. T. G. Fikes, R. L. Klatzky, S. J. Lederman, Effects of Object Texture on Precontact Movement Time in Human Prehension. *J Mot Behav* **26**, 325–332 (1994).

7. G. Diaz, J. Cooper, C. Rothkopf, M. Hayhoe, Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *J Vis* **13**, 20–20 (2013).

8. V. C. Paulun, R. W. Fleming, Visually inferring elasticity from the motion trajectory of bouncing cubes. *J Vis* **20** (2020).

9. V. C. Paulun, T. Kawabe, S. Nishida, R. W. Fleming, Seeing liquids from static snapshots. *Vision Res* **115**, 163–174 (2015).

10. F. Schmidt, V. C. Paulun, J. J. R. van Assen, R. W. Fleming, Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues. *J Vis* **17** (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. A. C. Schmid, K. Doerschner, Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *J Vis* **18** (2018).

12. V. C. Paulun, F. Schmidt, J. J. R. van Assen, R. W. Fleming, Shape, motion, and optical cues to stiffness of elastic objects. *J Vis* **17** (2017).

13. J. J. R. van Assen, P. Barla, R. W. Fleming, Visual Features in the Perception of Liquids. *Current Biology* **28**, 452-458.e4 (2018).

14. T. Kawabe, K. Maruya, R. W. Fleming, S. Nishida, Seeing liquids from visual motion. *Vision Res* **109**, 125–138 (2015).

15. W. Bi, P. Jin, H. Nienborg, B. Xiao, Manipulating patterns of dynamic deformation elicits the impression of cloth with varying stiffness. *J Vis* **19**, 1–18 (2019).

16. W. Bi, A. D. Shah, K. W. Wong, B. Scholl, I. Yildirim, "Perception of soft materials relies on physics-based object representations: Behavioral and computational evidence" (2021) https:/doi.org/https://doi.org/10.1101/2021.05.12.443806.

17. W. Bi, B. Xiao, Perceptual constancy of mechanical properties of cloth under variation of external forces in *Proceedings of the ACM Symposium on Applied Perception, SAP 2016*, (Association for Computing Machinery, Inc, 2016), pp. 19–23.

18. C. Aliaga, C. O'sullivan, D. Gutierrez, R. Tamstorf, 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*, (2015), pp. 41–46.

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

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

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

22. I. Yildirim, K. A. Smith, M. Belledonne, J. Wu, J. B. Tenenbaum, Neurocomputational Modeling of Human Physical Scene Understanding Indicates equal contribution in *CCN*, (2018).

23. G. Buckingham, J. S. Cant, M. A. Goodale, Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight. *J Neurophysiol* **102**, 3111–3118 (2009).

24. V. C. Paulun, G. Buckingham, M. A. Goodale, R. W. Fleming, The material-weight illusion disappears or inverts in objects made of two materials. *J Neurophysiol* **121**, 996–1010 (2019).

25. M. Denil, *et al.*, Learning to Perform Physics Experiments via Deep Reinforcement Learning. *arXiv preprint arXiv* **1611.01843.** (2016).

26. J. Wu, E. Lu, P. Kohli, W. T. Freeman, J. B. Tenenbaum, Learning to See Physics via Visual De-animation. *Adv Neural Inf Process Syst* **30** (2017).

27. D. Zheng, V. Luo, J. Wu, J. B. Tenenbaum, Unsupervised Learning of Latent Physical Properties Using Perception-Prediction Networks. *arXiv preprint arXiv* (2018) https://doi.org/10.48550/arXiv.1807.09244.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

28. P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, K. Kavukcuoglu, Interaction Networks for Learning about Objects, Relations and Physics. *Adv Neural Inf Process Syst* **29** (2016).

29. J. Wu, I. Yildirim, J. J. Lim, W. T. Freeman, J. B. T. Bcs, Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning. *Adv Neural Inf Process Syst* **28** (2015).

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

31. M. Nusseck, J. Lagarde, B. Bardy, R. Fleming, H. H. Bülthoff, Perception and prediction of simple object interactions in *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, (ACM, 2007), pp. 27–34.

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

33. T. Kawabe, S. Nishida, Seeing jelly in *Proceedings of the ACM Symposium on Applied Perception*, (ACM, 2016), pp. 121–128.

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

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

36. , *Perception of Space and Motion* (Elsevier, 1995) https:/doi.org/10.1016/B978-0-12-240530-3.X5000-7.

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

38. D. M. Eagleman, *et al.*, Time and the Brain: How Subjective Time Relates to Neural Time. *The Journal of Neuroscience* **25**, 10369–10371 (2005).

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

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

41. Y. Morgenstern, D. J. Kersten, The perceptual dimensions of natural dynamic flow. *J Vis* **17**, 7 (2017).

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

43. A. E. Welchman, J. M. Lam, H. H. Bülthoff, Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences* **105**, 12087–12092 (2008).

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

45. E. Ludwin-Peery, N. R. Bramley, E. Davis, T. M. Gureckis, Limits on simulation approaches in intuitive physics. *Cogn Psychol* **127**, 101396 (2021).

46. S. J. Gershman, E. J. Horvitz, J. B. Tenenbaum, Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science (1979)* **349**, 273–278 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.24.534031; this version posted March 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. G. Gigerenzer, P. M. Todd, *Simple Heuristics that Make Us Smart* (Oxford University Press, 2001).

48. D. Mrowca, *et al.*, Flexible Neural Representation for Physics Prediction. *Adv Neural Inf Process Syst* **31** (2018).

49. D. L. Mann, H. Nakamoto, N. Logt, L. Sikkink, E. Brenner, Predictive eye movements when hitting a bouncing ball. *J Vis* **19**, 28 (2019).

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

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

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

53. K. P. Burnham, D. R. Anderson, Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* **33**, 261–304 (2004).

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