Perception of Material Kinematics

Lorilei M. Alley

M.Sc. Universita degli Studi di Trento (2015)

B.A. Rutgers University, New Brunswick (2013)

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Fachbereich 06

This dissertation is submitted for the degree of

Dr. rer. nat

September 2019

Abstract

Many objects that we encounter have 'typical' material properties that are related to their specific affordances: spoons are hard, pillows are soft, and Jell-O is wobbly. Over a lifetime of experiences and interacting with these objects, 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 particularly striking examples 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. We take this finding a step further and ask, when people are making judgments about the material properties of an object, where on the object do they look when making such judgments? Which areas are most informative for which material judgments? Can a linear classifier predict this? Ultimately, understanding which regions of an object are critical for material perception will provide insight into how the human visual system solves the 'problem' of material perception rapidly and efficiently.

Prüfungsamt Declaration

“I hereby declare that I have prepared the thesis at hand independently and without undue aid or the use of any resources other than those indicated within the thesis. All parts of my thesis taken either verbatim or analogously from the published or unpublished works of or based on oral communications with others are indicated as such. Regarding all aspects of my scientific inquiries as they appear in my thesis, I have upheld the tenets of good scientific practice as laid out in the "Satzung der Justus-Liebig-Universität Giessen zur Sicherung guter wissenschaftlicher Praxis" and complied with the precept of ethics, data protection and animal welfare. I declare that I have neither directly nor indirectly given monetary or any other valuable considerations to others in connection with the thesis at hand. I declare that I have not presented the thesis at hand, either in an identical or similar form, to an examination office or agency in Germany or any other country as part of any examination or degree. All materials from other sources as well as all works performed by others used or directly referenced within the thesis at hand have been indicated as such. In particular, all persons involved directly or indirectly in the development of the thesis at hand have been named. I agree with the screening of my thesis for plagiarism via offline or online detection software.”

Dedication

To the ones who said it could never be done.

(Edgar Guest, 1950)

Here is this mass of jelly - three pound mass of jelly - that you can hold in the palm of your hand, and it can contemplate the vastness of interstellar space, it can contemplate the meaning of infinity, and it can contemplate itself contemplating the meaning of infinity.

--V.S. Ramachandran, “A Journey to the Center of Your Mind”

Acknowledgements

- **Katja Doerschner**, for offering me the opportunity to pursue a Dr. rer. Nat. in material perception.

- **Roland Fleming**, for the support and feedback throughout the Dr. rer. Nat.

- **Jan Koenderink** and **Andrea van Doorn** for their invaluable support and feedback as I followed them around the world, from Rutgers, to Leuven, to Trento, to Gießen. I am very thankful for their friendship and continued support of my work.

- **Shin'ya Nishida and the entire NTT Communication Science Laboratories** for hosting me in Atsugi and for invaluable support and enthusiasm for our work. I hope to incorporate the skills learned there in many other research forays.

- **Flip Phillips**, for the support and feedback throughout the entire PhD and for motivating me to keep going.

- **Katerina Kamprani**, whose installation 'The Uncomfortable' inspired this thesis and entire body of research.

- **Matteo Valsecchi**, for the support and friendship throughout the entire PhD, and for reminding me to appreciate the small things.

- **Matteo Toscani,** for the support and friendship throughout the entire PhD.

- **Naoki Kogo,** for the support and friendship throughout the entire PhD, whether it was in Belgium, Florida, the Netherlands, or Japan. Thank you for your support and wisdom over the years.

- **To TK** and **Angela:** Vielen Dank, dass Sie auf dieser Doktorandenreise gute Freunde sind.

- **Robert Ennis,** for the continued assistance throughout the experiments described in this thesis.

- **Alexandra Schmid,** for contributions to the studies described in Chapter 2.

- **To B-O. W:** Thank you for putting up with me on the rougher days, and celebrating on the good ones.

- **To our HiWis,** thank you for collecting the data from many of these experiments- I (literally) could not have done this without you!

- **To my (over 300) participants**: Thank you for participating in our experiments—this research would not have been completed without you!

- **Lavazza** and **Licher,** whose products got me through the writing of this thesis (and related manuscripts).

- To my cat, Kedi, for ‘trying to help’ with this thesis by sitting on the keyboard ☺

- To all those I have not yet thanked, please know that I appreciate your support throughout the Dr. rer. Nat.

Contents

List of Figures

Asterisk (*) denotes animations exist, which correspond to figures.
Animations can be found on Zenodo (see References)

List of Tables

Overview

Here I present a broad overview of all experiments presented in this thesis, and include highlights of the main findings from each experiment. Portions of the experiments described in Chapter 2 has been posted as a preprint on biorXiv and is currently under review at Journal of Vision.

Surprising Materials Experiment

Here we present a novel set of stimuli that depicts familiar objects deforming in unexpected ways (a cloth shattering, a teacup wrinkling, etc.). These unexpected deformations allow us to probe an understanding of typical kinematic behaviors of materials. 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.

Highlights:

- *Surprise Motion (A Pilot Experiment):* We demonstrate 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.

- *Full-Motion/Full Color Rating Experiment:* We extend our findings from the Pilot experiment, and show that results can be replicated with a richer set of stimuli.

- *Greyscale Rating Experiment:* Investigates the potential influence of (typical) color on our previous findings, and shows that similar results for color and greyscale stimuli support previous findings of top-down influences of typical color on object perception.

Control Experiments

Here we describe a series of experiments that aim to clarify and extend our findings from the primary experiments described in Chapter 2.

Highlights:

- *'Questions' Experiment:* Confirms that participants are identifying the stimuli as we intended.

- *Color vs. Grey (2 AFC) Experiment:* Investigates the Ratings-based method of response, and confirms that typical color is an informative cue to object identity.

- *'Touch' Experiment:* Confirms that shape (and the deformation thereof) is a dominant cue to object identity.

Eye Tracking Experiments

How do existing priors (which allow one to make predictions) affect where people look when judging stiffness or lightness? Where you look depends on the attribute judgement you are making. Where do people look when predicting future states of an object/making a material judgement? Can a linear classifier reliably predict which task a participant is performing, based on their fixations alone?

Highlights:

- *'Pilot Blobs':* Confirms that fixations are drawn to the most informative regions of the object for material property judgements in dynamic scenes: For lightness, fixations remain near the brightest region of the object. For stiffness, fixations are drawn toward the regions of greatest motion.

- *'Punching Bag':* Extends the findings of the 'Pilot Blobs' stimuli. Confirms that a linear classifier is able to successfully predict (by the fourth fixation) which task the participant is performing (stiffness or lightness). The fixation landing position can be reliably used to classify the task.

CHAPTER 1: Introduction

Overview

In everyday life, we encounter a multitude of objects and surfaces, and their corresponding materials. With almost no effort, we can identify multiple categories of materials (textiles, liquids, non-breakable rigid items, etc.), decide to which category of material a novel material belongs (drinkable versus non-drinkable liquids, wearable vs. non-wearable textiles), and even infer properties of a material that cannot be directly observed, whether the material is a non-newtonian fluid (e.g. Oobleck) (Seuss, 1949), beer, or the fur of Truffula trees (Seuss, 1972).

With little awareness, we interact with these objects and materials successfully, making inferences rapidly without pausing to notice the material of which an object is composed (e.g. "The egg is fragile so I will grip it carefully, but the soccer ball is elastic and will form its shape again so I will kick it"). Such decisions are largely based on utility—can this object (which is shaped like a cup) be used to hold water, based on its material? Material perception, then, is used as a way to know how to interact with an object. Predictions about future states of an object allows us to interact successfully, and yet, we select the correct interaction behavior consistently. This ability to identify material rapidly means that we can use this information to perform tasks other than lifting-- we can also slide across a freshly-waxed bowling floor, avoid a falling knife, or time the velocity of a flying rubber racquetball to win the game.

Traditional views of material perception have focused on optical properties of static objects. In this thesis, we take a different, more cognitive approach to material perception than is usually studied. The work described here is the first (to our knowledge) to violate basic expectations regarding familiar object shape and kinematics to uncover basic assumptions regarding real-world physics about familiar objects.

To that end, this thesis will address the following questions, providing new insight into prediction of object kinematics:

1. What is the role of object knowledge and priors on material perception in motion?
2. What regions/features of an object provide critical information regarding the judgment of material properties?

In this chapter, I will describe material perception (and the influence of shape and motion) broadly, using the following examples to provide context:

Figure 1. Rainboots. This figure, taken from Katerina Kamprani's 'The Uncomfortable', depicts a pair of rainboots made of waterproof rubber material. In the first few hundred milliseconds, object recognition allows us to identify this object as a pair of rainboots. However, we predict future states of this object, and realize that the violation of shape (open-toed rain boots) in this object would violate their utility as rainboots. This leads to a noticeable sense of surprise. *Reprinted with permission.*

This figure, taken from Katerina Kamprani's 'The Uncomfortable' (Kamprani, 2012, *reproduced with permission*), depicts a pair of rainboots made of waterproof rubber material. In the first few hundred milliseconds, object recognition allows us to identify this object as a pair of rainboots. However, we *predict future states* of this object, and realize that the violation of shape (open-toed rain boots) in this object would violate their utility as rainboots. This leads to a noticeable sense of surprise.

Motion graphics artists have taken this observation on-board, and leveraged our ability to predict future states of objects by directly violating material properties. In Figure 2, below, the artist builds up an expectation in the viewers: The (rubber) balloon is approaching the (spiky) cactus and will suddenly 'pop' on contact with the spines. At the end, the expectation is violated when the *cactus* pops *instead of* the balloon.

Figure 2. Three frames from "Preposterous" by Florent Porta. Artists have played with our expectation of how objects and their materials should behave. In this study, we compare material perception for falling objects that deform in surprising and unsurprising (i.e. expected) ways. Retrieved from VImeo. *Reprinted with permission.*

Taking inspiration from these artworks, this thesis is a first step into the investigation of the interaction of material, motion, and shape, and the role that expectations and inferences have on material perception.

Object Representation (In-Brief)

What is object representation and how does it work? In this chapter, I will briefly describe why it is important to investigate how object representation plays a role in interacting with objects in daily life.

Kahneman et. al's 1992 object file theory proposes that 'object files' exist to collect, store, and update information for specific objects over time via a process called feature integration. This can include properties like form, color, shape, and size. One of the challenges of object recognition research in present day is to investigate what the inputs to object recognition are, and to develop an understanding of all possible inputs. In this thesis, we argue that these inputs to object recognition also influence our perception of the kinematic properties of materials, and that the material of which something is composed (as well as its expected kinematic behavior) are part of object representations.

Feature integration in ventral stream object files integrates information like *form* and *color* into an intermediate representation (Marr & Nishihara, 1978; Biederman, 1987). This intermediate representation is then combined with information like *motion, speed, shape, size, and direction* from the dorsal stream (Murata et. al, 2000). The resulting outcome is a 'full object representation' (Marr & Nishihara, 1978; Biederman, 1987), which is then used for object recognition.

Combining material and shape: With respect to material properties and shape, Pinna & Deiana (2015) argue that (identification of) shape and material depend on the same source. They posit that if the shape comes first, the "material" must explain the variation. Based on these object representations, semantic knowledge, and repeated exposure to familiar objects, we have expectations about object outcomes. The 'Touch' experiment described in this thesis answers exactly this question.

In work on scene perception and (typical) object location, the work of Kaiser, et.al. (2018) confirms that over a lifetime of interacting with objects, learning that objects occur at specific locations leads to improved perceptual processing when objects appear in their typical locations. Is this then true for pairings of object shape and material? Does the visual system process objects more quickly when they are presented in their typical color or material? Experiments contained in this thesis begin to investigate these questions.

Is it possible then, that expected kinematic behavior is *also* part of object representations? We would argue that our findings suggest that kinematic behavior *must also* be included in the list of inputs that the visual system relies on when making judgments about objects and the materials of which they are composed.

Now that we have broadly discussed object recognition, we must consider what judgements people can make regarding the material of which they are composed.

What types of judgements can people make about material?

A large amount of work has been done concerning the 'thingish' qualities of objects (that is, what they are, what we name them, where they are found). In contrast, a smaller amount of work has been done to investigate the 'stuffish' qualities of objects (that is, what they are made of) (Adelson, 2001). Tables are often made of wood, spoons are often made of metal, oobleck (Seuss, 1949) is made of cornstarch and water. To understand judgments about material perception broadly, we must first understand what kinds of judgements people can make about objects and their materials.

They can categorize. Fleming, Wiebel, and Gegenfurtner (2013) showed participants close-up photographs of materials from a variety of categories and asked them to rate certain attributes (hardness, glossiness, and prettiness). Even though subjects were not told *a priori* that that the stimuli were from various material categories, ratings clustered into categories, suggesting that subjects used visual judgements of image properties to classify materials.

They can make judgments about whether an object is Real or Fake. Sharan et. al. (2014) presented a series of real and fake objects to observers, asking them to identify the object category (dessert, flowers, fruit), and to categorize whether the image shown contained a real or fake object. They found that categorization judgments were faster and more accurate than real versus fake judgments, suggesting that object and material categorization engage distinct mechanisms.

They can infer properties and characteristics that cannot be directly observed. Fleming (2014) describes the usefulness of inferring properties of an object based on its material appearance—the characteristic look of precious metals and gemstones often yields large sums of money. Such an estimate is given, largely based on surface appearance. Indeed, companies rely on our ability to perceive material properties, and place a substantial amount of effort in making the optical properties look as pleasing as possible. However, vanAssen et. al., (2016) makes the following notable exception: Estimates of temperature are *not* inferred from the observation of surface properties or viscosities of liquids, suggesting that our ability to infer material qualities is not unbounded.

They can make judgements of gloss, translucency, and roughness. In glossiness, Nishida and Shinya (1998) showed that observers can judge specular reflectances of surfaces that appear glossy. Fleming, Dror, and Adelson (2003) later showed that this gloss-judgement ability generalizes across lighting conditions. Motoyoshi and Matoba (2012) demonstrated that the statistical characteristics of the illumination has systematic effects on perceived glossiness. Specular reflectance inferences are affected by both binocular disparity and motion information (Blake & Bülthoff, 1990; Doerschner et al., 2011; Hurlbert, Cumming, & Parker, 1991; Koenderink & van Doorn, 1980; Muryy et al., 2013; Wendt, Faul, & Mausfeld, 2008), as well as the properties of highlights (Beck & Prazdny, 1981; Berzhanskaya et al., 2005; Fleming, Torralba, & Adelson, 2004; Kim, Marlow, & Anderson, 2011; Marlow, Kim, & Anderson, 2012; Todd, Norman, & Mingolla, 2004). To investigate the hypothesis that the visual system uses (the combination of) certain characteristics as a heuristic for the estimation of surface parameters, Marlow et. al. (2012) asked participants to rate properties of highlights (rather than estimate surfaces) and found that assigning various weights to the properties accounts for glossiness judgements, suggesting that what participants actually do is compare properties of characteristic signatures of a material quality (for highlights, size, contrast, distinctness, etc.). Judgments of apparent glossiness actually reflect the extent to which the surface manifests salient specular reflections.

They can extrapolate haptic cues from visual cues. Bergmann-Tiest & Kappers (2007) have shown that visual estimates of roughness relates to haptic impressions of roughness. Ho, Landy, and Maloney (2006) demonstrated that judgments of surface roughness are biased by the illumination under which they are observed, suggesting that glancing illumination angles make surfaces appear rougher than frontal illumination.

By no means are these the only judgements that people can (and do) make regarding material properties. It remains a challenge for material perception research to continue investigating exactly what (and how) the visual system (and higher cognitive processes) takes into consideration when making material property judgements.

Judgements about material are rapid and simple

Identifying the material of which an object is composed happens rapidly and is (perceptually) simple. With seemingly no effort, we rapidly (and successfully) select the correct object for the task at hand (toothpaste instead of antibacterial cream for our toothbrush, although the shapes are surprisingly similar), fixate the relevant portion of an object or scene for the goal we are trying to achieve (e.g. moving our head to check for specular highlights to avoid slipping on black ice), and interact with these objects successfully.

Sharan et al. (2014) have shown that the speed for judgements of material perception can be equally as fast as judgements concerning object categorization or scene perception. In their 'RapidCat' experiments, they show that material categorization is simple, that object knowledge plays a role in the judgement being made, and that categorization is rapid—material category judgements can (successfully) be made after only a brief exposure to a stimulus.

Similarly, in studies of color perception and categorization, Witzel et. al., (2011) have shown that color category judgements are rapid and simple. In their study, they show that a correlation exists between reaction time and diagnosticity of the color of the object, suggesting that expectation facilitates processing speed of object identity. Our color-versus-greyscale results are consistent with this finding.

Inferences based on shape and material: Naïve/Intuitive Physics

Higher-level (top-down) cognitive processes also influence object perception. Referring to the previous 'rainboots' example (Figure 1): In the first few hundred milliseconds of observing the rainboots, their shape and material seems to be consistent with our experience with waterproof footwear. However, upon closer inspection, we simulate future states of the object (e.g. "I am imagining myself wearing this pair of rainboots") and through this process, suddenly realize that our toes and feet would get wet if we were to wear these boots in the rain (rendering their utility moot). Here, our expectations have been violated and a distinct sensation occurs (Itti and Baldi (2010), "units of 'wow'"). This distinct sensation may be related to the necessity of correcting a prediction error. This phenomenon is not limited to mental imagery, but rather, applies to all judgments of objects in the world: A precariously stacked pile of dirty dishes will fall, a Jenga tower may or may not tip over if a block is removed, and if you pull the one apple out of the pile at the supermarket (on which, all other apples are resting for stability), the pile will collapse. Battaglia et. al., (2013) have proposed a mechanism by which such expectations about object physics may develop. They propose the existence of an 'intuitive physics engine' (IPE) as a cognitive architecture that "allows for the estimation and simulation of object interactions by encoding properties like elasticities, rigidities, surface characteristics, and velocities". Akin to a rendering algorithm, humans utilize a process of repeated 'mental simulations' as a means of predicting future states and events involving objects (See Ullman, et.al., (2017) for a review).

Consistent with this literature, our results demonstrate for the first time that prediction mechanisms for future states of objects, materials, and their kinematics are much more sophisticated than previously appreciated, and suggests a central role of prediction in studies of material perception that has not been previously considered.

What is the Role of Motion in Material Perception?

Studies of the visual perception of 'stuffishness' thus far have primarily looked at the material properties of objects at rest. But the world is not static-- objects fall to the ground, popsicles melt, we deform fruits and vegetables with our hands to tell if they are ripe and ready to eat. Based on our interactions with (and observations of) objects in the world, we develop expectations regarding the expected outcome of an object in motion, based largely in part on its material qualities. We interact with objects and deform them with hands and tools, giving us an impression of material composition that is much stronger (and more reliable) than that of a static image alone. We expect dropped plates to shatter, popped bike tires to deflate, and gooey grilled cheese to hang between two slices of bread when pulled apart. Perceiving the shape and motion of an object, the deformations that arise from the collision of an object with a solid surface, and recalling the predictions that are associated with certain shapes and materials allows an individual to navigate the physical world, avoid danger, and remain alive.

