Visual perception of surprising materials in dynamic scenes

Authors:

Lorilei M. Alley*[1], Alexandra C. Schmid[*+1], Katja Doerschner[1,2]

[1] Justus Liebig University Giessen

[2] Bilkent University

[*] Equal contribution

[+] Lead contact and corresponding author

Visual perception of surprising materials in dynamic scenes

Abstract

Many objects that we encounter have ‘typical’ material qualities: spoons are hard, pillows are soft and jelly dessert is wobbly. Over a life time of experiences, strong associations between an object and its typical material properties may be formed, and these associations not only include how glossy, rough or pink an object is but also how it behaves under force: we expect knocked over vases to shatter, popped bike tires to deflate, and gooey grilled cheese to hang between two slices of bread when pulled apart. Here we ask how such rich visual priors affect the visual perception of material qualities and present a particularly striking example of expectation violation. In a cue conflict design, we pair computer-rendered familiar objects with surprising material behaviors (a linen curtain shattering, a porcelain teacup wrinkling, etc.) and find that material qualities are not solely estimated from the object’s kinematics (i.e. its physical (atypical) motion while shattering, wrinkling, wobbling etc.); rather, material appearance is sometimes “pulled” towards the “native” motion, shape, and optical properties that are associated with this object. Our results, in addition to patterns we find in reaction time data, suggest that visual priors about materials can set up high-level expectations about complex future states of an object and show how these priors modulate material appearance. Understanding how high-level expectations are integrated with incoming sensory evidence is an essential step towards understanding how the human visual system accomplishes material perception ([1], [2], [3], [4], [5], [6]).

Results & Discussion

Visual perception is not a one-way (bottom-up) road; how we process visual input is influenced by expectations about the sensory environment, which develop from our previous experience and learning about existing regularities in the world, i.e. associating things or events that co-occur. Expectations have been shown to facilitate visual processing in the case of priming, to modulate the frequency of a particular percept in bi-stable stimuli, and to change our interpretation of ambiguous stimuli (see [7], or [8] for a review). However, the stimuli used in these experiments have been fairly simple (static images of objects), and it has been shown that learning associations can also include fairly complex phenomena. For example, recently, [9] (also [10], [11]), showed that humans can learn to predict how different liquids flow around solid obstacles (also see other examples for predicting of motion trajectories of rigid objects, e.g. ([12], or [13], or [14]). While the authors attributed human performance to an ability to “reason” about fluid dynamics, here we explicitly test whether existing *perceptual expectations* about material properties can set up complex predictions about future states, and whether – and to what extent – these expectations influence material appearance.

The role of predictions or associative mechanisms in material perception is not well understood [15] [16], [17]]. When perceiving qualities like how hard, crumbly, wobbly, gelatinous, or heavy a material is, we may integrate dynamic and pictorial cues through stored *prior associations* with familiar objects; for example a glossy, translucent wobbling cube might resemble Jell-O, whose associated high wobbliness may bias the percept towards a more gelatinous appearance when compared to a matte object with the same mechanical deformation (Figure 1A). Alternatively, optics may modulate the extent to which dynamic shape deformations can be perceived (e.g. the specular highlights on the translucent object create more image motion [18], [19] compared to its matte counterpart, making it appear wobblier, and therefore more gelatinous). A combination of associative and modulatory mechanisms is also possible. The effect of prior knowledge (i.e. associations) on how different types of sensory input are integrated has been looked at more formally within the framework of Bayesian models (e.g [20]; [21] [22]; [23]). In particular, cue conflict scenarios have proven extremely useful to generate insights about the complex interplay of prior selection and the weighting of sensory input in the perception of object properties (e.g. [24]. Here, we use an experimental paradigm analogous to cue conflict to create gross violations of expectations about materials (Figure 1B; Figure 4A and B). We show that the qualities of “surprising” materials are perceived different to expected ones that behave the same (Figure 3), and that surprise leads to increases in processing time of the stimuli (in line [25], [26], or recently [27], Figure 4C and D). Furthermore, our method provides a general technique to differentiate the extent to which material qualities are directly estimated from material kinematics versus being modulated by prior associations from familiar shape and optical properties.