Just as basketballs are necessarily made of rubber, the 'stuff' (see Adelson 2001 for a definition) that an object is made of (and its corresponding shape) indicates how an object will deform. Recent studies have shown that image motion significantly affects the perception of surface optical properties (Doerschner et al., 2011; Doerschner, Kersten, & Schrater, 2011; Marlow & Anderson, 2016; Oren & Nayar, 1997; Sakano & Ando, 2010; Wendt, Faul, Ekroll, & Mausfeld, 2010; Yilmaz & Doerschner, 2014). For example, Doerschner et.al. (2011) identified that specific motion cues (coverage, divergence, and 3-D shape reliability) are potentially utilized by the visual system to distinguish between moving shiny and textured matte surfaces that, when presented as static images, appeared identical. Such 'Material-from-Motion' (MfM) approaches have been previously only studied in the context of specular and matte bumpy spherical objects (Hartung and Kersten 2002; Doerschner et. al., 2011).

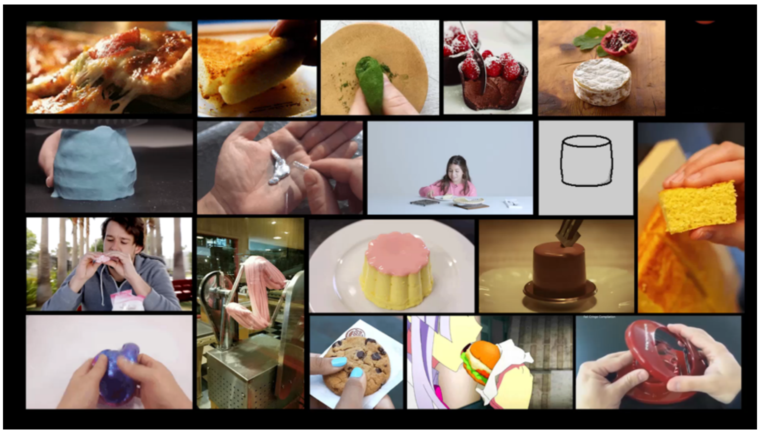


Figure 3. Materials-in-Motion. Studying the kinematics of materials is critical in material perception. Shown is a single frame taken from a multi-frame video. The percept of material is much stronger when observed in motion, than from a single static image alone. Taking pizza as an example, more information about the quality of the pizza is gained by watching the crust bend and the ease with which the cheese is pulled apart (than from a static image alone).

Motion can also be a powerful cue in the perception of rigid and nonrigid moving objects (see Figure 3). We see an object falling through the air and have expectations about the outcome of an object, should it come into contact with a rigid surface or another object. If a coffee mug is dropped from a height, it is likely to shatter into pieces. In contrast, if a metal spoon is dropped from a height, it will likely bounce, but remain intact. To navigate the world visually, we must rapidly classify the material of an object and make a judgement about whether and how to catch the object: you might jump out of the way to avoid a falling glass, dropped knife, or spilled liquid (if hazardous), but make little effort to avoid a dropped pencil. Fleming et. al., 2013 has shown that humans are able to rapidly classify the material of an object and make judgements about its material properties.

What motion cues do observers use when making judgements?

Deforming objects to provide cues to material properties is an intuitive behavior: Many videos show humans independently dropping or squishing (compressing) things to give the viewer an indication of material quality from motion (small cakes, soggy sandwiches, cooked steaks), suggesting that the motion and deformation of rigid and non-rigid materials plays a role in the identification of material composition. To predict outcomes of forthcoming actions in a real-world scene, we base our expectations about the motion of an object on our previous experiences with that object, as well as taking into account our knowledge about the expected size and weight of that object. How and where will an object fall? Although there has been a significant amount of work concerning the visual perception of objects and properties like size (Long, et. al., 2016) or weight (Gallivan et. al., 2014), a relatively small amount of work has been done to investigate the effect of motion on the perception of object material.

This deforming-objects phenomenon also extends to liquids (objects which do not have their own shape, but rather, take the shape of their container). In terms of viscosity (e.g. observing the velocity of a liquid) Kawabe et. al. (2015) finds that the critical feature for the perception of liquid viscosity is local motion speed, and that for impressions of liquidity, image statistics are related to spatial smoothness. This is in line with the findings of our experiments, as observing the motion of the deformation of the objects (in terms of speed and degree of spread) provide additional cues to the material composition of the object. Extending this to the perception of elasticity in cubes of translucent jelly, Kawabe and Nishida (2016) conclude that observers use information from the deformation of both the outline of the object, as well as motion resulting from optical deformations. Indeed, using deforming and bouncing cubes Paulun et. al., (2017) find that the magnitude of deformation is the dominant cue in softness perception. Given that observers make use of various cues when making material property judgments, what then is the influence of priors, expectation, and prediction?

Influence of Prediction and Expectation on Perception

We have described recognizing objects. We have described recognizing materials. We must also consider how *expectations* affect perception of objects and their materials.

Expectation is essential for perception in a number of ways. It allows us to quickly and efficiently disambiguate signal from noise when making a perceptual judgement (Bar, 2004), and has been shown to increase speed and accuracy when detecting visual stimuli (Posner, 1980). One of the most striking examples of the effect of expectation on object perception is Hansen et. al.'s 'blue banana'. In their study, Hansen et. al., (2006) showed that object memory (and by extension, the resulting expectation) affects the perceived color of objects: When asked to judge the color of a perfectly gray banana, observers rated the banana as somewhat yellow (despite a profound lack of color information), suggesting an influence of expectation (based on familiar object identity) on the perception of object color.

What is the relationship between existing expectations and reaction time? Summerfield and DeLange (2014) describe expectation as "changes in neural activity that occur before and during the presentation of a visual stimulus" and suggest that expectation, prediction, and the updating of a representation work together to navigate a visual scene. In tandem with this description, DeLoof et. al. (2016) and others have found effects of expectation on reaction time, in that if something is *expected* to occur in a location yields faster reaction times-- that is, if you are expecting a stimulus to appear in a certain location, you are faster to respond, given that your attention is deployed in the 'correct' (expected) location (to which you are already attending). Reliance on expectation and memory is critical for rapid processing. Pinto, et. al., (2015) have found that expectations facilitate entry of visual stimuli into awareness, in terms of both detection of the presence of a stimulus and identification of the stimulus identity. Heuristics in the visual world such as these are advantageous for rapid perception and categorization of items ('is this safe to eat?', 'Is that animal a threat to my survival?')

The effect of expectation also extends to other domains, and is especially interesting in terms of linguistics. When gaining meaning from a sentence, (Federmeier & Kutas, 1999) have found a phenomenon in event-related potentials (ERPs) where contextually unexpected words in a sentence elicit an N400 effect as a result of surprise due to a violation of expectation ("He took his coffee with milk and mustard" versus ". . . and sugar"). The violation of the expectation that was created by knowing that mustard is not usually combined with one's coffee generates such an N400 effect. The studies described in this thesis extrapolate this finding to material kinematics: While we did not conduct an EEG experiment as part of this thesis, we would expect to find the N400 effect as a result of our stimuli.

Ultimately, combining research on object representation, motion perception, material perception, and the top-down influence of expectation will allow us to gain a broader perspective on how the material properties of deforming objects are perceived. Furthermore, an analysis of eye movements will allow us to further investigate critical features and regions of an object that the human visual system takes into consideration when making such judgements.

The Role of Fixations in Material Property Judgements

Given the knowledge regarding object representations and the materials of which these objects are composed, the second half of this thesis asks where people fixate when making material property judgements, and specifically, if these fixations vary as a function of task/material judgement. If so, can a linear classifier predict which material judgement task the participant is performing? To investigate these questions, it is important to consider visual perception broadly, as well as factors that influence eye movements.

Factors Influencing Eye Movements (In-Brief)

Visual perception is the process of retrieving, integrating, and acting upon information from the visual world. In everyday life, we are able to effortlessly select the relevant information in a scene to successfully navigate the world, attending to certain regions of the visual scene without much conscious awareness (if any). This ability to rapidly select, attend to, and act upon relevant information in a scene (to the exclusion of all others) affords us the ability to survive.

Bottom-Up vs. Top-Down: In investigating how observers fixate a particular region of a scene, a number of questions and resulting theories have been proposed: Are eye movements simply deployed to the properties of scene features (bottom-up), or are they driven by higher-level, more top-down factors? Low-level visual features (e.g. color and contrast) have been shown to influence eye movements (Findlay, 1981; Zelinsky et. al., 1997), but recent research contradicts the idea that scene properties are the *only* factors driving attention within a scene. In contrast, recent studies have proposed a more utility-based approach, suggesting that a series of interacting control loops are what drives eye movements. This involves integrating the input of "saliency, object recognition, value, and plans to saccadic target selection" (Schütz, et. al., (2011), as eye movements are predominantly goal/task oriented. If we accept that eye movements are predominantly task-oriented, how does attention and saliency affect where to attend in a scene?

Attention and Saliency: Carrasco (2011) describes attention as a selective visual process whereby the human perceptual system is constrained by a limitation on cognitive and brain resources and thus, the ability to process vast amounts of information. Competing for limited resources, the brain must selectively attend to various aspects of the visual scene for efficiency. In tandem, *visual saliency* is the tendency for the visual system to attend to features of a given scene which draw the greatest amount of attention (Itti & Koch, 2000), often defined via a Saliency Map. Itti and Koch define a Saliency Map as 'an explicit two-dimensional map that encodes the saliency or conspicuity of objects in the visual environment' (Itti & Koch, 2000). Such Saliency- map-based theories (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998) assume that attention is deployed to the location with the highest map activation. While a selection of research presents evidence that saliency map-based models account for a percentage of the variance in participant eye movements (Foulsham & Underwood, 2008; Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Underwood & Foulsham, 2006), Saliency Maps alone cannot account for the findings in the literature. Therefore, we must also consider the role that task and top-down object knowledge plays in identifying which areas of the scene are most relevant for the task at hand.

How are Eye Movements Directed as a Function of Task?

Static Scenes. In studies of static scenes, Yarbus (1967) recorded eye movements while observers viewed a painting and were asked to judge various properties of the painting (e.g. wealth/or social status of the people depicted in the painting). He concluded that the eye movements had different trajectories depending upon the task that the participant was asked to perform. In tandem with this finding, Tatler (2010) successfully replicated the results of Yarbus (1967), and further investigated the effect of task instructions on eye movements. They asked observers to examine a painting ("Girl From the Volga"), seven times, providing a range of seven different sets of instructions each time (e.g. "make a series of judgements about the scene"/" remember aspects of the scene"/ "look at the picture freely"). They found that altering the instructions given to the observer significantly impacted their goal-directed eye movements.

Such an 'attention for immediate task'-based approach (otherwise termed 'what you see is what you need' (Triesch, et. al., 2003)) proposes that the information useful for a task is sampled "just in time" (Ballard et. al., 1995), selecting what information can be obtained at any given timepoint. In a real-world food-preparation task, Land and Hayhoe (2001) conducted eye tracking studies while participants made tea or a peanut-butter and jelly sandwich, as a way to investigate the role of eye movements during real-world actions and task completion. They found that fixations on an object during food preparation tasks were directed toward specific objects relevant for each subtask (knife, lid of peanut butter jar, etc.). These fixations were found to vary as *a function of* the task/sub-task being performed, and could be classified into segments (locating, directing, guiding, and checking). Inputs to the task at hand (identifying the object, identifying the location) as well as corresponding fixations, suggested that eye movements are directly influenced by the goal or sub-goals to be performed. But how are these task-goals additionally influenced by expectations and scene priors?

Additional Influence of Expectations on Fixation Behavior ("Scene Grammar"): Consistent with top-down approach previously discussed, it is also important to consider the effects of top-down influences on eye movements in scenes. Studies of "Scene Grammar" (the semantic content and arrangement of familiar objects in scenes, see Drashkow, et. al., (2017)), demonstrate that eye movements are not only affected by the task the participant has been asked to perform, but also by the expectations set forth by the statistical regularities of the world: Birds fly in the sky, benches sit on the ground, and the sun at sunset (illumination source position) can be found at the horizon. As a result, eye movements tend to be drawn toward regions of the visual field relevant for the selected information. This phenomenon has been discussed not only in psychophysics literature, but also in that of imaging and event related EEG responses-- In recent fMRI-based literature, Kaiser and Cichy (2018) found that typical visual-field locations of objects in scenes facilitate cortical processing, and find that fMRI decoding is more reliable when objects are presented in their typical locations in the visual field. In EEG literature, the N300/N400 ERP response has been shown to occur not only for semantic processing in sentences (Federmeier & Kutas, 1999) but also for the holistic processing of objects in scenes: Drashkow et. al. (2018) have shown that the N300 and N400 effect can also be found when processing objects in inconsistent locations in visual scenes (e.g. finding toilet paper in the dishwasher). These findings support the hypothesis of high-level cognitive processes on eye movements.

Taking these findings together, we aim to investigate how the above-described factors affect eye movements in the material perception domain, and ask how these factors influence eye movements when making material perception judgements.

Eye Movements in Material Perception Tasks :

Material Perception Tasks: In early studies of material perception, Sharan et. al. (2008) conducted a series of experiments to investigate the role of fixations in material property judgements of albedo and gloss. Using the theory that observers may fixate on regions of high contrast or skewness when making gloss judgements, they asked how fixations differ when making surface property judgments, versus shape judgements. They conclude that observers look in different places during shape and material perception tasks, finding that these eye movements were “non-random, correlated with one another, and similar for both tasks” Sharan et. al. (2008). This suggests that while regions of objects may bring about eye movements during both shape and material judgment tasks, these regions cannot be predicted by low-level factors (mean luminance, local contrast, local skewness or local energy).

Using lightness and stiffness judgements, we extend these findings in the material perception domain, and ask how fixations differ as a function of task. As this question has been broadly investigated via studies of lightness, we utilize a similar paradigm to investigate these questions.

Eye Movements for Lightness and Stiffness Estimation

Lightness perception. Lightness perception necessarily entails the estimation of the reflectance of a surface. Such estimation by the visual system involves the processing of the luminance distribution from a surface. While lightness estimation is arguably not a frequent *independent* judgement made in interacting with the world in daily life, the visual system routinely makes use of this information in order to make basic assumptions about objects in a scene. Certain regularities about the world (e.g. the ‘light-from-above’ prior) allow us to make basic, consistent predictions about objects in scenes, and interesting optical illusions occur when these priors are violated (‘light from above’ illusion, see Ramachandran, 1998). Toscani et al., (2013a) conducted an experiment to investigate how observers fixate when making lightness judgements. They found that when asked to judge surface lightness (and were permitted to freely fixate), participants tend to focus their fixations on the brightest parts of the surface. When participants were constrained in their fixation location, however, assessments of lightness varied more widely, suggesting that observers are using an efficient sampling strategy for judging lightness in scenes.

Eye Movements in DYNAMIC Scenes. These investigations are sufficient for considering fixations on static images; however, the world is not static. Objects in our visual field move and change, and we must be able to track moving objects in order to decide whether to interact with (or rapidly avoid) them. While the influence of static scenes on gaze behavior has been widely investigated, there is a growing body of research which suggests that dynamic components of scenes have a strong effect on fixation behavior (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Itti & Koch, 2000; for review, see Schütz, Braun, & Gegenfurtner, 2011). The present thesis will attempt to investigate these questions broadly.

Combining lightness judgements and dynamic scenes, Toscani et. al. (2016) asked participants to make lightness matches under dynamic illumination conditions for a moving (rendered) 3D object. Tracking a rendered 3D object, participants made lightness matches on a moving target. The authors concluded that the perception of the 3D object differed as a function of both the direction of motion and the position in the light field. For lightness constancy, they observed an increase in lightness constancy ability under free-viewing conditions (compared to trials where participants were forced to fixate). From these findings, the authors concluded that as dynamic scenes and non-uniform light fields pose a significant challenge for the visual system, eye movements are directed meaningfully to improve lightness constancy.

Recent research Smith and Mital (2013) asked whether viewing task influences gaze during dynamic scene viewing. As previous work has found that gaze tends to cluster near regions of greatest motion ("attentional synchrony", Smith and Mital, 2013), Smith and Mital asked observers to free-view videos of real-world scenes, or identify from where the video had been filmed. Using static scenes as a control, they found that free-viewing lead to greater attentional synchrony, longer fixations, and increased gaze to people and areas of high flicker (compared to the 'identify' task). The authors suggest that task has a significant influence on gaze behavior, as gaze clusters near the region of greatest motion. Given this knowledge that eye movements are goal directed (and that they are directed toward different regions of a scene based on task), can we infer the task that the subject is performing, using fixation positions?

Inverse Yarbus: Inferring Task From Eye Movements

Inverse Yarbus: Inferring Task From Eye Movements: As previously discussed, Yarbus (1967) investigated how fixations differ as a function of task. An inverse Yarbus process, then, would be to infer the task from eye movements. This could be applied in the case of webpage design, to indicate what information a user is trying to acquire. But whether this can be done is controversial. In 2012, Greene, Liu, and Wolfe tried to train humans and machines to solve an inverse Yarbus using scan paths, but found that they could only infer the task at chance level. If this could be accomplished, investigations could extend beyond basic summary statistics like the number of fixations, mean fixation duration, mean saccade amplitude, or the portion of the image covered by fixations. In 2014, Haji-Abholhassani and Clark developed a probabilistic method to infer the visual task of the viewer. Drawing from concepts from Hidden Markov Models (HMM), they were able to successfully predict the coordinates of fixation locations given the task the observer was asked to perform. Can we extend these findings in the material perception domain?

Fixations in Dynamic Scenes with Material Property Task Judgements: Present Study

Taking these findings into consideration, we created a simplified set of stimuli with stiffness and lightness tasks that would allow us to investigate 1) Where participants fixate when making kinematic (dynamic) material property judgements; and 2) (In contrast to the finding of Greene et al (2012) and consistent with Inverse Yarbus), can a linear classifier reliably predict which task a participant is performing, given the fixations as input?

Ultimately, this thesis attempts to tackle questions regarding the interaction between material, motion, shape, and top-down priors on material property judgements, and asks how fixations during such material judgements differ as a function of task. We utilize a violation-of-expectation paradigm to investigate whether materials and kinematics are part of the representations of familiar objects, and, if so, if we can measure how typical a certain behavior is for a given familiar material/object. We ultimately suggest that the perception of material properties is heavily influenced by a number of different factors, and that the human visual system efficiently makes use of the available information to successfully navigate the world as a result of these judgements.