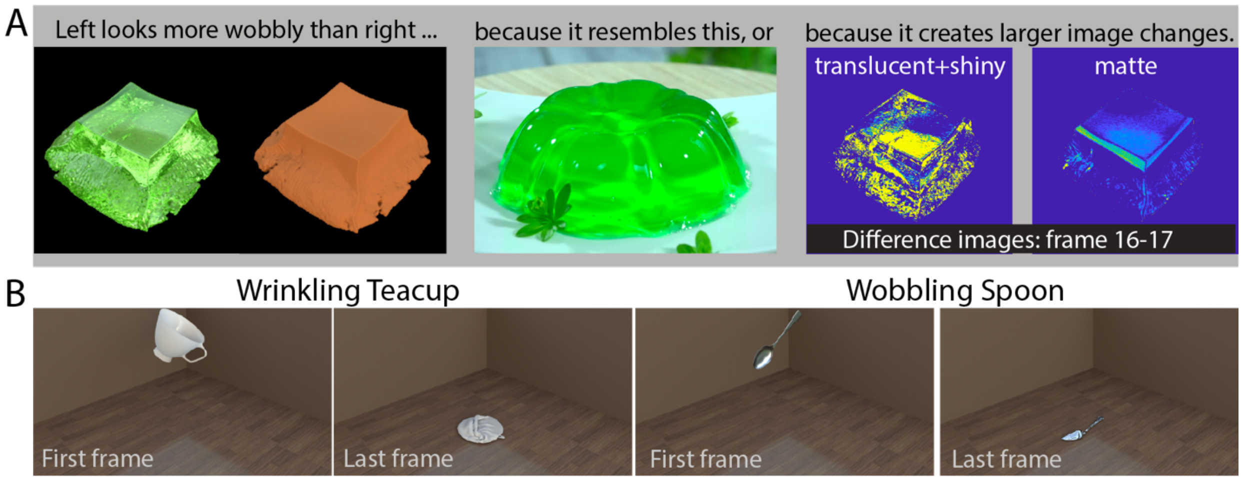

Figure 1. Contribution of prior associations and image cues on perceived material qualities. A. The perception of material qualities (such as gelatinousness) can be influenced by prior associations between dynamic optics, shape, and motion properties. Watching the green (left) object deform may evoke an association with green Jell-O, and may therefore be perceived as wobblier and more gelatinous than the matte object, despite both objects wobbling in identical ways (as shown in Movies S1 and S2). Alternatively, the green object may be perceived as wobblier due to larger image differences between frames, and potentially higher motion energy, as illustrated on the right. The difference in motion energy in greyscale images of the translucent object is about seven (6.8) times larger than that of the matte one, purely due to the difference in optical properties between these two objects. **B.** Shown are first and last frames from two animations used in our cue conflict design, where we pair familiar objects with atypical motions. Here, our expectations about the material behavior have been violated, leading to an experience of surprise [26]

Observers watched animations of objects (one per trial) fall to the ground and behave in one of five ways: rigidly (non-deforming), wobbling, shattering, wrinkling, or splashing. Objects behaved either as expected, based on their shape and surface properties (a porcelain teacup shattering, translucent Jell-O wobbling, etc.; Expected motion condition), or they behaved in a surprising way (a linen curtain shattering, a brass key wobbling, etc.; Surprising motion condition). Generally (but not always), hard objects in the Surprising condition behaved as soft or liquid ones and vice versa. On each trial, observers rated one of four material attributes on a continuous scale (blocked): hardness, gelatinousness, heaviness, and liquidity (Figure 2A). “Mechanical” qualities, like hardness and liquidity, have been of increasing interest in recent studies of deformable materials [17], [15], [28]; [29], [30] and are likely to be directly estimated from material kinematics [but see [16]– who found an additive effect of optics], whereas gelatinousness and heaviness might not be estimated directly from mechanical deformations and are potentially influenced more by associations with familiar shape and optics.

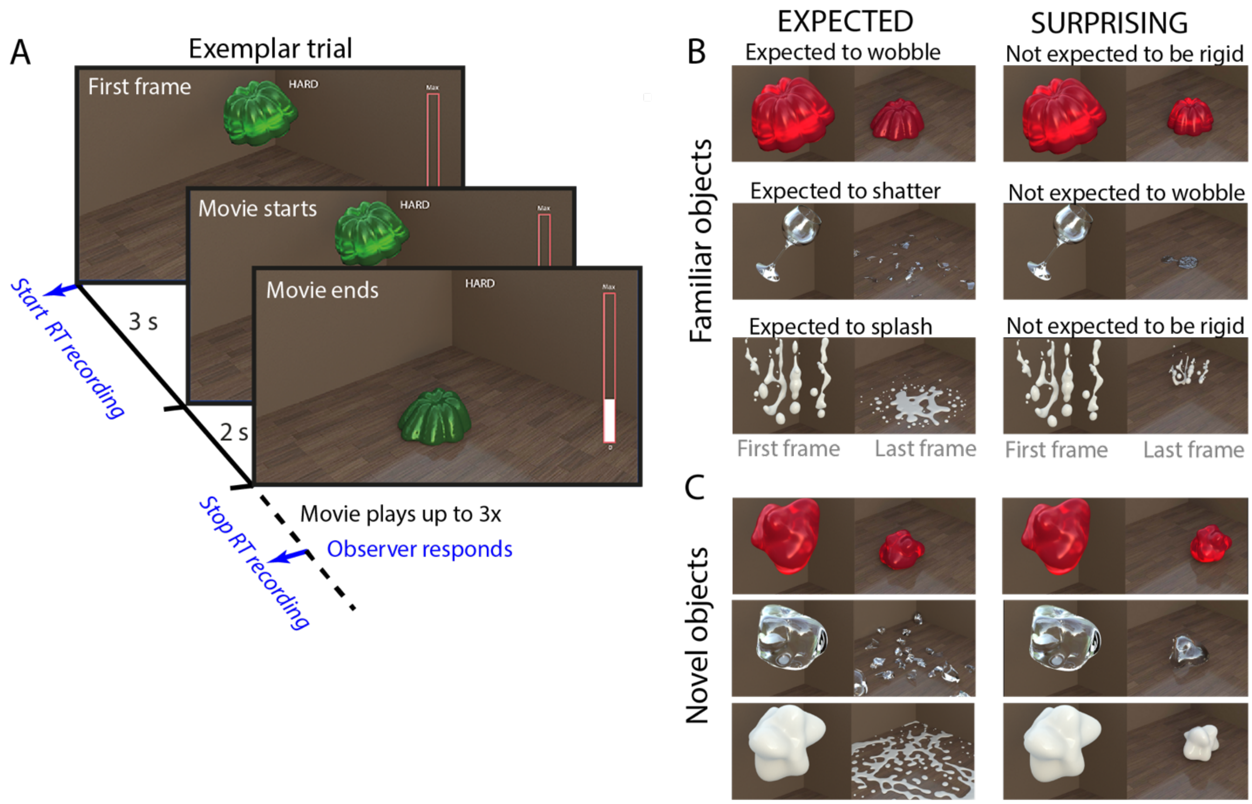

Figure 2. Trial and stimuli. A. An exemplar trial. **B.-C.** A subset of familiar objects (B) and corresponding novel objects (C) used in the experiments. Familiar objects could either 'behave' as expected or in a surprising manner. Note that this distinction (expected surprising) is only meaningful for familiar objects. Note that individual scenes are scaled to maximize the view of the object (First frame), or to give a good impression of the material kinematics (Last frame). Figures S1 and S2 show corresponding views for the entire stimulus set.