Publication Notice:

In this thesis, I describe the first 8 experiments conducted during the course of the Dr. rer. Nat. work, which seek to investigate the perception of object kinematics. Portions of the studies described in Chapter 2 have been posted to biorXiv, and are currently under review at the Journal of Vision for publication.

Chapter 2: Surprise Motion Experiment

Portions of text contained within this chapter have been posted to biorXiv as a preprint and are currently under review at the Journal of Vision for publication. This text should be cited as (Alley, Schmid, and Doerschner, 2019). * Denotes shared first authorship.*

Experiment 2.1: Surprise Motion: A Pilot Experiment

In this study, we use a violation of expectation paradigm to investigate how predictions about object deformations based on object knowledge influence how we perceive the material of an object. In order to manipulate object familiarity, we rendered two types of objects: Familiar objects, for which there exist strong predictions about their mechanical material properties, and Novel, unfamiliar objects, for which few strong predictions should exist. For each of these object types, we rendered two motion sequences that showed how objects deformed when being dropped on the floor: an ‘Expected’ sequence, in which objects’ deformations were consistent with the observer’s prior beliefs about the mechanical properties of the material, and a ‘Surprising’ sequence in which they were not (see Figure 4). The Novel objects inherited their optical and motion properties from a corresponding Familiar object. Participants rated the objects in these movies on various material attributes. While we expected to find substantial differences in ratings and reaction times between Expected and Surprising events for Familiar objects, we also expected to find attenuated differences for Novel shapes. Note that we expected *some* differences between Expected and Surprising deformations also to occur for the unfamiliar, Novel shapes due to their optical cues alone and/or the fact that they all were bounded 3D objects, which might have elicited a prior expectation (e.g. a rigidity prior) about these objects’ mechanical material properties. To investigate the interaction of shape, material, and motion, we created and rendered a basic set of stimuli on which to test our initial hypotheses of reaction time and object familiarity effects.

Methods

Stimuli

Three Categories of Shapes

Three categories of stimuli were created in the open-source software Blender 2.77a (Stichting Blender Foundation, Amsterdam, NL): Six Familiar Objects (Figure 4a), Novel Objects (Figure 4b), which are globally convex, ‘noisy’ three-dimensional objects (Phillips, 2004), and Regular Objects (platonic solids, cylinder, sphere, torus), Figure 4c. The Familiar Object condition consisted of objects that people frequently encounter on an everyday basis (chairs, cloth items, spoons, etc). From repeated exposure to (and interactions with) these objects, observers are likely to have experience with their functions and utilities, and so an expectation should exist about how these objects behave and deform.

The 3D meshes for the chair, glass, and spoon objects were obtained from Free3D (see References) and TurboSquid (see References); we created the meshes for cloth, honey, and jelly objects by hand. Novel Objects were novel shapes that observers had never seen before. Thus, we expected that observers should have no expectations about how a Novel Object would deform when it hits the ground. The Regular Objects, included as a second control condition, are familiar, but we did not expect a strong association between the 3D shape and the object's kinematics, as these objects do not have specific affordances (Gibson, 1979). Each object subtended on average 4.49 degrees of visual angle during the first frame.

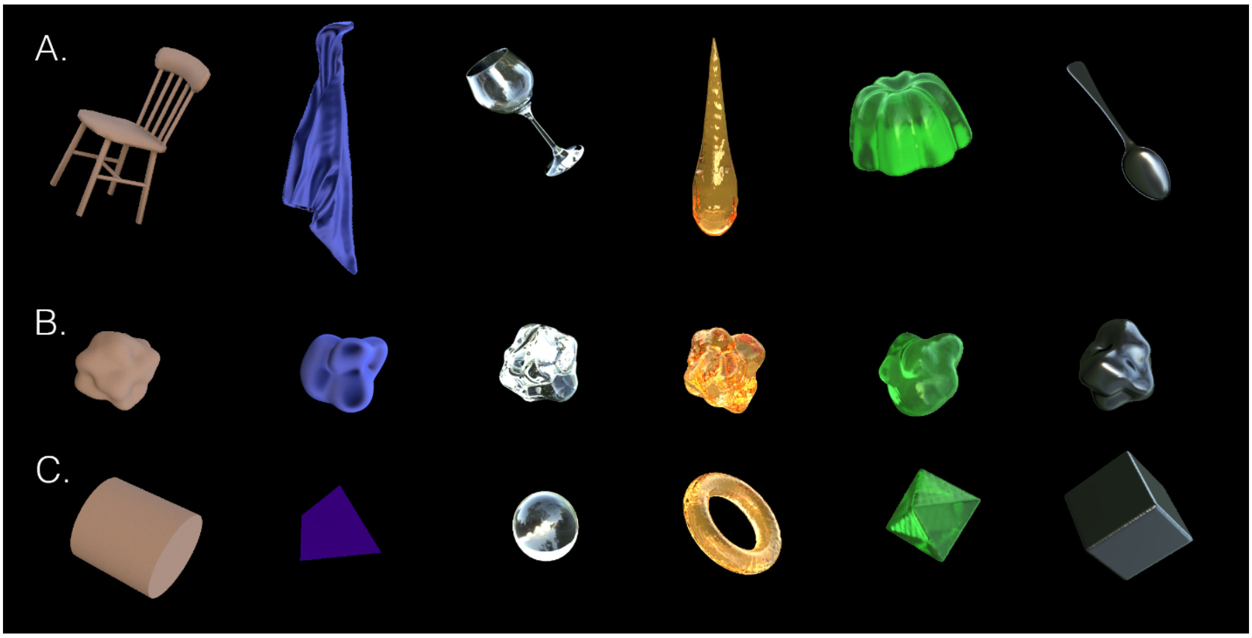


Figure 4 a-c. The six stimuli used in the Pilot Experiment. Row A shows the six rendered Familiar Objects. The optical properties of these six objects were matched with a Novel Object (row B), and additionally matched with a Regular Object (row C). For Novel Objects, we recycled the same six shapes but rotated them to a different view for each condition. *Animations contributed by A. Schmid/Figure created by L. Alley.*

Material Properties

Novel Objects and Regular Objects (each) inherited the material properties from a corresponding familiar object (Figure 4) - which included optical (BSSDF) as well as kinematic properties. For example, the Chair Novel Object would look brownish and fall rigidly, while the Green Jelly Novel Object would look translucent greenish and wobble after impacting the ground.

Animations

We expected observers to have some prior knowledge about the typical behavior of everyday objects and liquids when they fall on the ground. Upon impact with the floor, familiar objects (and corresponding control objects) deformed in a manner that was either Expected (e.g. a *wrinkling* cloth) or Surprising (e.g a *shattering* cloth) relative to the observer's expectations, yielding two motion conditions in the experiment: 'Expected' and 'Surprising'. We selected the surprising motion and object pairings in an effort to maximize surprise. Figure 5 gives an overview of how each object deformed in the two motion conditions. Note that the *Expected* condition for Novel Objects and Regular Objects is not expected to generate a surprise effect. Still we keep the terminology in order to be able to match up the results with the corresponding motion condition for familiar objects. All scenes were illuminated with the 'Campus' HDR probe (Debevec, 1998) and rendered with a fixed camera position. 36 two-second long animations (6 Familiar Objects, 6 Novel Objects, 6 3D shapes) consisted of 48-frame videos depicting rendered objects falling and impacting a virtual surface, with black background and floor. Exactly how each object would deform was determined 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).

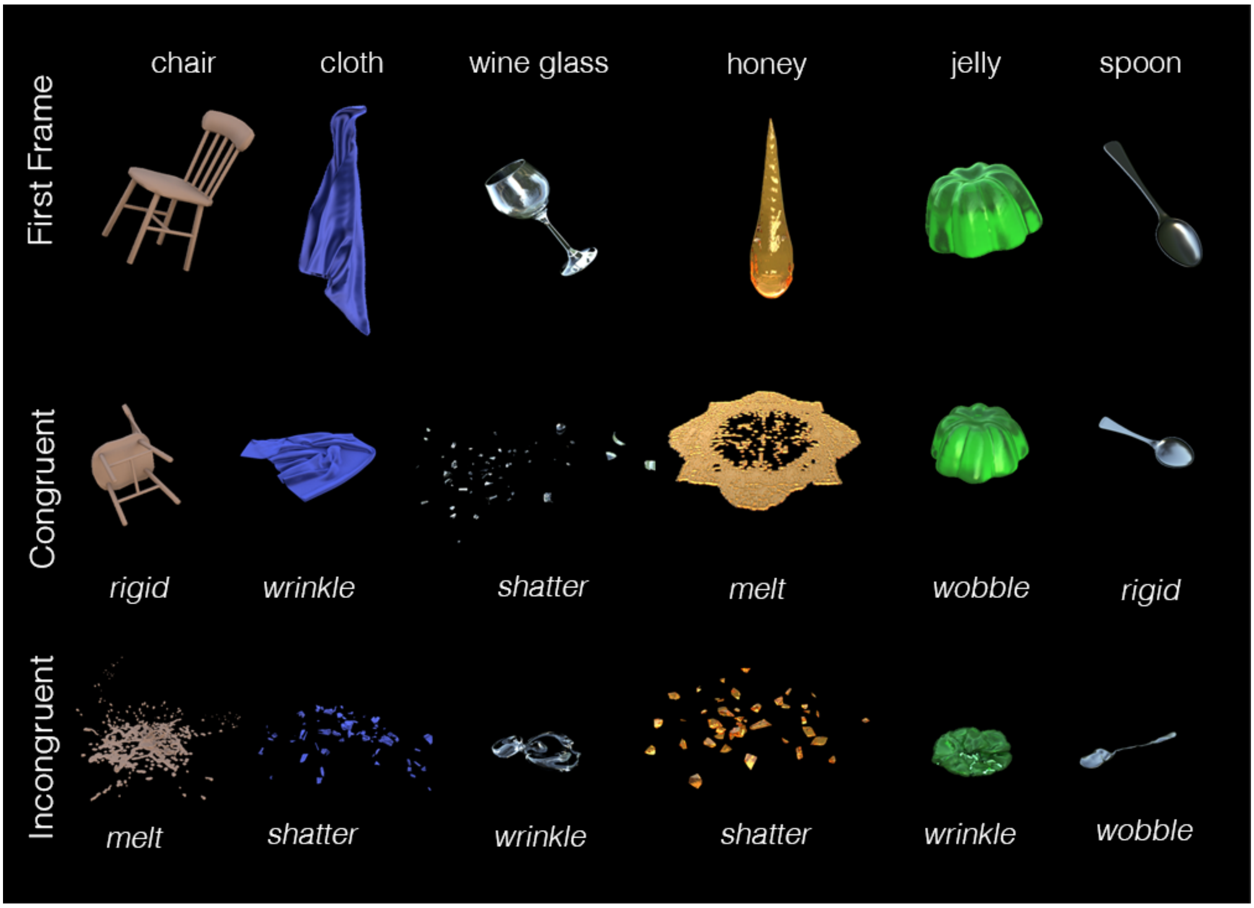


Figure 5. All Familiar Objects and their matched Expected and Surprising outcomes. Labels indicate the deformation method for each object. Here, the labels 'Congruent' and 'Incongruent' correspond to congruency with expectations: 'Expected', 'Surprising'. *Animations contributed by A. Schmid/ Figure created by L. Alley.*

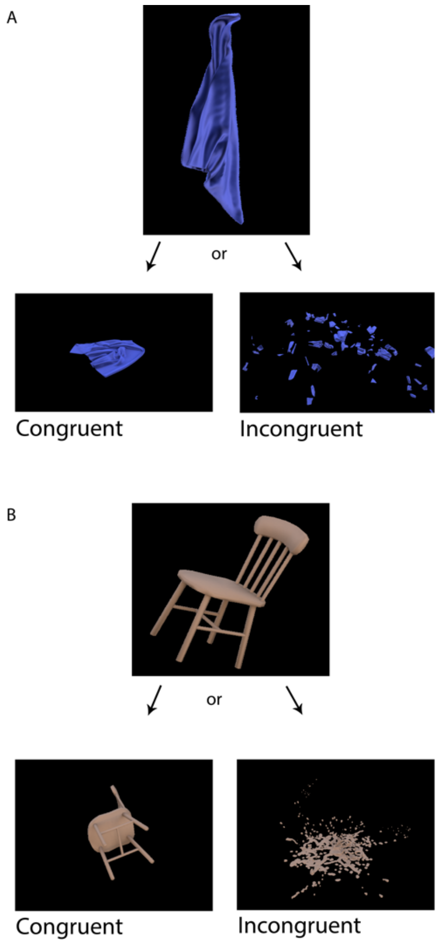


Figure 6a-b. Outcomes of Objects. Two examples of the expected/surprising outcomes of the Familiar Objects. Here, the labels 'Congruent' and 'Incongruent' correspond to congruency with expectations: 'Expected', 'Surprising'.

Apparatus

The experiment was coded in MATLAB 2015a (MathWorks, Natick, MA) using Psychophysics Toolbox (version 3.8.5), 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

An instruction screen was presented prior to the start of the experiment. Participants were asked to rate the stimuli with respect to four attributes (one per block): 'Hard', 'Gelatinous', 'Heavy', and 'Liquid'. Before each block, an explanation of the question posed to the participant was presented on-screen (e.g. "Rate how HARD each object looks. A setting of zero means not at all hard (soft). A setting of Max means that the object looks extremely hard."). Observers were instructed to be as accurate, but *also* as fast as possible when making their response.

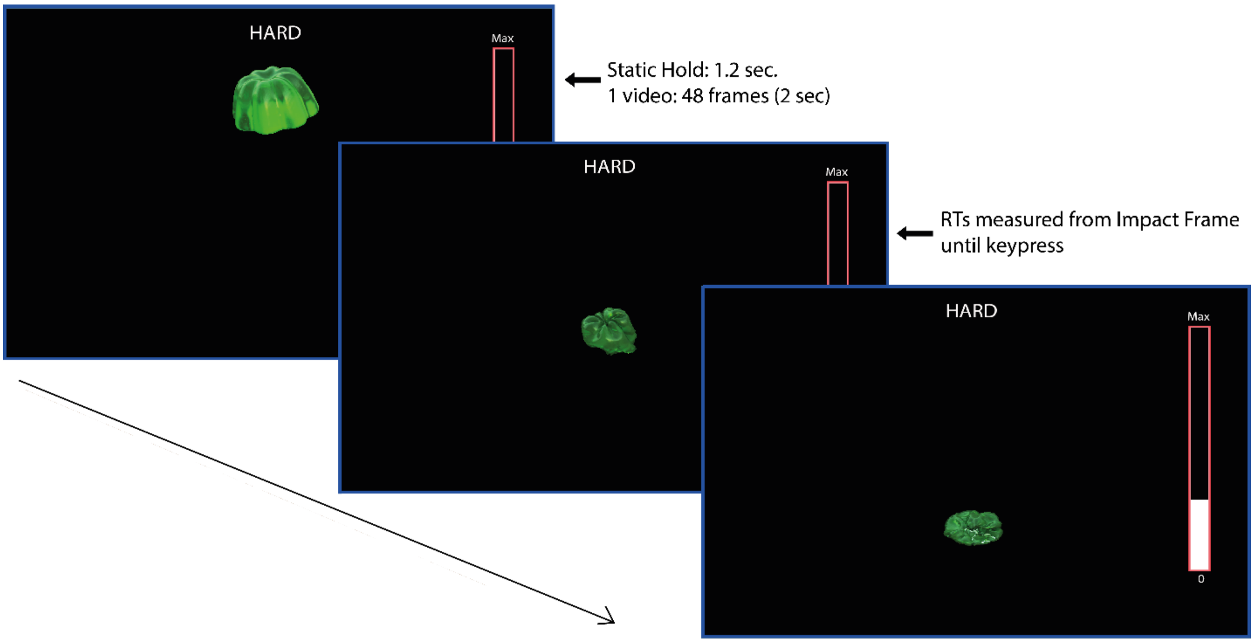


Figure 7. Example of a single trial. The first frame was shown for 1.2 seconds. The object then fell and deformed in a manner that was either Expected or Surprising relative to observers' expectations. The participants adjusted the bar at the right to rate the object on the given attribute, which was present throughout the trial at the top of the screen.

On every trial, an object (Familiar, Novel Object or Regular) appeared and remained still for 1.2 seconds to allow the participant to identify the object and to potentially activate corresponding expectations (see Figure 7). The object then fell and impacted the floor, deforming either as expected, or in a surprising way that violated participants' expectations. The participants indicated their response by using the mouse to adjust a visual rating bar (zero to maximum) and a response could be made at any time point. A press of the spacebar would lock in their response and proceed immediately to the next trial. Participants were allowed to take breaks between blocks. The experiment contained 144 trials in total (2 motion conditions $\times$ 3 object categories $\times$ 4 rating blocks $\times$ 6 object identities). The order of stimuli in a block was randomized, but participants completed blocks in the same order ('Hard', 'Gelatinous', 'Heavy', 'Liquid').

Participants

Ten naïve participants from JLU Giessen (mean age 30.9; six female) participated in Experiment 1; nine were right-handed. All participants had self-reported normal or corrected-to-normal vision. The experiment followed the guidelines set forth by the Declaration of Helsinki, and participants were naïve as to the purpose of the experiment. Five native German participants completed the German-language version of Experiment 1; all others completed the English-language version of the experiment, regardless of native language. Nine participants were experienced observers who participated on a voluntary basis; one participant was reimbursed at a rate of **€**8/hour. All participants provided written informed consent.

Analysis

Each object was held static for 1.2 seconds. The impact time point varied between objects, but occurred as early as .45 seconds after the first frame. Provided an observer was already happy with the setting of the rating bar and would only press 'space bar' to record their choice, it would take approximately .75 seconds. Thus, any response that occurred before 2.4 seconds (1.2 seconds + .45 seconds +.75 seconds) was discarded (four cases). We then computed mean and standard deviation of the reaction time data across all conditions, and excluded RTs that were longer than 2 standard deviations of the mean (62 cases, about 4.3 % of the data). We computed a sign test at an alpha level of .01 on the RT distributions, comparing median RTs for Surprising and Expected conditions. To measure how surprising a motion outcome was, we averaged RT data across observers and computed *reaction time difference scores (RD)* for every condition:

RD= Unexpected outcome RTs - Expected outcome RTs. **Equation 1**

Results

Reaction Times

Overall, - and as expected - reaction times were longer for the Familiar Object Unexpected motion condition than for any other condition. In Figure 8, we see that the corresponding histogram looks slightly more platykurtic than any of the other histograms. Sign tests between the Expected and Surprising RT distributions revealed that the increase in RT for the Familiar Object condition was significant ($p<0.0001$), but not for the two control conditions (Novel Objects ($p=0.897$), Regular Objects ($p=0.948$).