Each familiar object was paired with a matched unfamiliar, novel-shaped (to the observer) stimulus (Figure 2B and C, and Figures S1 and S2) that was rendered with the same surface properties and material behavior. We anticipated that Novel objects would not elicit strong prior expectations about how they will deform, and as such, their ratings would reflect an estimation of material properties from material kinematics, including any low-level effects of dynamic optical properties. Expected, Surprising, and the matched Novel (Expected & Surprising) objects were rated by the same observers within the same experiment, with a random order of presentation within each attribute block. A separate group of observers rated static images of the first frame of the Familiar and Novel objects (where each object was still intact).

The four attributes that observers rated can best be estimated when observing an object interacting with another one and should be difficult to judge from still images alone. Therefore, the First Frame ratings of Familiar objects could be used as a measure of *prior knowledge* about the material qualities from familiar shape and optics associations. Ratings of moving Novel objects, on the other hand, could be used as a measure of how much the image motion, generated by the kinematics of the material ('sensory' route), influences the rating (no influence of shape, equating for the effect of familiar optical properties). These two conditions make similar predictions (i.e. yield similar ratings) when the material behavior is expected (e.g. wobbling red

Jell-O; Figure 3A) but make *different* predictions (i.e. yield different ratings) when the material behavior is surprising (e.g. rigid (non-deforming) red Jell-O, Figure 3B), as evidenced by the following: First Frame ratings (from prior associations) differ significantly from Novel object motion ratings (from direct estimation) for 48 out of 60 (80%) in the Surprising motion condition, compared to only 24 out of 60 (40%) in the Expected motion condition (see Table S1). We operationally define and measure the "prior pull" as the distance between Familiar and Novel object motion ratings in the direction of First Frame ratings (Figure 3C).