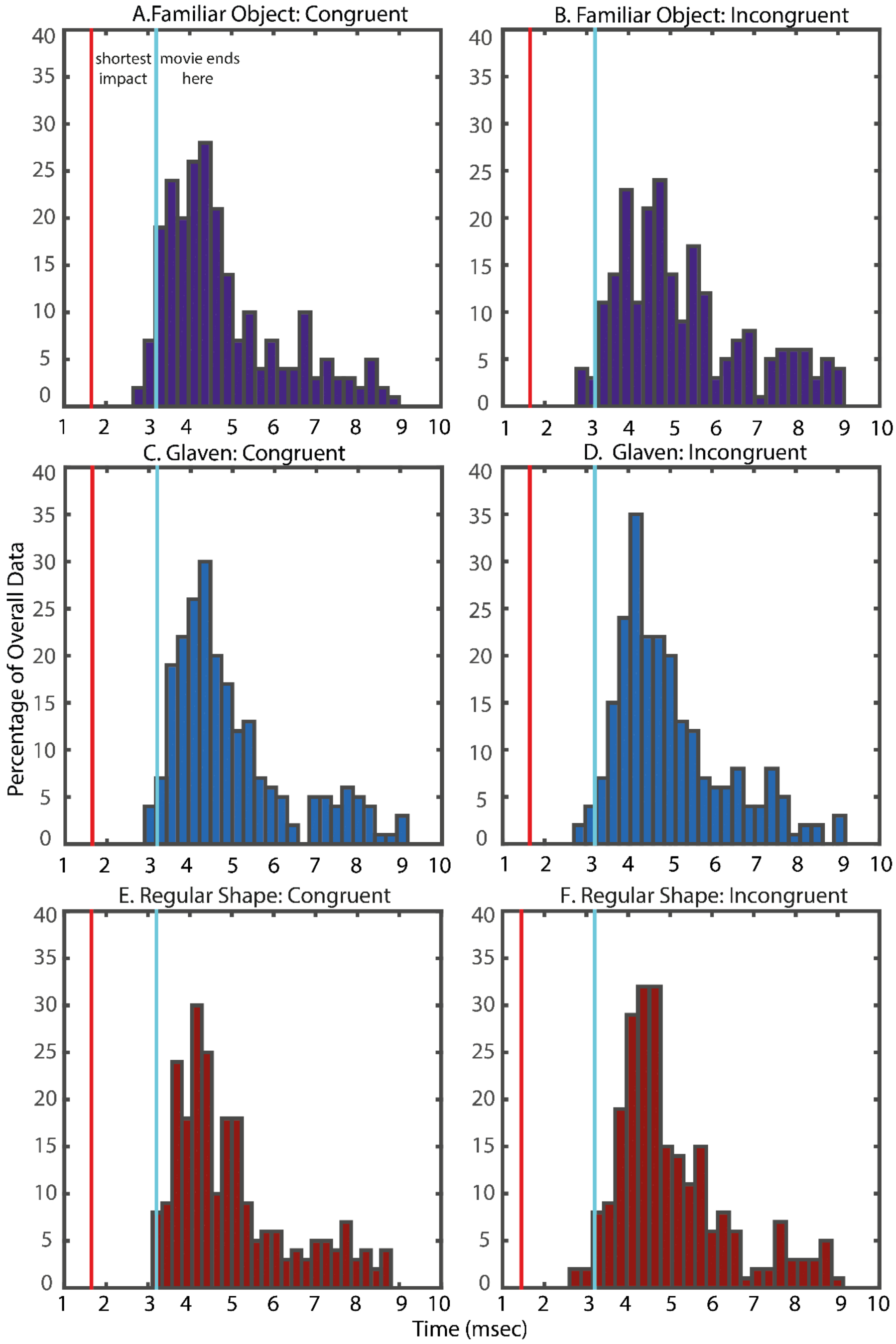


Figure 8 a-b. Histograms of all combinations of object type and motion condition. Figure B shows that, consistent with our hypothesis, reaction times were longer for the Familiar Object Surprising motion condition than for any other condition. (Here, the labels 'Congruent' and 'Incongruent' correspond to congruency with expectations: 'Expected', 'Surprising'.)

Reaction Time Differences

Although participants were asked to rate the materials of these objects using the four rating attributes, we were primarily interested in reaction time as a measure of how surprising the outcome was. Figure 9 plots the reaction time difference scores for all objects and rating attributes. Positive values of the RD indicate responses that were slower for the Unexpected condition than for the Expected one (e.g. responding to a shattering droplet of honey slower than a dripping one). Conversely, negative values on the RD indicate responses that were faster for Unexpected than Expected motion (e.g. responding to a melting chair faster than one that fell over rigidly).

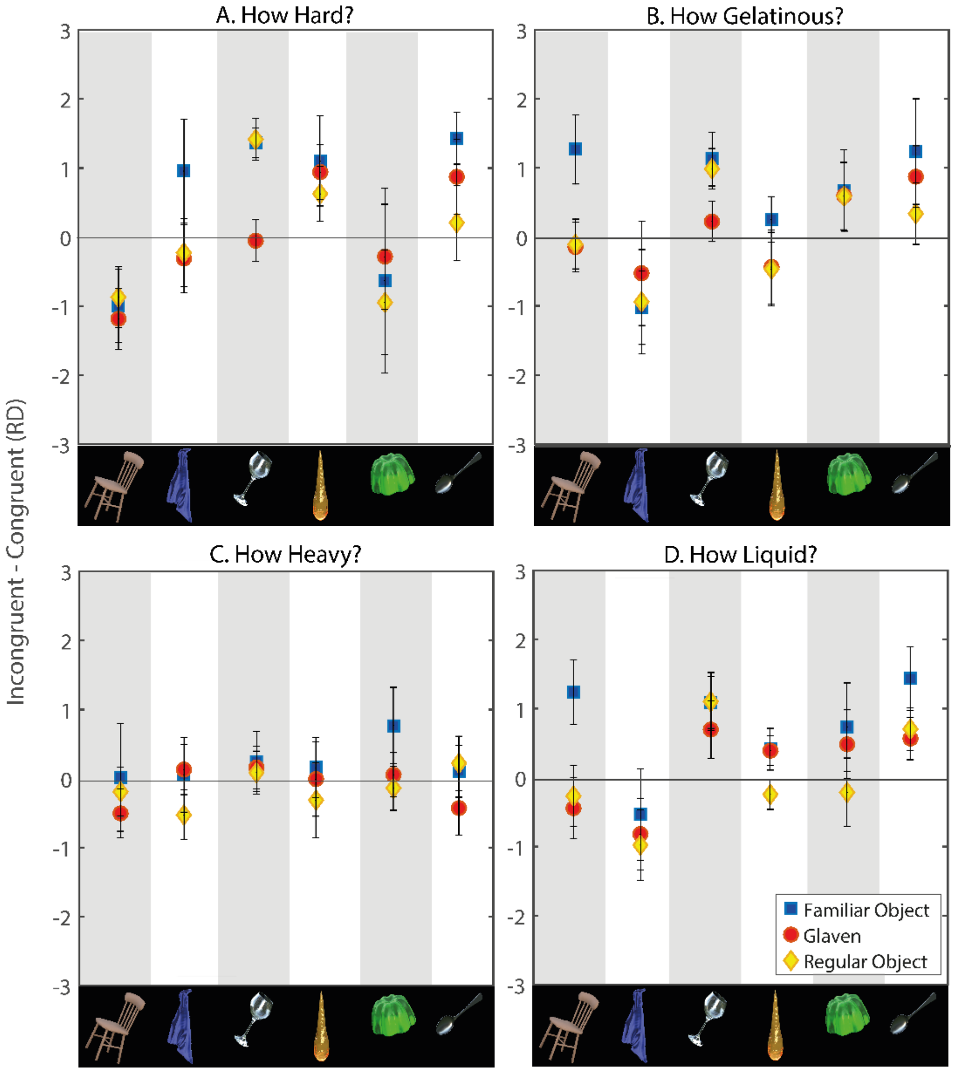


Figure 9 a-d. Reaction Time Difference for each object identity. Presented separately for each rating question. Errorbars are 1 Standard Error of the mean. Here, the labels ‘Congruent’ and ‘Incongruent’ correspond to congruency with expectations: ‘Expected’, ‘Surprising’; ‘Glaven’ corresponds to ‘Novel Object’.

In Figure 9 we see that for most Familiar Objects (blue squares), observers responded more slowly when outcomes were surprising, and faster when outcomes were expected. This can be seen by their positive reaction time difference scores. In contrast, the RD scores of the Novel Object and Regular Object control conditions (Figure 9 a-d) were lower. This was confirmed with a sign test that showed that Familiar Object RDs were greater than both Novel Object and Regular Object RDs in 19 out of 24 conditions (p<.01).

Ratings

Participants were rating objects sensibly and agreed with one another. Figure 10 shows a selection of object identity and rating question combinations. Familiar Objects that fell rigidly to the ground were in fact rated as more rigid (i.e. a rigid spoon is rated *harder* than a wrinkling one (Figure 10a); a wobbling jelly is rated *more gelatinous* than a wrinkling one (Figure 10b), and a melting chair is perceived as *more liquid* than a solid one (Figure 10c)). This suggests that participants use all provided motion information when judging materials. However, some characteristics, e.g. heaviness (Figure 10) were hard to assess for observers, regardless of whether the motion was Expected or Surprising. Ratings across object categories (Familiar Objects, Novel Objects, Regular Objects) were quite similar, which suggests that observers based their rating primarily on the provided motion information and *not* on the shape of the object.

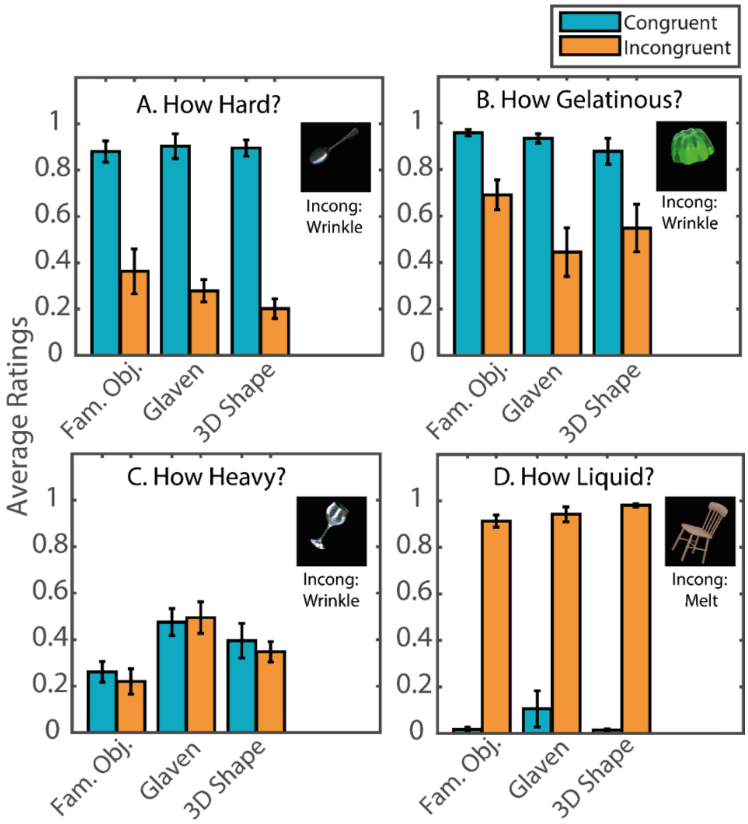


Figure 10 a-d. A selection of ratings. Ratings for Expected and Surprising conditions are shown for a subset of the stimuli. Here, 'Glaven' corresponds to 'Novel Object'.

Figure 11a shows that participant responses strongly correlated with one another, suggesting that the perception of the material qualities queried is similar across participants. This is further illustrated in Figure 11b, which shows that correlations between participants were high for three out of four attributes. Again, heaviness seemed to be difficult to visually assess for our stimuli.

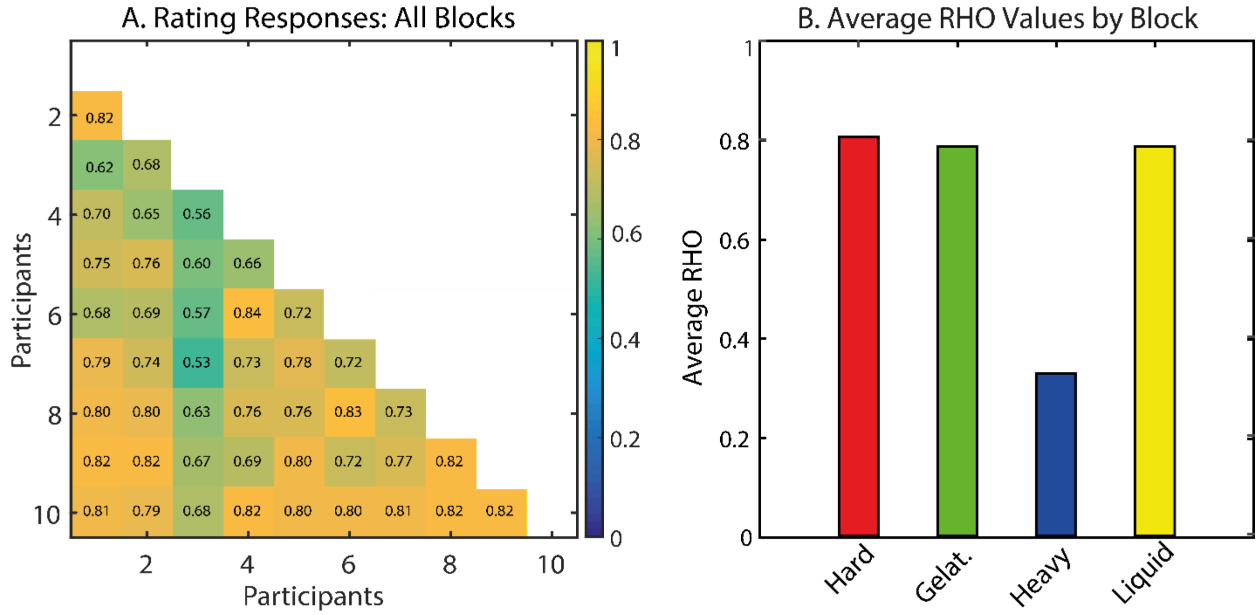


Figure 11 a-b. Correlation matrix. Average rating values given by all 10 participants**.** Here we show the correlation matrix of rating responses, with corresponding graph of average Pearson's RHO values by block (rating attribute).

Discussion

In the present pilot experiment, we created a novel set of stimuli containing six objects that were shown to fall and impact the ground, and deform on impact in a manner that was either expected or unexpected, given observers knowledge about familiar objects and their commonly associated materials. We used reaction time as a measure of the surprisingness of an object's behavior. For comparison and control stimuli, we developed two equivalent conditions of objects: Novel Objects (called Glavens) and 3D objects. For these conditions, participants should have limited expectations about how these objects will deform on impact.

Reaction Times. As predicted, we find that participants have slower reaction times when judging surprising outcomes of familiar objects, compared to expected ones. We did not observe such a difference in reaction time for the Novel Object and Regular Object conditions. We believe that for these object classes, observers did not have strong expectations about how objects should behave upon hitting the ground. Although the latter two classes are bounded shapes, it is unlikely that strong intuitions about kinematic behavior are present for these objects and their paired optical material properties. These findings may also be due to inherent expectations based on shape and material properties (e.g. the glass ball could have been interpreted by participants as a solid glass marble, and the cylinder a wooden dowel (neither of which would be expected to break upon impact)).

Object Representations. An object representation ('Object File') is a grouping of properties which contains everything we know about the features of an object. Existing literature describes that features like color, shape contours, size, and optical properties are part of an object representation. Referring broadly to Treisman and Kanwisher's (1998) six types of object representations, consider a green jelly: The shape and optical properties identify that it is a jelly, that it is green, that it has a prototypical size and shape, and a glossy surface. We are able to imagine the object from other angles and distances, we know that it belongs to the category of desserts, we may know what it tastes like, and we can represent affordances. These properties are all part of the object representation.

Extended reaction times to Surprising events is consistent with object representation literature, and would suggest that our Surprising outcomes 'violate' the existing object representation, necessitating an 'updating' of the object representation: Our results suggest that when an outcome is surprising (that is, violates our expectations), a longer amount of time is required to process the information presented, and perform an 'online' correction for the expectation that was violated. In tandem with this finding is the observation that ratings tended to increase between static first frame and motion, suggesting that the observation of motion affects the rating value overall. A jelly that was rated as gelatinous based on its shape and optical properties alone was rated as more gelatinous following the observation of the expected motion.

The current study is to our knowledge the first to demonstrate a reaction time effect with respect to expectations of kinematics and material qualities of Familiar Objects. It is possible that the rating judgement is made as a result of observing the amount of deformation involved, which requires the observation of motion, and cannot be seen from single static frames in isolation. The observation of velocity is also a critical factor—the amount of material, density, and speed of deformation all contain visual cues to material (e.g. something that falls more quickly is denser and therefore potentially more heavy). In terms of object files and the object representations they create, such a finding suggests interplay between shape and optical qualities/material, and may further suggest that kinematic properties of an object are contained within the post-dorsal stream integration (see Fattori, 2005) of the intermediate object representation.

Ratings: In this pilot experiment, we predominantly use the ratings to confirm that participants were rating the objects sensibly. Looking at patterns in the ratings, we find that Novel objects (Glavens) and 3D objects were rated similarly, suggesting (in tandem with the reaction time results) that these object classes do not elicit strong expectations regarding how an object will deform.

Limitations. This stimulus set was created as proof-of-paradigm, and as such, did not control for many low-level factors such as object size or degree of deformation/object splash size. This has the potential to negatively affect our results, as objects/object splashes that are larger may attract additional attention, and consequently, longer reaction times. In the Incongruent condition, *chair* and *cloth* began to deform prior to impact. The size of the objects (in degree of visual angle) varied between objects, as did the size and spread of the deformation (the Novel object's liquid substances had a large splash, whereas the splash of the 'Honey' Familiar Object was small). Due to observed difficulties in language comprehension for non-native English speakers, all participants in Experiment 2 were native speakers of German language, and the experiment was given entirely in German.

The experiments that follow are an effort to further investigate the interaction of optical properties, material, motion, and shape, properly control for additional factors, and to tease apart the influence of material priors on the percept of the kinematic properties of objects as a whole.

Experiment 2.2: Surprise Motion: Typical-Color Experiment

After confirming the hypothesis that objects that behave in *unexpected* kinematic ways elicit longer reaction times than those who behave in ways that do not violate basic laws of physics, we sought to investigate how the 'deformation method' of these Familiar objects (shatter, wrinkle, melt, etc). affect the perception of (and expectations about) these objects. To investigate this question, we created a second, more tightly controlled set of stimuli that allowed us to parametrically investigate differences between material categories. Unlike the previous set of stimuli, this set contains objects that impact the ground on the same frame, allowing us to use the point of impact with the floor as a critical time point in the analysis. This second set of stimuli additionally provided scene context and contained a counterbalanced set of 15 objects, containing three objects that behaved as expected for each category (see Table 1). The analysis that follows is an attempt to use these stimuli to investigate the relationship between expectations, shape, optical properties, and material kinematics.

Methods

Stimuli

Objects. We used two types of objects: Familiar and Novel. Figure 12a-c gives an overview of all Familiar objects used in the experiment. In order to create a stimulus set with 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, metal key, metal spoon), wobbling (red and green jelly, custard), and wrinkling (linen, velvet, and silk curtain). All objects were rendered in Blender (Blender 2.77a, Stichting Blender Foundation, Amsterdam, NL) with their 'typical' optical material properties, e.g. metallic-looking key, green transparent jelly, or a silky- appearing curtain). Objects were rendered in a 'room' with brown walls and a hardwood, mid-gloss floor, and were illuminated by the 'Campus' environment map (Debevec, 1998). The 3D meshes for the chair, glass, and spoon were obtained from Free3D (see References) and TurboSquid (see References); the remaining were created by hand.

Novel shapes ('Glavens', Phillips, 2004) inherited their optical and mechanical properties from the corresponding Familiar object (Figure 12a). Each familiar object was paired with a matched unfamiliar, novel-shaped stimulus (Figure 13) that was rendered with the same surface properties and material behavior. Consistent with the previous experiment, 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).

Scene: Following from the work of Bian & Andersen (2010) on the perceptual advantage of the presence of a ground surface in the representation of visual scenes, the stimuli used in this experiment included context via the presence of two rendered wooden side panels and mid-gloss wooden floor (see Figure 12 below).

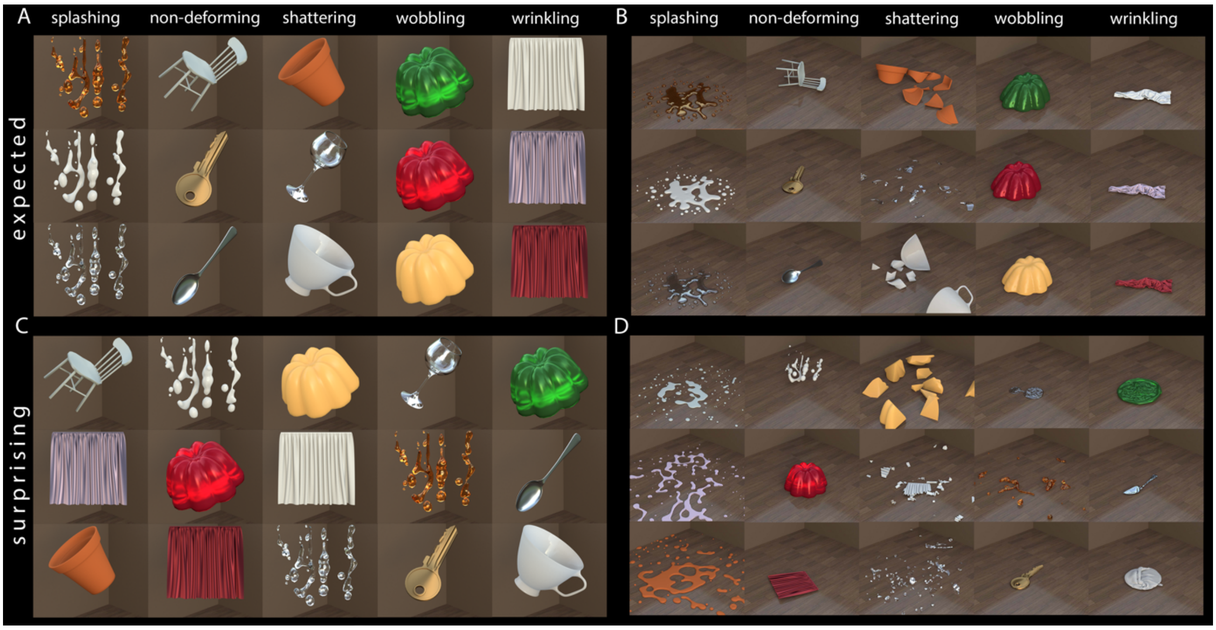


Figure 12a-c. Familiar Object Stimuli. Panel A shows Familiar objects grouped according to their material mechanics. Panel B depicts corresponding last frames of animations that show how a given object fully deformed. Panel C shows how each object will deform in the surprising condition, and Panel D shows the corresponding last frames of the surprising animation. We refer the reader to Zenodo (see References) for the corresponding animations. *Animations contributed by A. Schmid.*

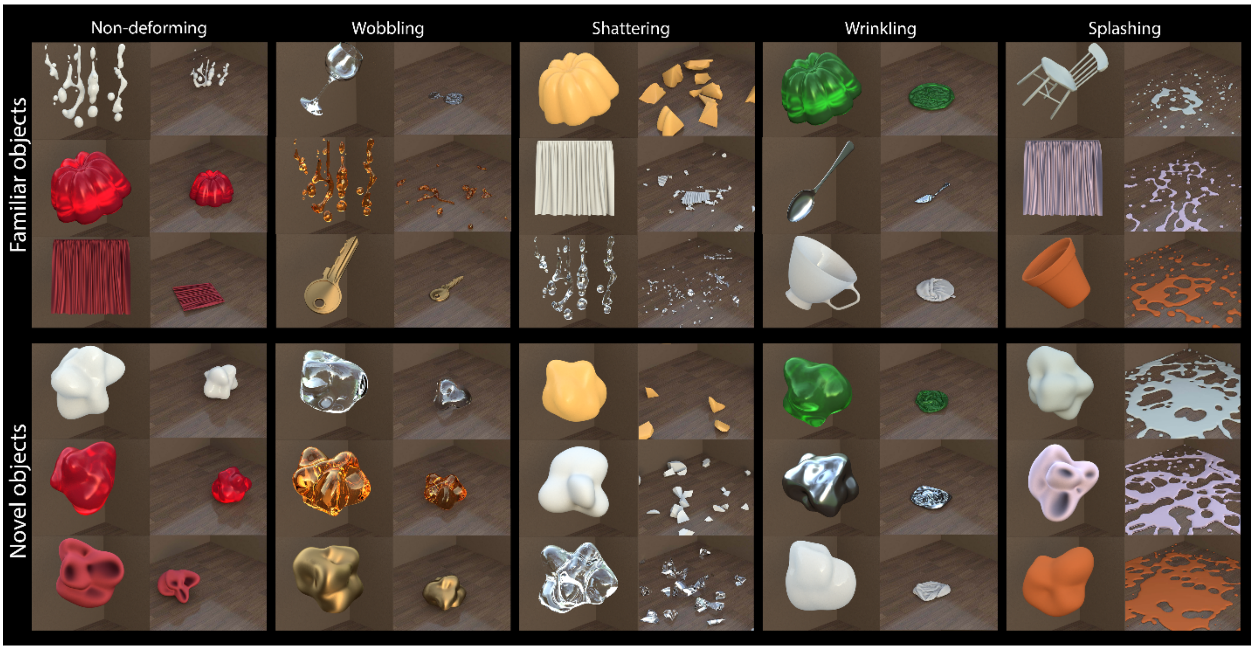


Figure 13. Novel Stimuli. The lower half of this figure shows corresponding Novel objects (also) grouped according to their material mechanics. The right-hand column of all categories depicts corresponding last frames of animations that show how a given object fully deformed in the Surprising condition. We refer the reader to Zenodo for the corresponding animations. *Animations contributed by A. Schmid.*

Deformations. For each object, we rendered short movies that showed how an object fell to the ground and how it interacted with the floor (Figures 12 and 13). In order to manipulate surprise in our experiment, an object could either behave as expected (e.g. a glass would shatter), or it could inherit the mechanical material properties from another object, e.g. a chair would splatter like milk upon impact (Figures 12 and 13). We created corresponding Expected and Surprising movies for Novel objects. Table 1 shows which objects inherited which material mechanics in the Surprising condition. All experimental movies can be downloaded on Zenodo (link in references).

	Splashing	Non-deforming	Shattering	Wobbling	Wrinkling
Expected	Honey droplets	Painted chair	Clay pot	Green jelly	Linen curtain
	Milk droplets	Brass key	Wine glass	Red jelly	Silk curtain
	Water droplets	Metal spoon	Teacup	Custard	Velvet curtain
Surprising	Painted chair	Milk drops	Custard	Wine glass	Green jelly
	Clay pot	Red jelly	Linen curtain	Honey droplets	Metal spoon
	Silk curtain	Velvet curtain	Water droplets	Brass key	Teacup

Table 1. Overview of objects and conditions in the experiment. Columns show the five categories of typical material mechanics (material class) in the experiment; rows show which objects occurred in the Expected and Surprising conditions. Every object in the Expected condition would also appear exactly once in the Surprising condition, where it would be rendered with very atypical mechanical properties, e.g. the silk curtain would Splash. To control for object familiarity, every cell in this table had a corresponding Novel object (Novel Object) paired with it. Novel Objects had no familiar shape, but inherited all optical and mechanical material properties from a corresponding Familiar object.

Animations. Each movie consisted of 48 frames, depicting an object suspended in air, which then fell to the ground. Impact occurred exactly at the 11th 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 (see Figures 12 and 13). 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 (MathWorks, Natick, MA) using the Psychophysics Toolbox extension (version 3.8.5, Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), 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 (while still maintaining accuracy). 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. After this the movie clip repeated two more times (without the hold at the beginning) if necessary. 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 14a). 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. 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 14a). "Mechanical" qualities, like hardness and liquidity, have been of increasing interest in recent studies of deformable materials and are likely to be directly estimated from material kinematics, whereas gelatinousness and heaviness might not be estimated directly from mechanical deformations and are potentially influenced more by associations with familiar shape and optics.

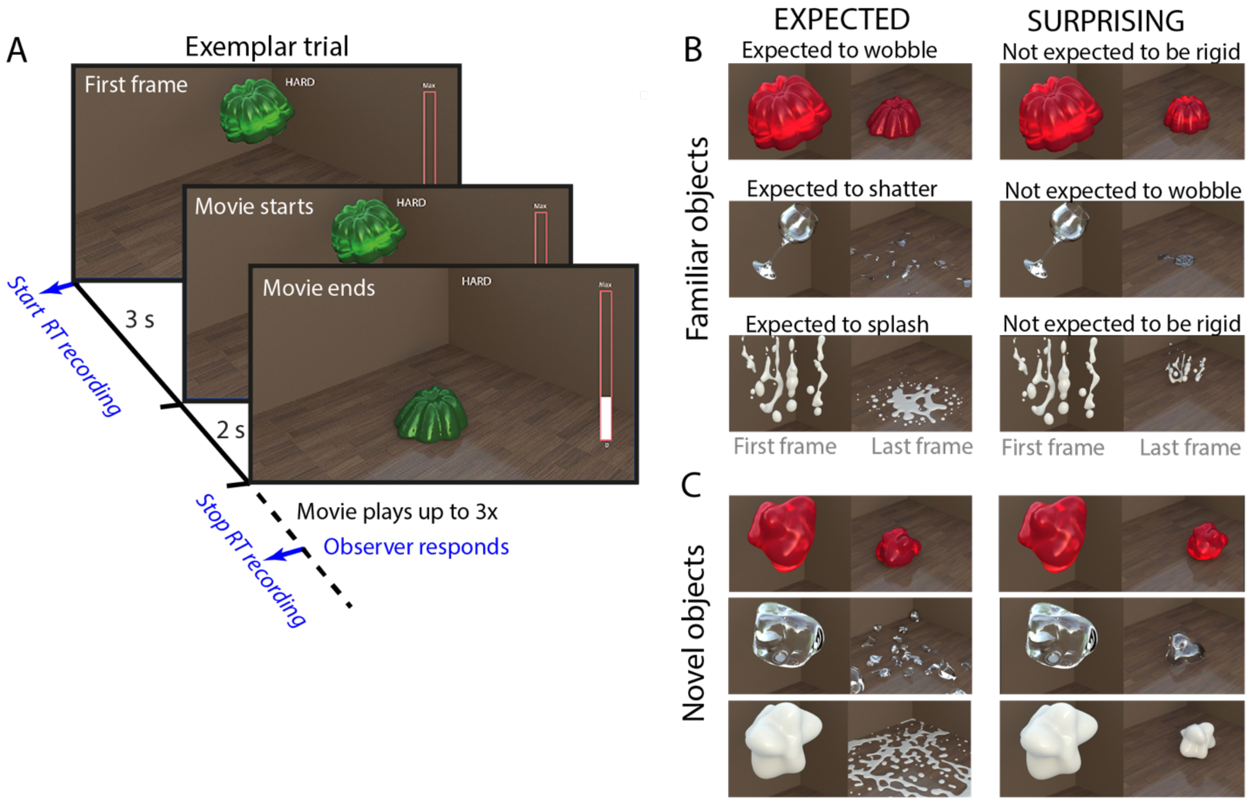


Figure 14 a-b. Trial and stimuli. A. An exemplar trial. **B.-C.** A subset of familiar objects (B) and corresponding novel objects (C) used in the experiments. Familiar objects could either behave as expected, or in a surprising manner. Note that this distinction (expected vs. surprising) is only meaningful 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). *Animations contributed by A. Schmid.*

Participants

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.

Static Frame Experiments: First Frame (FF) and Last Frame (LF)

Corresponding Static Frame Experiment for Analysis: We consider the First Frame of each video to be a measure of expectations based solely on optical material properties and shape (in the absence of motion). Analyzing the data by comparing full-motion data and first frame results will allow us to tease apart the influence of motion on the perception of object material. In order to perform such analyses, a separate group of observers were asked to perform the same rating task when presented with static images alone. The data that follows is used to compute indices comparing the nature of the interaction between the priors elicited by optical material properties and shape ('First Frame' information) and the material kinematic behavior. We *do not consider data and analyses from these experiments in isolation*, but rather, in the context of the relationship between this data and the corresponding motion analyses. We refer the reader to the analyses that follow in Chapter 2 for the corresponding analyses and results.

Here we describe the parameters and setup for the three time-point control experiments, which allow us to use different time points in the video, as different inferences about the objects can be made at each of these time points. These data form the basis of the equations described in Chapter 2, which are included in the Predictability Scores and Effect of Expectation Scores for data analysis. In this section, we are specifically interested in using the First Frame of the video as a measure of pre-existing object knowledge.

Methods

Stimuli

The stimuli used in this experiment were static images of the First Frame and Last Frame of each video, presented for five seconds (the same duration as one presentation of the video plus static hold duration).

Apparatus

The apparati used in this experiment were identical to that of the ones previously described in this chapter. We refer the reader to the previous experiment for these details.

Task and Procedure

The task and procedure used in this experiment were identical to that of motion experiment described previously. We refer the reader to the previous experiment for these details. Here, the participants rated only a single frame of each video (First or Last), rather than observing the full motion.

Participants

15 participants (mean age 26.40; 13 female) participated in the First Frame version of this experiment. 13 participants were right-handed and all had self-reported normal or

corrected-to-normal vision. Six participants participated in both the 'Full Motion' experiment described in this chapter and the First Frame experiment described here. In these cases, the First Frame experiment was always performed first to avoid the influence of motion on the perception of the static images.

15 naive participants (mean age 24.8; 10 female) participated in the 'Last Frame' control experiment; 8 were right-handed. All participants were native speakers of German language, and had self-reported normal or corrected-to-normal vision. All participants provided written informed consent, and were reimbursed at a rate of €8 per hour.

The experiments followed the guidelines set forth by the Declaration of Helsinki, and participants were naive as to the purpose of the experiment. All participants were native speakers of German language, and the experiment was given entirely in German. All participants provided written informed consent and were reimbursed at a rate of €8 per hour.

A Note on Results and Discussion

As the data and results from these experiments are used to compute indices comparing the nature of the interaction between the priors elicited by optical material properties and shape ('First Frame' information) and the material kinematic behavior, we *do not consider data and analyses from these experiments here in isolation*, but rather, in the context of the relationship between this data and the corresponding motion analyses. We refer the reader to the analyses that follow in this chapter for the corresponding analyses and results.

Analysis I: Full-Motion

Rating Differences Analysis: Expected versus Surprising

To measure the influence of object knowledge on perceived material properties, we computed rating differences between Expected and Surprising conditions. For each object type (Familiar, Novel), each attribute (how hard, how gelatinous, how heavy, how liquid), and each material class (Splashing, Shattering, Non-deforming, Wobbling, Wrinkling), rating differences were calculated between the three objects with Expected outcomes and the three (different) objects with Surprising outcomes (ratings were averaged over the three objects before difference scores were computed). Thus, how an object deformed was the same, e.g. it would splash, but which object (Familiar or Novel) would do the splashing is different in Expected (honey, milk, water) and Surprising conditions (chair, pot, curtain). We were agnostic about the direction in which the rating will change (e.g. will a Splashing chair look more liquid, or harder, or heavier than Splashing milk?), so we took the absolute value of this difference score as an index of the effect of expectation on ratings (Effect of Expectation Index, $\boldsymbol{\epsilon}$):

$$\epsilon = |\ \mathbf{avg.ratings\ Expected - avg.ratings\ Surprising}\ | \qquad \textbf{Equation 2}$$

This resulted in 40 difference scores for each observer, 20 in the Familiar object condition and 20 in the Novel object condition. If object knowledge influences perceived material properties, then ratings should differ between Familiar objects that behave as expected, and Familiar objects that behave in a surprising way. Familiar objects should be larger than $\boldsymbol{\epsilon}$ for Novel objects ($\boldsymbol{\epsilon}_{\text{Familiar object}} > \boldsymbol{\epsilon}_{\text{Novel object}}$). A binomial sign test was used to compute the likelihood of obtaining k or more instances in which $\boldsymbol{\epsilon}$ (averaged across observers) was greater for Familiar versus Novel objects.

Reaction times Analysis: Reaction Time Difference

Time taken to make each judgment (reaction time) was also 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. After this, we computed reaction time difference scores (τ_D) as follows:

$$\boldsymbol{\tau_D} = \mathbf{av.reaction\ time\ Expexted - av.reaction\ time\ Surprising.} \qquad \textbf{Equation 3}$$

Similar to the rating data, reaction times were averaged over the three objects in each material class before difference scores were computed, which led to 40 difference scores for each observer, 20 in the Familiar object condition and 20 in the Surprising object condition. These difference scores were averaged across observers, and a binomial sign test was used to compute the likelihood of obtaining k, or more instances in which the τ_D was greater for Familiar versus Novel objects.

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.

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 criterion (~1.3% too fast, ~4.7% too slow).