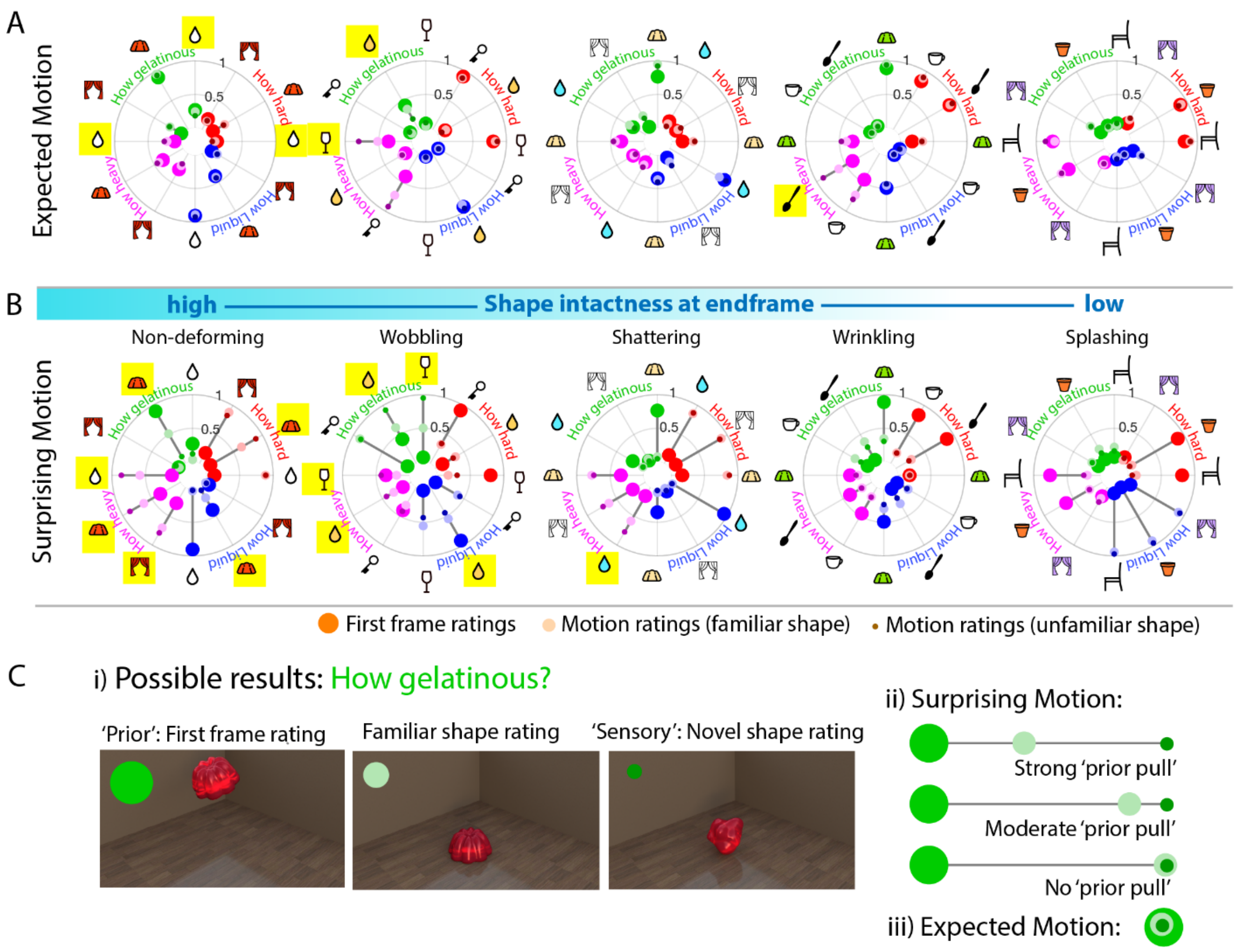

Figure 3. Material quality ratings and prior attraction. A. Average observer ratings for different questions about material qualities e.g. how hard, liquid, heavy or gelatinous an object appears. Icons symbolize individual familiar objects (chair, key, cup, pot, glass, spoon, blue droplet: water, yellow droplet: honey, white droplet: milk, violet curtain: silk, red curtain: velvet, white curtain: linen; yellow custard, red and green Jell-O). Each question and corresponding data are coded in the same color (red: hard, blue: liquid, purple: heavy, and green: gelatinous). Ratings could vary between 0 (lowest) and 1 (highest). On average, ratings from all 3 conditions (i.e. First Frame Familiar objects (large dots), typically-behaving Familiar objects (medium dots) and corresponding (moving) Novel objects (small dots)) tended to overlap. Organization of objects follows that of B. **B.** Same as A, but here, ratings of atypically-behaving familiar objects are plotted as medium-sized desaturated dots (now organized by motion), and ratings of corresponding Novel objects - i.e. unfamiliar shapes-- which inherit their optical and kinematic qualities from a familiar object - as small dark dots. Critically, we wanted to investigate whether the ratings of atypically-moving familiar objects are 'pulled' towards the prior. This clearly could be only the case if ratings on a given quality in First frame and Surprising motion experiments differ substantially, as indicated by the extent of the dark gray lines between these two types of data points. A prior 'pull' occurs when the ratings of atypically-moving familiar and novel objects do not overlap (significant cases

highlighted in yellow; also see Table S1); if instead they overlap completely, the object prior did not exert any significant influence on the rating. Overall, the more the familiar object remained intact in the surprise motion condition, the more likely the prior exerted an influence over the material appearance. **C. i)** Illustrates how we measure how much the rating of an atypically-moving familiar object (middle image) overlaps with the rating of a material-matched moving novel object (right image), or conversely, how much it is pulled towards ratings of a static view of the familiar object (left image). **ii)** Shows possible results: For example, seeing an image of red Jell-O in its classical shape, observers tend to expect that it is quite gelatinous. When they see an object with the same optical properties that falls and does not wobble when it hits the floor, they rate it - unsurprisingly - as very non-gelatinous. When a classically-shaped red Jell-O falls on the floor and doesn't wobble, observers could either rate it similar to the novel object -after all it doesn't wobble at all - or it could be rated as somewhat more gelatinous, despite the sensory input, possibly because prior experience influences the appearance, making observers perceive wobble when there isn't. **iii)** When the familiar object moves exactly as expected, and when there is no strong influence of shape familiarity on material judgements, all three ratings will overlap.

The yellow highlighted cases in Figure 3 show that this "prior pull" occurred twice as much in the Surprising (10 cases) than the Expected condition (5 cases, see Table S1), and more in conditions where the object was still intact and recognizable at the end of the movie (objects that behaved rigidly or wobbled). Prior pull in the Expected condition also occurred where estimation (sensory input) and associative (prior knowledge) accounts made different predictions (gelatinousness of the honey, heaviness of the wine glass, Figure 3A). In these cases, the "expected" cases were not so expected – this may be related to shape properties, or the size of the splashing of the liquids. Although we controlled for the effects of image motion from optics (e.g. specular highlights), perhaps other low-level image differences exist between Familiar and Novel objects that could be driving differences in ratings. To rule this out, we modelled the data using differences in size between Familiar and Novel objects in the first and last frames (pixel area difference, see STAR methods) and calculated differences in motion energy (Figure S3). Such a model performs extremely poorly (R^2 = 0.025, p>0.05). This suggests rating differences are not caused by differences in object size or image motion.

That only a subset of our objects exhibits a "pull" towards the prior is consistent with literature that shows when sensory input is unambiguous, it will dominate the percept, and prior knowledge does not have an effect [31]. In our case, particular combinations of shape and optics (often involving "shape recognizability") at the end of the animation may lead observers to essentially "ignore" sensory information. For example, the rigid Jell-O is perceived as gelatinous despite not wobbling. Interestingly, prior associations with specific combinations of motion, shape, and optics at the end of the movie may "enhance" differences between Familiar and Novel object ratings (e.g. in the case of the wobbling glass, the particular combination of translucent optics forming what looks like a puddle might evoke strong representations of liquidity, so despite the wobble, it is not considered gelatinous).

One might argue that the reason why we see relatively few instances of significant prior pull is that Novel objects might too elicit expectations (after all, they are bounded shapes with particular optical properties). The correlations in Figure 4A between First Frame and motion ratings, however, suggests that, for the most part, this is not the case. Familiar object first frame ratings predict Expected motion ratings extremely well (Figure 4A, R^2=0.9), whereas novel object first frame ratings predict motion ratings very poorly (Figure 4B, R^2=0.07). Thus, for the most part, Familiar object first frame ratings are a good measure of observers' prior expectations about the material qualities of these objects. In contrast, Novel object motion ratings are an ideal measure of sensory input because the motion from the material behavior includes information about both the mechanical deformations, as well as any additional effect of image motion from optics [18].