Results

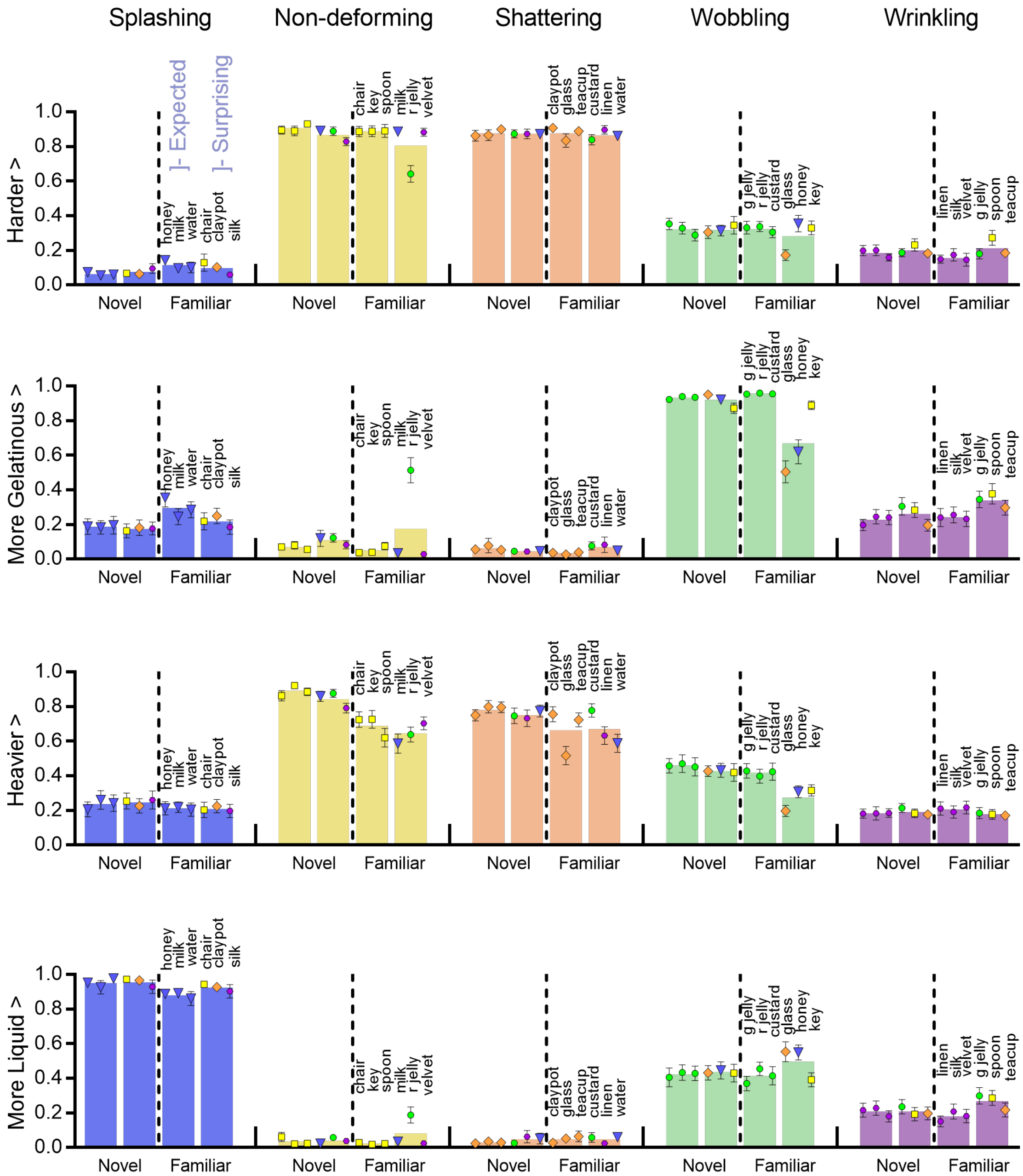


Figure 15. Ratings for all attributes. Error bars are one standard error of the mean. In general, ratings on all blocks 'make sense': splashing objects (blue bars) are rated as highly liquid, non-deforming objects (yellow bars) and shattering objects (orange bars) are rated not at all liquid, and wobbling objects (green bars) are rated in between. The directionality of some ratings is not straight forward on first view, when comparing

expected and surprising conditions, e.g. for splashing objects (blue bars). We might have predicted the splashing honey, milk, and water to be rated as less hard and more liquid than the chair, clay pot, and silk curtain that splash surprisingly – object knowledge about chairs, clay pots and silk curtains should "pull up" hardness ratings and "pull down" liquid ratings. However, we see the opposite pattern. Ratings of how gelatinous these objects appear may shed some light on this: splashing honey, milk, and water appear more gelatinous than all other splashing objects (even Novel objects). This suggests that the shape of the droplets in the initial frame do not generate compelling impressions of very runny liquids, but rather some thicker, more viscous substances (First Frame ratings in Supplementary Figure 5 confirm this idea). So, in the case of splashing stuff, object knowledge interferes in our Expected condition in a way that we did not anticipate. *Figure contributed by A. Schmid.*

Overall, the ratings of observers varied systematically with the mechanical material properties of Familiar and Novel objects. Figure 15, which shows average ratings for all attributes, illustrates that observers were able to perform the task sensibly. For example, non-deforming (yellow bars) and Shattering (orange) objects were rated as hard (top plot), and not gelatinous (second plot). Conversely, objects that splashed (yellow bars), Wobbled (green bars), and Wrinkled (purple bars) were rated less hard. Finally, objects that Wobbled (green bars) were rated as very gelatinous. Ratings for the other two attributes heavy and liquid varied systematically with the mechanical material properties of objects. Non-deforming and Shattering objects were rated as heavy, and Splashing objects as liquid.

There were a few deviations from this overall pattern, in that ratings for individual Familiar objects were substantially above or below their group mean, e.g. in the Surprising condition, hardness and gelatinousness ratings for the non-deforming red jelly tended to be lower and higher than the group average, respectively. Similarly, ratings for the wobbling glass and the wobbling key appear to be outliers. We discuss these cases below. Overall, Familiar and Novel objects tended to be rated similarly.

Following from an analysis of ratings, we sought an alternative method to compute the 'Expectedness' of our objects. By subtracting the ratings of the 'Surprising' outcomes from those of the 'Expected' outcomes, we are able to generate a measure of 'surprisingness', relative to all objects in the stimulus set.

Effect of Expectation Index (ϵ)

Figure 16 plots the Effect of Expectation Index (ϵ) for each attribute and each material class, averaged across subjects. In all but one condition (19/20), ϵ was greater for Familiar objects than Novel objects, sign test: $p<0.001$. Results from a paired t-test corroborate the finding that on average, ϵ was greater for Familiar objects than Novel objects, $t(24)=6.70$, $p<.001$, which also verifies the effect at an individual level. This supports our hypothesis that judgments of material qualities are not based purely on the observed material mechanics, but are also affected by knowledge about the object. Referring to Figure 16, in some cases the directionality of the above effect is directly interpretable. For example, the jelly, spoon and teacup that wrinkled surprisingly (Figure 15, purple bars) were rated as harder (top plot) than the linen, silk, and velvet curtains that wrinkled expectedly. Prior knowledge about spoons and teacups (and even jelly relative to cloth) being hard led to increased ratings of hardness compared to their soft curtain counterparts, despite all of these objects wrinkling. Thus, prior object knowledge about hardness "pulled" ratings of hardness towards this expectation. A similar interpretation can be made for hardness ratings of Shattering objects: the clay pot, wine glass and teacup that Shattered expectedly were rated as harder on average than the custard, linen curtain, and water that Shattered surprisingly; hardness ratings of the latter items were "pulled down" by the expectation of softness.

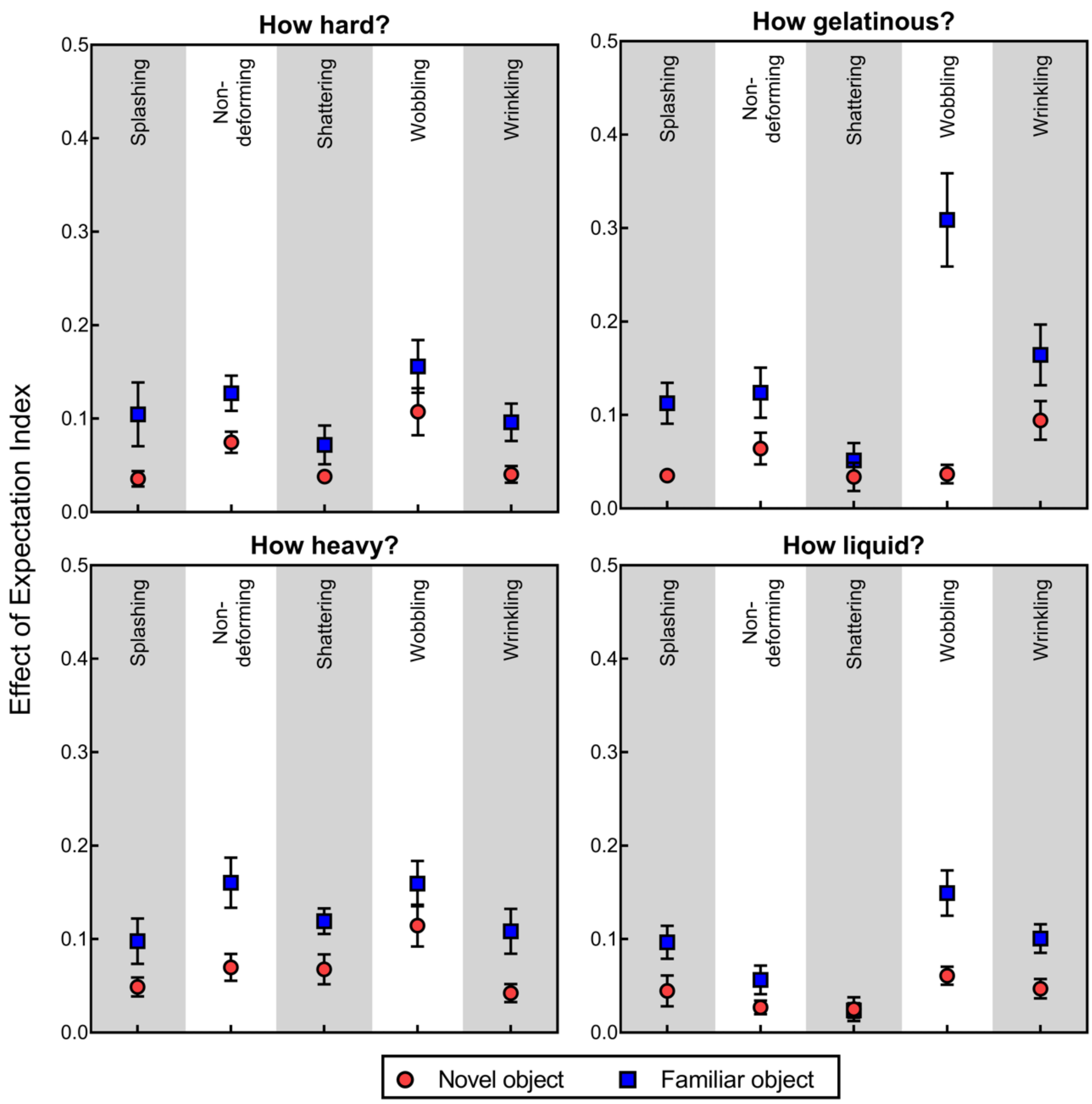


Figure 16. Effect of Expectation Index (ϵ) averaged across participants. ϵ was calculated as the absolute difference between ratings in the Expected and Surprising conditions (see analysis section for equation). Error bars are one standard error of the mean. *Figure contributed by A. Schmid.*

In other cases, the directionality of the rating differences requires more interpretation, which is why we were originally agnostic about the directionality of ϵ. The jelly, spoon and teacup that wrinkled unexpectedly (purple bars) were rated as more gelatinous (bottom plot) than the linen, silk, and velvet curtains that wrinkled expectedly. While object knowledge may have "pulled" up gelatinous ratings for the jelly, the spoon and teacup require a different explanation. It is possible that in these two conditions, the wrinkling action was not seen for what it is; shape and optics may have interacted (see Schmid & Doerschner, 2018) to make the spoon appear like melting metal and the cup like rubber, leading to increased ratings of gelatinousness. In these cases, object knowledge interfered with ratings in a different way than in the hardness rating case. However, in both cases, the material mechanics of the Novel objects were perceived more for what they were (i.e. Wrinkling materials) without the top-down influence of familiar shape, and hence the differences between the matched expected and surprising conditions are smaller.

The above suggests that top-down object knowledge influences perceived material properties. Yet, it is possible that low-level differences between our expected and surprising conditions could also be driving some of the effects. For example, the second panel in Figure 16 shows a high $\boldsymbol{\epsilon}$ for gelatinous ratings of wobbling objects (i.e. large difference between jellies/custard that wobble expectedly, and wine glass/honey/key that wobble surprisingly). Looking at the rating plots (Figure 15, bottom plots), we can see that the wobbling key is rated more similarly to the wobbling jellies/custard than the wine glass and the honey. It is unlikely that they key is less resistant to the effect of object knowledge than a wine glass or honey. In fact, based on the optical properties, it should be more resistant if anything (in the corresponding Novel object plots, the optical material properties of the key lead ratings to differ more from the jellies versus the glass and honey optics). Rather, it is possible that "low-level" differences between the objects in the Surprising condition are driving this difference. Inspecting the movies (link available in References) it appears that the key wobbles more (similar to the (Familiar) jellies and the custard), and the wine glass and honey appear to wobble less. Therefore, it is possible that in some cases low-level differences i.e. resulting from the specific interactions of object geometry and material mechanics, might influence the $\boldsymbol{\epsilon}$ scores.

However, we do not believe that ϵ results can be explained solely by such low-level effects. To continue the wobbling key example: it wobbles substantially, and yet, is still rated as harder and less gelatinous than the jellies, suggesting that an effect of object knowledge is present. Moreover, $\boldsymbol{\epsilon}$ is non-zero for Novel objects (Figure 16, red symbols). The shape and image motion generated by these Novel objects was nearly identical in the Expected and Surprising conditions, thus the differences in optical material properties must have driven the rating difference ($\boldsymbol{\epsilon}$). For example, in Figure 15 the non-deforming velvet Novel Object was rated as less hard than any of the other Novel Objects in that condition. Similarly, the wobbling glass Novel Object was rated as more gelatinous than any of the other Novel Objects, presumably due to optical qualities. The rating differences might reflect top-down differences in expectation caused by "material knowledge" (rather than object knowledge, also see Schmid & Doerschner (2018) or Schmidt, et al. (2017) for further discussion). Alternatively, it might be possible that 'bottom-up' differences in image motion caused by optics (e.g. presence or absence of specular motion) might have caused these rating differences. In order to obtain an independent estimate of the degree of expectation for objects in each material mechanics group - caused by either object or material knowledge alone - we developed an additional measure of the effect of object knowledge that removed the potential confound of low-level differences in image motion between stimuli.

Predictability index Π

We wanted to verify whether participants really build predictions based on object knowledge by comparing their ratings of objects that behave expectedly, to ratings of the first frames of these movies. If object knowledge does lead to strong predictions about material properties, then ratings of the first frame should be highly consistent with Expected conditions for Familiar objects, but not Novel objects. We had a separate group of observers (N=15) rate only the first frames of these movies, and compared these to our previous ratings from the Expected condition.

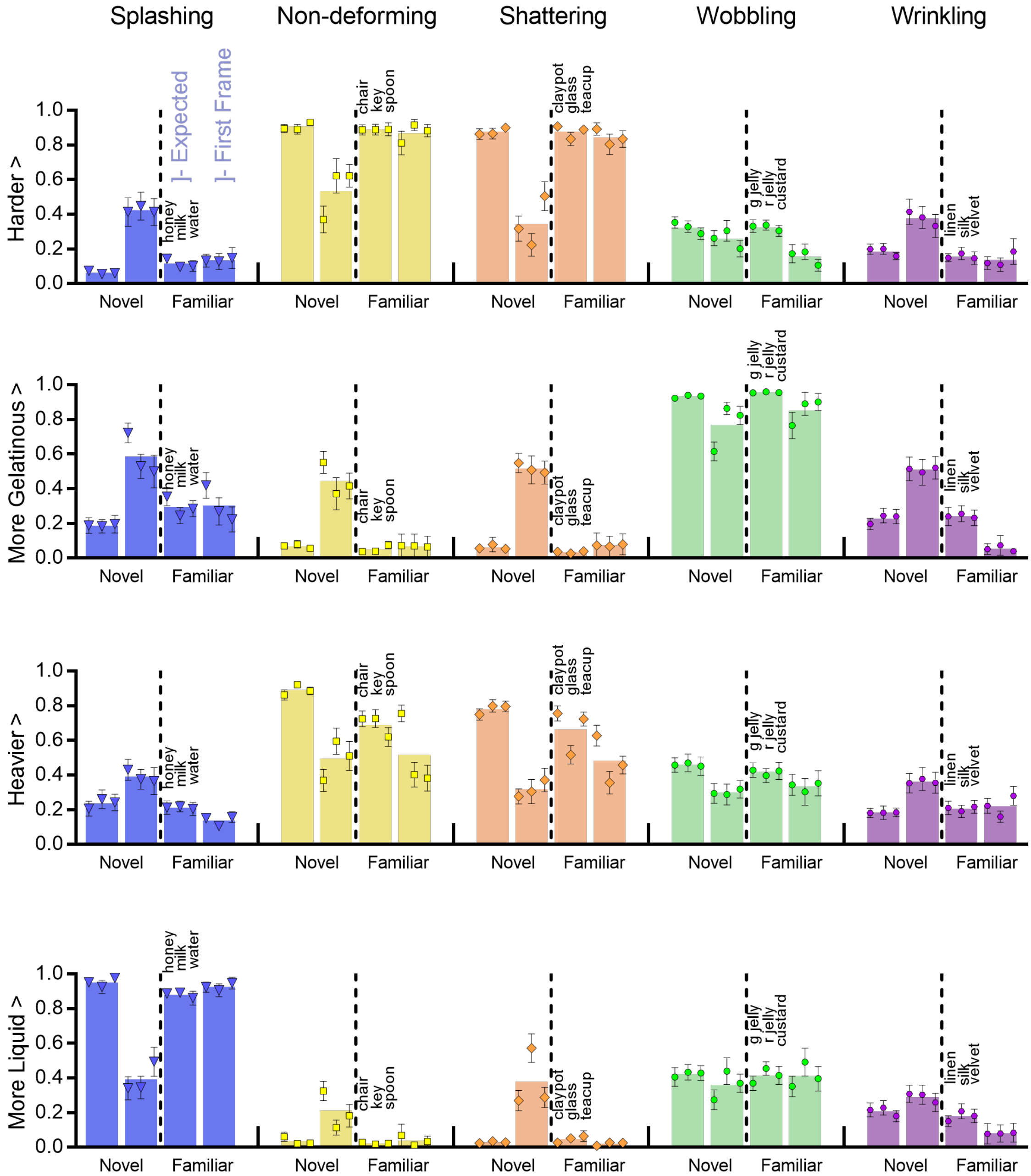


Figure 17. Ratings for Expected and First Frame conditions. Expected condition is the same data from before, First Frame ratings are from a new set of observers. Error bars are one standard error of the mean. *Figure contributed by A. Schmid.*

Figure 17 plots all ratings of the first frame next to the Expected movie condition for Novel (to the left of the dotted line) and Familiar objects. Notably, first frame ratings of Novel objects differ substantially from the other three conditions (Familiar first frame, Familiar Expected, Novel Expected), which are all more similar to one another. Novel and Familiar Expected object conditions tended to be rated similarly because objects in these conditions deformed in the same way (e.g. they splashed, shattered etc). The corresponding Familiar first frame condition was rated similarly to these two movie conditions, suggesting that object knowledge was used to build predictions about the material qualities of these objects: observers were able to (correctly) determine that the teacup is hard and the curtain is soft. Conversely, the First Frame of the Novel objects appeared to have generated less specific predictions about the material mechanics: the optical material properties were not enough to predict whether a clear transparent blob would be hard glass, or a water droplet suspended in air, or a bouncy silicon object, as indicated by more or less intermediate ratings for First Frame Novel objects on all attributes.

Following from an analysis of ratings, we sought an alternative method to compute the 'Predictability' of our objects. By subtracting the ratings of the 'Surprising' outcomes from those of the 'Expected' outcomes, we are able to generate a measure of how predictive an object's First Frame static appearance is of its expected deformation behavior, relative to all objects in the stimulus set. To quantify the "predictability" of the Expected conditions, we calculated a predictability index:

$$\boldsymbol{\Pi = 1 - |\,\text{av. ratings First Frame} - \text{av. ratings Expected}\,|} \qquad \textbf{Equation 4}$$

We subtract from one so that a higher number indicates the material outcome was more predictable. Like the other indices, we first calculated the average rating of the three objects in each material class, then computed the rating difference. Predictability scores are plotted in Figure 18. We performed a binomial sign test, which was used to compute the likelihood of obtaining k or more instances in which $\boldsymbol{\Pi}$ was greater for Familiar versus Novel objects. Figure 18 shows that Familiar objects were more predictable than Novel objects 18/20 times (sign test: $p < 001$). The very high predictability scores for Familiar objects suggests that participants really do build predictions based on object knowledge. Note that the predictability index for Novel objects was not zero, which suggests that observers generate some predictions about the mechanical properties based on the estimated optical properties of these objects, and/or based on associations they have with the general material properties of a bounded convex shape (e.g. a rigidity prior (Ullman, 1979; Grywacz & Hildreth, 1987; but also see Jain & Zaidi 2011).

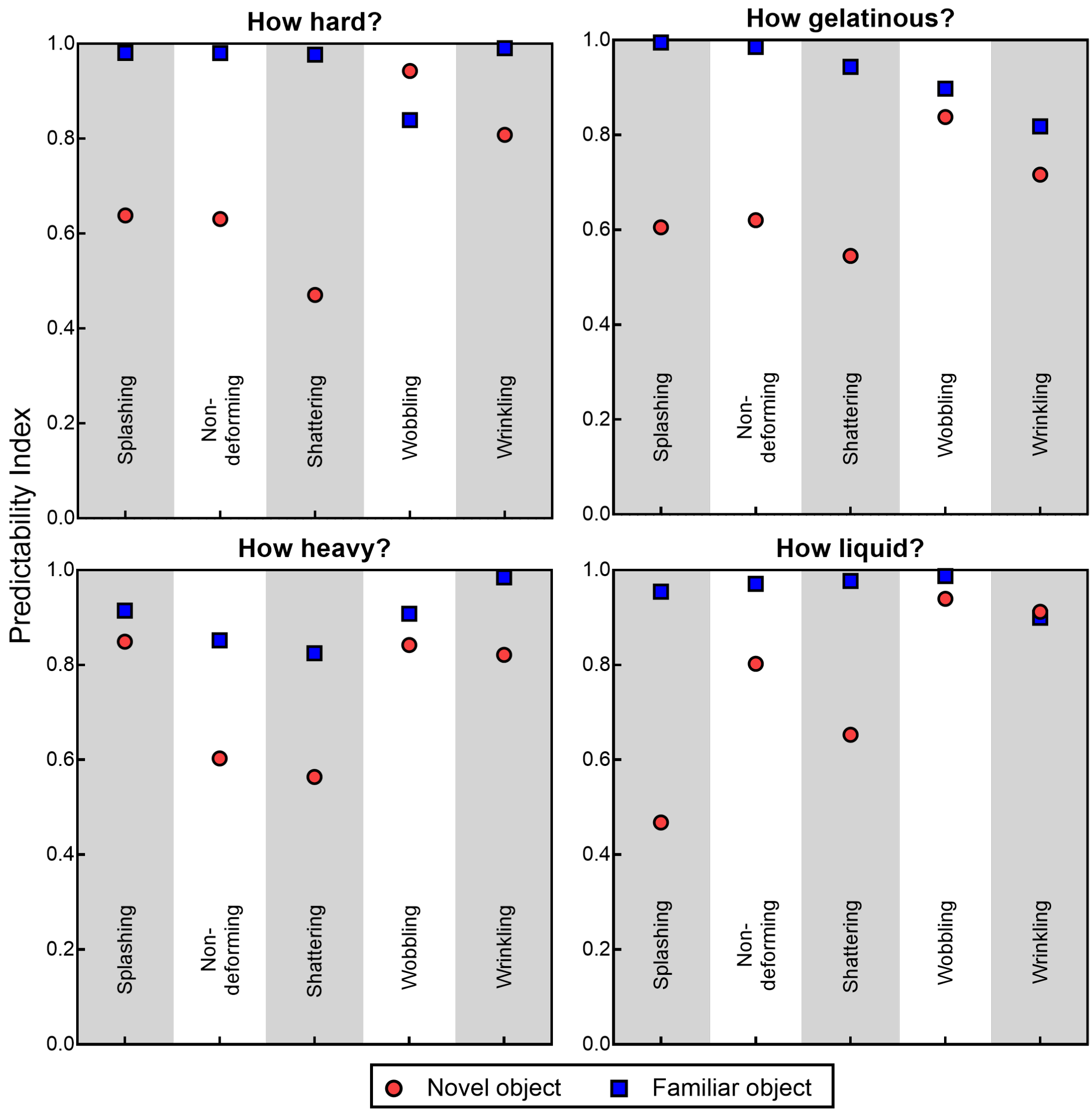


Figure 18. Predictability index *Π* calculated as one minus the absolute difference between first frame ratings and Expected condition ratings. There are no error bars because a separate group of observers rated first frames, so difference scores were computed between ratings averaged across observers. *Figure contributed by A. Schmid.*

Interobserver reliability

An alternative way to assess how predictable or Surprising the material mechanics were is to look at the consistency in ratings between observers. Average correlations between observer ratings are plotted in Figure 19, and were calculated as follows: first, observers' raw ratings for each object type (Familiar, Novel) and each experimental condition (Expected, Surprising, First Frame) were correlated with every other observer, and the average was computed. Ratings of the first frame of Novel objects show far less interobserver reliability than the other conditions. This further shows that for Novel objects, no specific expectation (at the group level) was generated about the material qualities of these objects. Conversely, observers were quite consistent when rating the first frame of Familiar objects, about as consistent as in the two movie conditions.

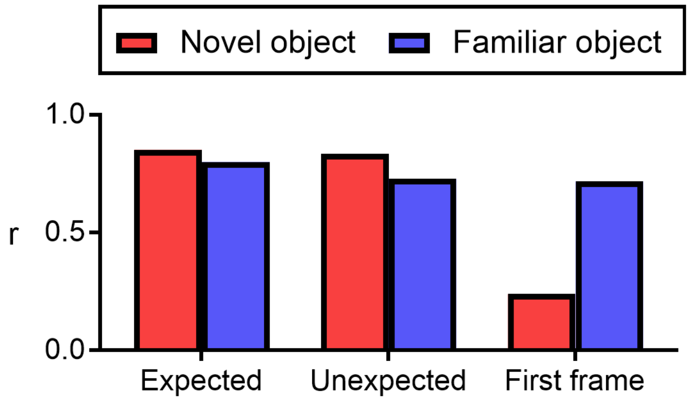


Figure 19. Average interobserver correlations. Ratings of the first frame of Novel objects show far less interobserver reliability than the other conditions. This might indicate that no specific expectation (at the group level) was generated about the material qualities of Novel objects. *Figure contributed by A. Schmid.*

Reaction Time Difference

As shown in Figure 14a, a trial started with 3 second period at which the first frame of a movie was shown, followed by 3 replays of the entire clip, each replay lasting for 2 seconds. Thus, the first replay of the movie ended at 5 seconds into a given trial. Figure 20 plots the reaction time difference (τ_D) between Surprising and Expected conditions. τ_D was greater than zero in 15 out of 20 conditions (i.e. Surprising events took longer to rate), sign test: p=0.041. Of these 15 conditions, the τ_D effect was stronger for Familiar objects versus Novel objects 12 times, sign test: p=0.035 (black stars in figure). In other words, when judging Familiar objects, observers took longer to rate objects that 'deformed Surprisingly' versus expectedly. This was not the case on average for Novel objects. Results from paired t-tests corroborate these findings at an individual level. On average, reaction times in the Surprising condition were longer than in the Expected condition for Familiar objects, t(24)=5.17, p<.001, but not Novel objects, t(24)=1.43, p=.17. Furthermore, $\tau_D$was larger for Familiar versus Novel objects, t(24)=2.17, p=.040.

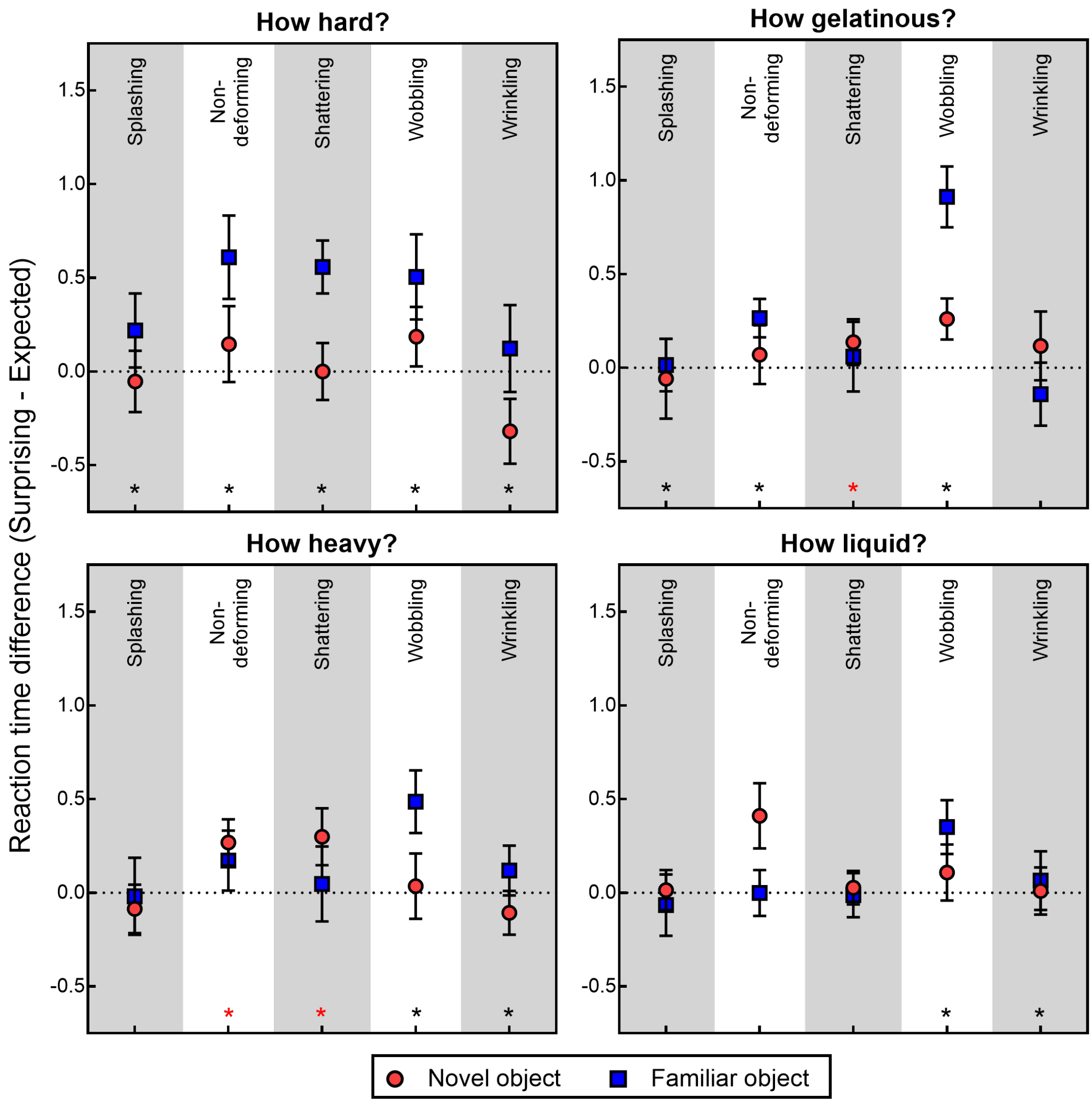


Figure 20. Reaction time difference (τ_D) between Surprising and Expected conditions averaged across subjects. A score above zero indicates Surprising trials took longer, and a score below zero indicates Expected trials took longer. Error bars are one standard error of the mean. *Figure contributed by A. Schmid.*

The increased reaction time when rating Familiar objects that behave in a Surprising way fits with our hypothesis that – in a Bayesian framework - these trials might involve the correction of a larger prediction error, when compared with trials that are consistent with expectations generated by object (or material) knowledge.

Given these mean indexes and Reaction Time Differences, we investigated whether a correlation existed for object categories and object types to probe the nature of the relationship between the delay in reaction time and the degree of Expectation of the object categories.

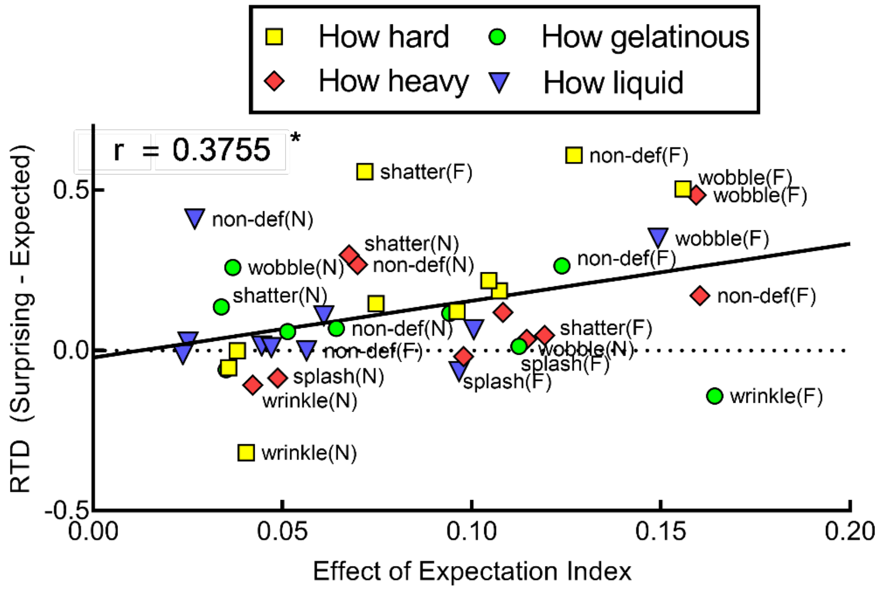


Figure 21. Correlation between ϵ and τ_D. The positive correlation was significant (r=.3755, p=.0185). We removed one outlier (how gelatinous is wobbling jelly: ϵ =0.31, τ_D=0.91), however including this outlier only made the correlation stronger (r=.5866, p<.001). A group of points are labelled with their material deformation: F stands for Familiar Object, N stands for Novel object. *Figure contributed by A. Schmid.*

Figure 21 plots the correlation between the Effect of Expectation Index and the Reaction Time Difference. Points are paired between object type and deformation method (shatter, wrinkle, wobble, etc.) for each question posed in the experiment. We find a significant positive correlation (r=.3755, p=.0185) between how expected an object deformation category was for that object, and the resulting delay in reaction time. We removed one outlier (how gelatinous is wobbling jelly: ϵ =0.31, τ_D=0.91) (However, including this outlier only made the correlation stronger (r=.5866, p<.001)). This suggests that the object categories with slower reaction times that behaved expectedly (below the line) (e.g. splashing Familiar object when asked about liquidity) generally have longer reaction times than their unexpected counterparts.

Analysis II/"Prior Pull":

After computing these indices, we sought to investigate the effect of the prior in making such material property rating judgements. First Frame ratings (see 'First Frame' images of Figure 22b) are used as a measure of objects as prior knowledge about how an object should deform (as a function of familiar shape and optics associations). In contrast, ratings of moving Novel objects are 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 14a) but make *different* predictions (i.e. yield different ratings) when the material behavior is surprising (e.g. rigid (non-deforming) red Jell-O, Figure 14b).

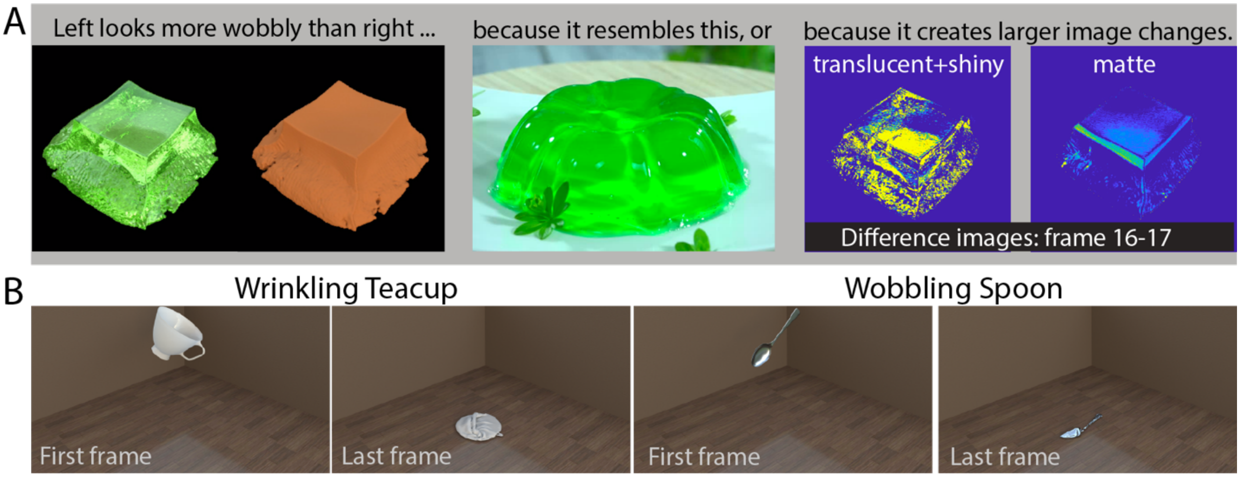


Figure 22. Contribution of prior associations and image cues on perceived material qualities. A. The perception of material qualities (such as gelatinousness) can be influenced by prior associations between dynamic optics, shape, and motion properties. Watching the green (left) object deform may evoke 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. 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.