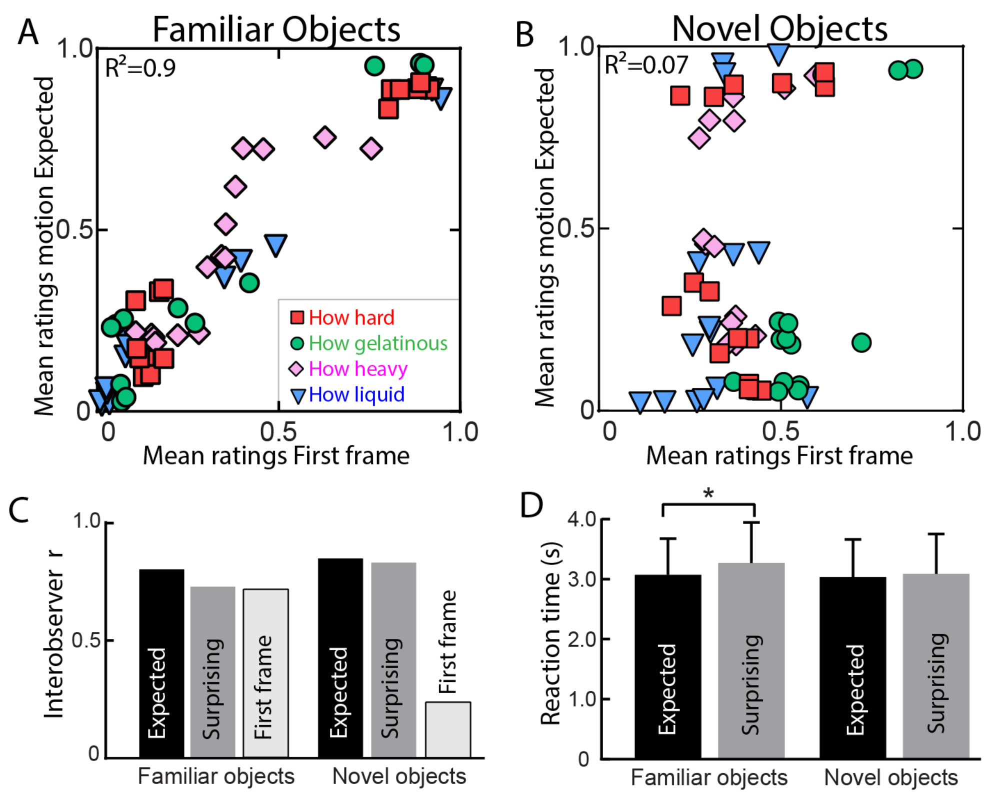

Figure 4. Prediction strength reaction time differences and interobserver correlations. A. Correlation between mean First frame ratings and mean Expected motion ratings for familiar and novel objects (**B.**) A high correlation indicates that the First frames (still images) of objects are highly predictive of the objects' kinematic properties, and thus are in good agreement with ratings in the Expected motion condition, where objects fall and deform according to their typical material kinematics. This is clearly not the case for novel objects, suggesting that these objects do not elicit strong prior expectations about how an object will deform. **C.** Average interobserver correlation for Expected and Surprising motion trials, as well as the First frame experiment. Note that only for novel objects, this latter correlation was quite low, suggesting that still images of unfamiliar objects do not elicit a strong prior in observers about the material qualities measured in this experiment. **D.** Reaction time data averaged across all observers for Expected (black) and Surprising trials (medium gray). Stars indicate significant differences p<0.005.

Given that there are a few cases where Novel objects do seem to generate correct predictions about the material outcome (those at the bottom left and top right of Figure 4B), and since some of the magnitude of the prior pull may be explained by shape recognizability at the end of the movie, we tested a linear regression model that predicted the direction and magnitude of Familiar and Novel object rating differences from prior pull from optics, familiar shape, and shape recognizability at the end of the animation (Figure S3B). We can model the data reasonably well (R^2 between 0.266 and 0.59 depending on the question), but not perfectly, potentially due to specific motion-shape-optics interactions [32].

We did not aim to test an exhaustive list of material attributes, but to determine whether effects of prior associations on visual input might depend on the type of material attribute judged, and on how the object behaves under external forces. We found that "mechanical" qualities like hardness and liquidity appear to be more directly estimated from material kinematics in "shape destroying" conditions (splashing, shattering, wrinkling) [[33]; [15]], but prior associations play a modulatory role (to the extent where material kinematics can even be ignored (e.g. red and green Jell-O)) when shape remains somewhat intact [32], Figure 1A, or see [34]. These latter conditions seem to create more of a cue conflict, and are more ambiguous. On the other hand, qualities like gelatinousness and heaviness (which are much more difficult to estimate directly from mechanical deformations) were more affected by familiar shape and optics associations [Figure S3B].

One might argue that the "prior pull" we demonstrated here is not perceptual but in fact is due to a particular 'cognitive strategy' of some observers (i.e. explicitly ignoring the motion information and thus rating material qualities of atypically moving familiar objects as they rated objects on the first frame, while other observers' ratings were 100% identical to novel object ratings). This would have resulted in bimodal rating distributions and/or low interobserver correlation in the object motion condition, neither of which we found (Figure S4 and Figure 4C, respectively). The prior 'pull' is a quite subtle effect, and for the majority of judgments, prior knowledge was not enough to 'outweigh' the strong material kinematics. In particular, when shape was not recognizable due to its deformation (splashing, wrinkling), the prior pull was essentially absent. This fits with the idea that if there is too much evidence against the prior, sensory input gets more weight [31].

Another argument *against* the cognitive strategy approach is supported by reaction time (RT) patterns in our experiments. A small but significant increase in RT in the familiar object surprising condition - which is the condition that most strongly juxtaposes prior expectation with sensory evidence - would be consistent with the idea of recurrent prediction error correction [27], Figure 4D, also see [25]). Importantly, we do not find evidence for a reaction time advantage in expected familiar objects condition (compared to novel, also consistent with [27]), which suggests that this increase in RT cannot simply be due to the fact that observers positioned the slider in advance to the 'wrong' (expected) position. If observers adopted such a strategy, we should have also seen faster RTs for expectedly moving familiar objects.

The human brain uses prior knowledge to continuously generate predictions about the visual input in order to make quick decisions and to guide our actions. Predictions about material properties are no exception to this: to avoid small daily disasters, we need to be able to predict how slippery and how heavy a cup is before picking it up. Our work shows that previously acquired object-material associations play a central role in material perception and are much more sophisticated than previously appreciated. Our results also offer an explanation to the seemingly conflicting findings in research that investigated the perception of material qualities of non-rigid objects. While some work proposes that perceived material qualities like softness are strongly influenced by motion and shape cues, which completely dominate optical cues [17], other work showed that *both* optical and mechanical cues affect estimates of viscosity [33] and yet other research concludes that optical properties dominates over image motion and shape cues when judging the stiffness of cloth [35]. Our results suggest that the prior on the material category (e.g.

rubber, liquids, cloth) determines how image cues (optical, motion, shape) are weighted and integrated with existing material knowledge to yield a specific material percept. Thus, the conclusions of previous research can be reconciled when the role of familiar shape priors is considered. This study extends a growing theme in the material perception literature that studying the perception of kinematic material qualities can serve as a tool to guide investigations of the neural mechanisms about material properties, as it provides insight into components (high and low level) that make up material perception as a whole.

References

[1] E. H. Adelson, "Human Vision and Electronic Imaging VI," 2001.

[2] L. T. Maloney and D. H. Brainard, "Color and material perception: Achievements and challenges," *J. Vis.*, vol. 10, no. 9, pp. 19–19, 2010.

[3] B. L. Anderson, "Visual perception of materials and surfaces," *Curr. Biol.*, vol. 21, no. 24, pp. R978–R983, Dec. 2011.

[4] R. W. Fleming, S. Nishida, and K. R. Gegenfurtner, "Perception of material properties," *Vision Res.*, vol. 115, pp. 157–162, Oct. 2015.

[5] R. W. Fleming, "Visual perception of materials and their properties," *Vision Research*, vol. 94. Pergamon, pp. 62–75, 01-Jan-2014.

[6] H. Komatsu and N. Goda, "Neural Mechanisms of Material Perception: Quest on Shitsukan," *Neuroscience*, vol. 392. pp. 329–347, 10-Nov-2018.

[7] K. Kveraga, A. S. Ghuman, and M. Bar, "Top-down predictions in the cognitive brain," *Brain and Cognition*, vol. 65, no. 2. NIH Public Access, pp. 145–168, Nov-2007.

[8] M. F. Panichello, O. S. Cheung, and M. Bar, "Predictive feedback and conscious visual experience," *Frontiers in Psychology*, vol. 3, no. JAN. Frontiers, p. 620, 21-Jan-2013.

[9] C. J. Bates, I. Yildirim, J. B. Tenenbaum, and P. W. Battaglia, "Humans predict liquid dynamics using probabilistic simulation," in *Cognitive Science Society*, 2015, pp. 172–177.

[10] J. Kubricht, C. Jiang, Y. Zhu, S.-C. Zhu, D. Terzopoulos, and H. Lu, "Probabilistic Simulation Predicts Human Performance on Viscous Fluid-Pouring Problem," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 2016, pp. 1805–1810.

[11] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum, "Simulation as an engine of physical scene understanding," *Proc. Natl. Acad. Sci.*, vol. 110, no. 45, pp. 18327–18332, Nov. 2013.

[12] J. I. Flombaum, B. J. Scholl, and L. R. Santos, "Spatiotemporal priority as a fundamental principle of object persistence," in *The Origins of Object Knowledge*, Oxford University Press, 2009, pp. 135–164.

[13] T. Gao and B. J. Scholl, "Are objects required for object-files? Roles of segmentation and spatiotemporal continuity in computing object persistence," *Vis. cogn.*, vol. 18, no. 1, pp. 82–109, Jan. 2010.

[14] J. F. Soechting, J. Z. Juveli, and H. M. Rao, "Models for the Extrapolation of Target Motion for Manual Interception," *J. Neurophysiol.*, vol. 102, no. 3, pp. 1491–1502, Sep. 2009.

[15] J. J. R. van Assen, P. Barla, and R. W. Fleming, "Visual Features in the Perception of Liquids," *Curr. Biol.*, vol. 28, no. 3, p. 452–458.e4, 2018.

[16] F. Schmidt, M. Hegele, and R. W. Fleming, "Perceiving animacy from shape," *J. Vis.*, vol. 17, no. 11, p. 10, Sep. 2017.

[17] V. C. Paulun, F. Schmidt, J. J. R. van Assen, and R. W. Fleming, "Shape, motion, and optical cues to stiffness of elastic objects.," *J. Vis.*, vol. 17, no. 1, p. 20, Jan. 2017.

[18] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten, "Visual motion and the perception of surface material," *Curr. Biol.*, vol. 21, no. 23, pp. 2010–2016, Dec. 2011.

[19] K. Doerschner, D. Kersten, and P. R. Schrater, "Rapid classification of specular and diffuse reflection from image velocities," in *Pattern Recognition*, 2011, vol. 44, no. 9, pp. 1874–1884.

[20] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling

of depth cue combination: in defense of weak fusion.," *Vision Res.*, vol. 35, no. 3, pp. 389–412, Feb. 1995.
[21] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, Jan. 2002.
[22] J. H. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy, "Combining sensory information: Mandatory fusion within, but not between, senses," *Science (80-.).*, vol. 298, no. 5598, pp. 1627–1630, Nov. 2002.
[23] D. Kersten, P. Mamassian, and A. Yuille, "Object Perception as Bayesian Inference," *Annu. Rev. Psychol.*, vol. 55, no. 1, pp. 271–304, Feb. 2004.
[24] D. C. Knill, "Learning Bayesian priors for depth perception," *J. Vis.*, vol. 7, no. 8, p. 13, Jun. 2007.
[25] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Res.*, vol. 49, no. 10, pp. 1295–1306, Jun. 2009.
[26] P. Baldi and L. Itti, "Of bits and wows: A Bayesian theory of surprise with applications to attention," *Neural Networks*, vol. 23, no. 5, pp. 649–666, Jun. 2010.
[27] B. M. Urgen and H. Boyaci, "When expectations are not met: unraveling the computational mechanisms underlying the effect of expectation on perceptual thresholds," *bioRxiv*, p. 545244, Feb. 2019.
[28] B. Xiao, W. Bi, X. Jia, H. Wei, and E. H. Adelson, "Can you see what you feel? Color and folding properties affect visual–tactile material discrimination of fabrics," *J. Vis.*, vol. 16, no. 3, p. 34, Feb. 2016.
[29] T. Kawabe, K. Maruya, and S. Nishida, "Perceptual transparency from image deformation," *Proc. Natl. Acad. Sci.*, vol. 112, no. 33, pp. E4620–E4627, Aug. 2015.
[30] W. Bi, P. Jin, H. Nienborg, and B. Xiao, "Estimating mechanical properties of cloth from videos using dense motion trajectories: Human psychophysics and machine learning," *J. Vis.*, vol. 18, no. 5, p. 12, May 2018.
[31] C. Summerfield and F. P. De Lange, "Expectation in perceptual decision making: Neural and computational mechanisms," *Nature Reviews Neuroscience*, vol. 15, no. 11. pp. 745–756, 15-Nov-2014.
[32] A. C. Schmid and K. Doerschner, "Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials," *J. Vis.*, vol. 18, no. 1, p. 14, Jan. 2018.
[33] J. J. R. van Assen and R. W. Fleming, "Influence of optical material properties on the perception of liquids," *J. Vis.*, vol. 16, no. 15, p. 12, Dec. 2016.
[34] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner, "Memory modulates color appearance," *Nat. Neurosci.*, vol. 9, no. 11, pp. 1367–1368, Nov. 2006.
[35] C. Aliaga, C. O'Sullivan, D. Gutierrez, and R. Tamstorf, "Sackcloth or silk?," in *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception - SAP '15*, 2015, pp. 41–46.

Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information for resources should be directed to and will be fulfilled by the Lead Contact, Alexandra Schmid (Alexandra.schmid@psychol.uni-giessen.de)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Motion Experiment: 25 participants (mean age 24.8; 18 female) participated in the experiment; 23 were right-handed and all had self-reported normal or corrected-to-normal vision. All participants were native German speakers, and the experiment was given entirely in German. The experiment followed the guidelines set forth by the Declaration of Helsinki, and participants were naïve as to the purpose of the experiment. All participants provided written informed consent and were reimbursed at a rate of €8 per hour.

First Frame Experiment: 15 participants (mean age 26.40; 13 female) participated in this experiment; 13 participants were right-handed and all had self-reported normal or corrected-to-normal vision. All participants were native German speakers, and the experiment was given entirely in German. The experiment followed the guidelines set forth by the Declaration of Helsinki, and participants were naïve as to the purpose of the experiment. All participants provided written informed consent and were reimbursed at a rate of €8 per hour.

Last Frame Experiment: 14 participants (mean age 26.85; 10 female) participated in the 'Static Last Frame' condition of Experiment 1; 12 were right-handed. All participants had self-reported normal or corrected-to-normal vision. All participants were native German speakers, and the experiment was given entirely in German. The experiment followed the guidelines set forth by the Declaration of Helsinki, and participants were naïve as to the purpose of the experiment. All participants provided written informed consent and were reimbursed at a rate of €8 per hour.

METHOD DETAILS

Stimuli: We used two types of objects: Familiar and Novel. Figures S1 and S2 give an overview of all Familiar objects used in the experiment. In order to create a stimulus set with a broad range of typical material classes, we choose 15 Familiar objects belonging to one of 5 material mechanics: Splashing (milk, honey, and water droplets), Shattering (wine glass, terracotta pot, porcelain teacup), non-deforming (wooden chair, brass key, metal spoon), wobbling (red and green jelly, custard), and wrinkling (linen, velvet, and silk curtain). All objects were rendered in Blender [1] with their typical optical material properties (e.g. metallic-looking key, green transparent jelly, or a silky- appearing curtain). Objects were rendered and shown in the context of a room consisting of brown walls and a polished hardwood floor. All scenes were illuminated by the "Campus" environment map [2]. The 3D meshes for the chair, glass, and spoon were obtained from Free3D [3] and TurboSquid [4]; the remaining objects were created by us. Unfamiliar shapes ('Glavens', [5]) inherited their optical and mechanical properties from the corresponding Familiar object (Figure S1 and S2).

For each object, we rendered short movies of these objects showing an object falling to the ground and interacting with the floor on impact (Figure 2A). In order to manipulate surprise in our experiment, an object could either behave as expected, e.g. a glass would shatter (Figure S1), or it could inherit the mechanical material properties from another object, e.g. a chair would splatter like milk upon impact (Figure S2). We created corresponding Expected and surprising movies for Novel objects (Figure S1, S2).

Animations: Each movie consisted of 48 frames, depicting an object suspended in air, which then fell to the ground. Impact occurred on the 11^{th} frame for all objects. The largest extent of the objects in the first frame varied between 6.91 (clay pot) and 12.6 (spoon) degrees visual angle (also see Figure 2 and S1). The largest extent of the objects in the last frame depended on the deformation, but varied between 48.85 degrees visual angle for Shattering/Splashing items and 4.29 degrees visual angle for rigidly falling items. Exactly how each object would deform was determined by us using a rigid body physics simulation carried out by the Physics Engine in Blender. For technical specifications, we refer the reader to the parameters listed in Schmid & Doerschner (2018).

Apparatus: The experiment was coded in MATLAB 2015a [6] using the Psychophysics Toolbox extension [7], and presented on a 24 5/8'' PVM-2541 Sony (Sony Corporation, Minato, Tokyo, Japan) 10-bit OLED monitor, with a resolution of 1024 x 768 and a refresh rate of 60 Hz. Videos were played at a rate of 24 frames per second. The participants were seated approximately 60 cm from the screen.

Task and Procedure: Observers were asked to watch a short video clip to the end and then to rate the object they saw on one of four attributes, as quickly as they could. We choose the attributes such that they would capture some aspect of the mechanical material qualities of the objects. For example, a Splashing object is likely to be rated as very *liquid*, and a non-deforming object not, a wiggling object is likely to be rated as very *wobbly* but a Shattering object not. In order to familiarize observers with the rating task, the use of the slider bar and the keypresses, they completed four practice trials with two objects that did not occur in the actual experiment and with two rating adjectives that also did not occur in the experiment (e.g. rate how shiny this object is).

The experiment was organized into four blocks, one block per attribute. Before the block started, the observer was familiarized with the rating question and then proceeded with a button press to start the trials. On every trial, a reminder of the question of this block remained at the top of the screen, e.g. '*hard*' (for 'How hard is the object?'), together with the first frame of a movie, which was held static for three seconds before the movie was played to the end. The movie clip then repeated 2 more times (without the hold at the beginning). Participants were asked to first watch the video until it finished (i.e. the first play through) and then to rate the object, as quickly as they could. They indicated their rating by using the mouse to adjust the height of a slider bar placed on the right side of the screen (Figure 2). A zero setting indicated the absence of an attribute, e.g. not *gelatinous* at all, while a maximum setting would correspond to the subjective maximum value of an attribute. The trial was completed when the observer pressed the space key on the keyboard, after which time the next trial would immediately begin. Reaction time was measured from the beginning to the end of a trial (between spacebar presses). The slider position of the previous trial was carried over the new trial in order to give the experimental interface a more natural feel to it.

Participants completed 240 trials in total (2 surprise conditions (Expected, Surprising) × 2 object familiarity (Familiar, Novel) × 4 attributes × 15 objects). Surprising condition and object type were the two relevant manipulations in the experiment. While the order of blocks was the same for all observers (*hard*, *gelatinous*, *heavy*, *liquid*), the trial order in each block was randomized.

The instructions to the participants (on-screen, prior to the start of the experiment) were the following:

> Rate how HARD each object looks. A setting of zero means not at all hard (soft). A setting of Max (corresponding to a numerical value of 1) means that the object looks extremely hard.
>
> Rate how GELATINOUS each object looks. A value of zero means not at all gelatinous. A value of Max means that the object looks extremely gelatinous.
>
> Rate how HEAVY each object looks. A setting of zero means the object looks very lightweight. A setting of Max means the object looks extremely heavy.
>
> Rate how LIQUIDY/FLUID each object looks. A setting of zero means not at all liquidy/fluid. A value of Max means that the object looks extremely liquidy/fluid.

First Frame/Last Frame Experiment: The tasks, setup, and procedures were identical to that of the motion experiment. In contrast to the Motion experiment, in this experiment, the first/last frame of each of the videos was held on-screen for the same duration as a single presentation of the video.

QUANTIFICATION AND STATISTICAL ANALYSIS

All experiments were performed in MATLAB using Psychtoolbox (v. 3.0.12) [6,7]. Analyses were performed in MATLAB and Excel. No observers were excluded from the analysis.

Data point exclusions: For the first 10 subjects, data points for the Novel object matched to the Surprising key object were excluded due to a rendering error that made the object behave rigidly rather than Wobble, as they should have in the matched-Surprising condition (10 subjects × 4 attributes = 40 data points excluded). For the remaining 15 subjects this error was fixed.

Reaction times: Time taken to make each judgment (reaction time) was measured. We reasoned that rating the material properties of materials that behave surprisingly might involve the reiterative correction of a prediction error by the visual system, and this error correction might be associated with an increase in reaction time when rating objects that behave in a surprising way. Before computing the difference in reaction time between Expected and Surprising conditions, we pre-processed reaction time data as follows: we subtracted the time to impact (3 seconds static first frame + 0.45 seconds to impact) from the raw reaction times so that a reaction time of zero would now indicate time of impact. Data points that were faster than 0.75 seconds after impact (fastest possible button press) were excluded. Response latencies that were longer than 2 standard deviations above the mean were also excluded. Following these exclusions, approximately 6% of the data were excluded for reaction times that were too fast or too slow according to this criterium (~1.3% too fast, ~4.7% too slow).

Linear Regression Model: We developed two models with the aim to account for the differences we observed in ratings of familiar and novel moving objects. Low- and high-level models had each exactly 3 predictors.

High-level Model: Each predictor in the high-level model was multiplied by a weight G that took into account the *predictability* of the "behaviour" of a given stimulus (computed as the difference first frame and motion ratings in the Expected condition) and the *extremeness* of the rating (computed as the twice the difference between first frame ratings in the Expected condition and .5). Thus, predictors in the high-level model were:

X1 (=X1*G), where G= *predictability*extremeness*: the *optics & bounded shape prior* (computed as the difference between First frame ratings of Novel objects in the Expected condition and moving stimuli ratings of Novel objects in the Expected condition),

X2 (=X2*(1-G)): the *prior-pull* (computed as the overall differences between First frame ratings of Familiar objects and corresponding ratings of moving Novel objects (across both, expected and surprising conditions)

X3 (=X3*(1-G)): *last frame shape recognizability* (computed as the overall differences between Last frame ratings of Familiar and Novel objects (across both, expected and surprising conditions).

Low-level Model: Predictors in the low-level model were:

X1: *motion energy differe*nce (this was computed in two steps):

> a) take the sum of the absolute value of all consecutive image differences, starting with the impact frame, e.g. sum(abs(f11-f12), abs(f12-f13), …, abs(f39-f40)), [1], for all experimental conditions/stimuli (60 in total) & normalize these values by object size (number of pixels corresponding to the object on frame 11 ME_n,
>
> b) for each object in the expected and surprising conditions, compute the differences of $ME_{Familiar}$ and ME_{Novel} conditions

X2: *first frame object size* (computed as number of pixels in the first frame corresponding to the object),

X3: *last frame object siz*e (computed as numbers of pixels in the last frame corresponding to the object).

Stimuli

https://www.dropbox.com/sh/bmlf4z03mgfqfna/AABUSwubHleEHp_l1lYkgKkEa?dl=0]