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. 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.

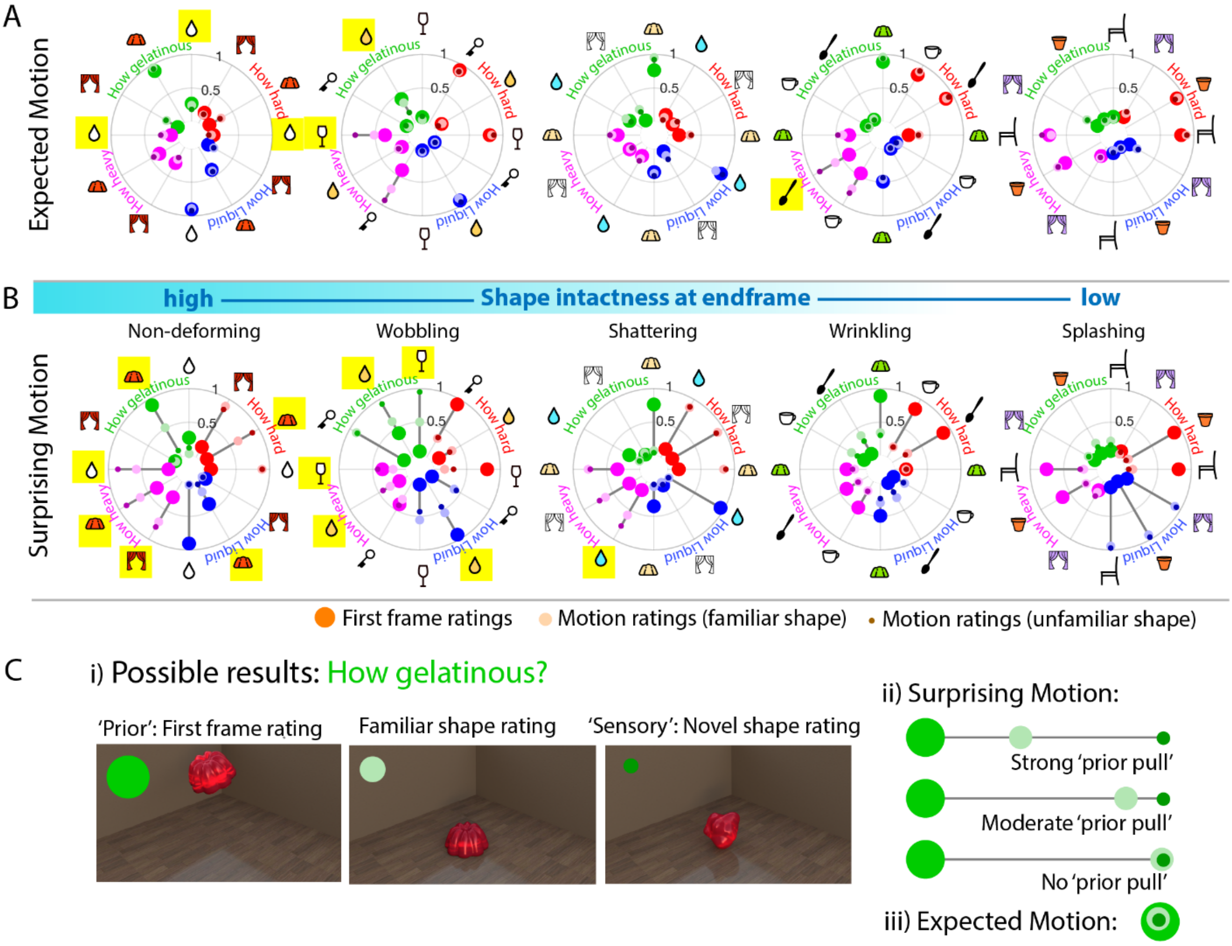


Figure 23 a-c. Material quality ratings and prior attraction. A. Average observer ratings for different questions about material qualities e.g. how hard, liquid, heavy or gelatinous an object appears. Icons symbolize individual familiar objects (chair, key, cup, pot, glass, spoon, blue droplet: water, 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. *Portions of figure provided by A. Schmid.*

The yellow highlighted cases in Figure 23a show that this "prior pull" occurred twice as much in the Surprising (10 cases) than the Expected condition, 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 13a). 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) and calculated differences in motion energy. 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.

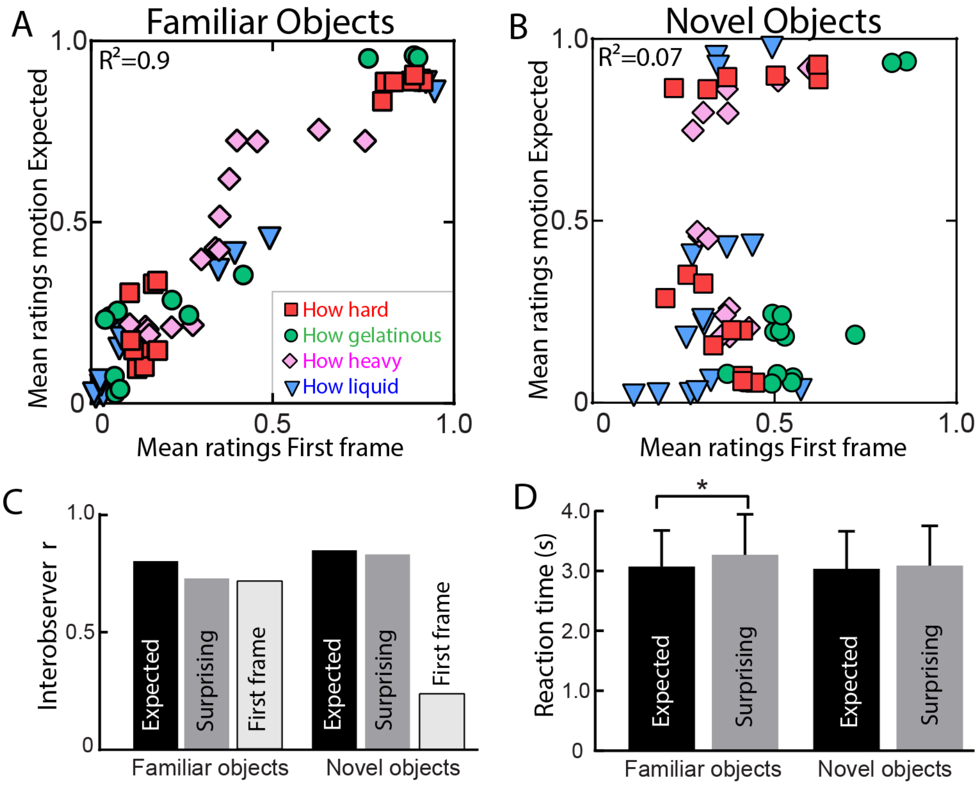


Figure 24. Prediction strength reaction time differences and interobserver correlations. A. Correlation between mean First frame ratings and mean Expected motion ratings for familiar and novel objects (**B.**) A high correlation indicates that the First frames (still images) of objects are highly predictive 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. *Regression analysis/figure provided by A. Schmid.*

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 15B), 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. We can model the data to an extent (R2 between 0.266 and 0.59 depending on the question), but not perfectly, potentially due to specific motion-shape-optics interactions.

Discussion

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), 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. 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.

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. 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.

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. Importantly, we do not find evidence for a reaction time advantage in expected familiar objects condition, 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, other work showed that *both* optical and mechanical cues affect estimates of viscosity, and yet other research concludes that optical properties dominates over image motion and shape cues when judging the stiffness of cloth. 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.

We probed observers' predictions about the mechanical properties of materials by using dynamic computer-rendered scenes and a violation-of expectation paradigm. We created an index that captures the differences in material judgements between Expected and Surprising object deformations, and find that the value of this index is larger for Familiar than Novel objects. Moreover, when expectations about object 'behavior' are violated, it takes observers longer to rate the object's mechanical material properties. Overall, predictions about mechanical properties of materials were not just activated by the familiar shape of an object, but could – to a lesser extent – also be activated by optical material properties alone. Taken together, our results imply that many perceived material qualities are not solely determined by the retinal stimulation, but instead are rather susceptible to cognitive influences.

Conclusion

This work shows that expectations regarding kinematic/mechanical properties of materials can be activated by the familiar shape of an object, but also by the optical qualities of a surface. Our results imply that perceived material qualities are not only determined by the retinal stimulation, but instead can also be susceptible to cognitive influences, such as expectations and memory. Of particular interest is the effect that the 'typical' color of objects has on perceived object properties. The experiment described below investigates this question by removing the typical/expected color from our Surprising stimulus set.

Experiment 2.3: Greyscale Rating Version

In the previous experiment, we found that both the familiar shape and optical material qualities of an object play a role in the identification of object identity. These results suggested that the perception of object identity is affected not only by perception alone, but is also influenced by prior knowledge about objects. This led us to investigate the influence of color (e.g. 'typical material color' c.f. Biederman, 1987) on the perception of our stimuli. In the absence of color, participants must rely on shape cues and object priors to make such material judgements.

To investigate the effect of optical properties and color on all previous results, we repeated the rating experiment using greyscaled versions of the stimuli previously described. We hypothesized that optical properties would provide more cues to object identity and therefore, we would find faster responses to colored stimuli than to greyscaled ones. This hypothesis is supported in the literature-- Price and Humphreys (1989) found that surface information such as color was particularly helpful for recognizing structurally similar objects, and reported superior latency performance on colored photographs and computer images, compared to greyscale ones. For such objects, naming was more rapid with colored slides and *appropriate* colored line drawings than it was with line drawings or *inappropriately* colored line drawings. Using response latencies as a method, Davidoff and Ostergaard (1988) and later, Brodie (1991) uncovered a naming-latency advantage for colored over black and white depictions, finding that the advantage for color depictions was equally present for manufactured objects and natural ones. Wurm et. al. (1993) further concluded that colored depictions of food items *congruent* with the expected color of the food item were named faster than greyscale (*incongruent*) depictions of the same objects. Following from this literature, the following experiment sought to investigate the influence of color when making kinematic material property judgements.

Methods

Stimuli

The stimuli used in this experiment were taken from the motion experiment described in Chapter 2. The stimuli were greyscaled using the rgb2grey function in MATLAB. The Greyscale version of the First Frame data used in the analyses for this experiment were collected consistent with those described in the previous experiment.

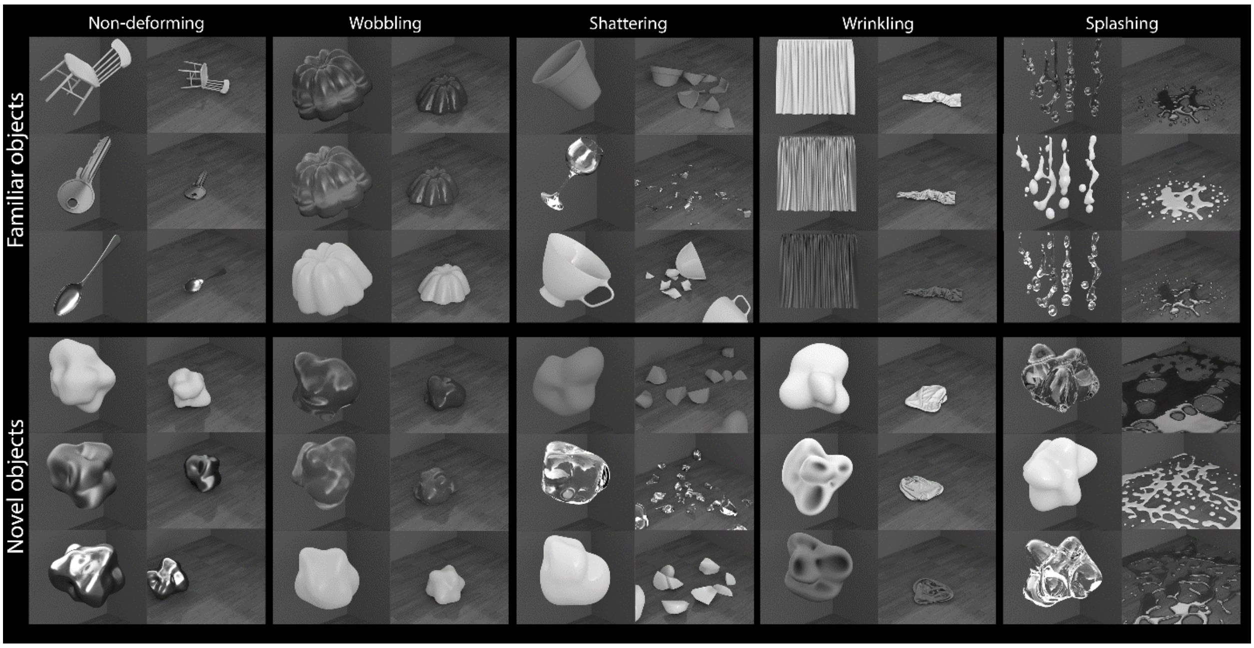


Figure 25. Greyscaled Familiar Object Stimuli. Consistent with the previous stimuli, here we show (in greyscale) Familiar objects grouped according to their material mechanics. The right-hand column of all categories depicts corresponding last frames of animations that show how a given object fully deformed in the Surprising condition. We refer the reader to Zenodo (see Reference List) for the corresponding animations.

Apparatus

As in the previous experiment, the experiment was coded in MATLAB 2015a (MathWorks, Natick, MA) using Psychophysics Toolbox (version 3.8.5), 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

The task and procedure used in this experiment were identical to that of the motion experiment previously described. We refer the reader to the previous experiment for these details.

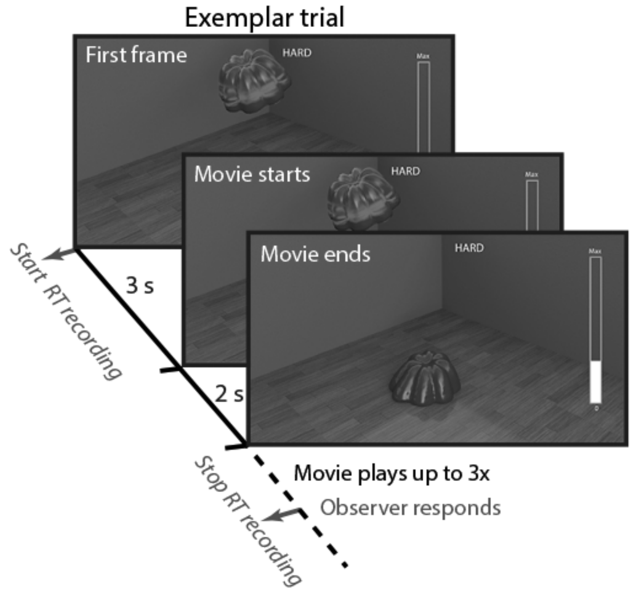


Figure 26. Example of a single greyscaled trial. The design of the experiment was identical to that previouslx described in Chapter 2. The first frame was shown for three seconds. The object then fell and deformed in a manner that was either Expected or surprising. The participants adjusted the bar at the right in order indicate their rating of the object's material appearance. The attribute on which to rate the object was always present throughout the trial.

Participants

20 naïve participants from JLU Giessen participated in the experiment. All participants had self-reported normal or corrected-to-normal vision. The experiment followed the guidelines set forth by the Declaration of Helsinki. All participants provided written informed consent and were reimbursed at a rate of **€**8/hour.

Results

Analysis I

For clarity, we reproduce the equations below (as discussed in the previous chapter).

Rating Differences, Expected versus Surprising: To investigate the effect of the removal of color when making various material property judgments (consistent with the previous experiment), we computed rating differences between Expected and Surprising conditions. For each object type (Familiar, Novel), each attribute (how hard, how gelatinous, how heavy, how liquid), and each material class (Splashing, Shattering, Non-deforming, Wobbling, Wrinkling), rating differences were calculated between the three objects with Expected outcomes and the three (different) objects with Surprising outcomes (ratings were averaged over the three objects before difference scores were computed). Thus, how an object deformed was the same, e.g. it would splash, but which object (Familiar or Novel) would do the splashing is different in Expected (honey, milk, water) and surprising conditions (chair, pot, curtain). As before, (given the increased ambiguity in differentiating objects) we were agnostic about the direction in which the rating will change (e.g. will a Splashing chair look more liquid, or harder, or heavier than Splashing milk?), so we took the absolute value of this difference score as an index of the effect of expectation on ratings (Effect of Expectation Index, $\boldsymbol{\epsilon}$):

$$\epsilon = |\ \textbf{avg. ratings Expected} - \textbf{avg. ratings Surprising}\ | \qquad \textbf{Equation 2}$$

Consistent with the full-color stimuli, this resulted in 40 difference scores for each observer, 20 in the Familiar object condition and 20 in the Novel object condition. If object knowledge influences perceived material properties (even in the absence of typical/expected color), ratings should differ between Familiar objects that behave as expected, and Familiar objects that behave in a surprising way. Familiar objects should be larger than $\boldsymbol{\epsilon}$ for Novel objects ($\boldsymbol{\epsilon}_{\text{Familiar object}} > \boldsymbol{\epsilon}_{\text{Novel object}}$). A binomial sign test was used to compute the likelihood of obtaining k or more instances in which $\boldsymbol{\epsilon}$ (averaged across observers) was greater for Familiar versus Novel objects.

Reaction times

Consistent with the previous rating experiment, the time taken to make each judgment (reaction time) was also measured. We again reasoned that, as in the full-color stimuli, 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. Reaction Time Difference Scores (τ_D) were then computed as follows:

$$\boldsymbol{\tau_D} = \textbf{av. reaction time Expexted} - \textbf{av. reaction time Surprising}. \qquad \textbf{Equation 3}$$

Consistent with the rating data of the full-color rating experiment, reaction times were averaged over the three objects in each material class before difference scores were computed, which led to 40 difference scores for each observer, 20 in the Familiar object condition and 20 in the Surprising object condition. These difference scores were averaged across observers, and a binomial sign test was used to compute the likelihood of obtaining k, or more instances in which the τ_D was greater for Familiar versus Novel objects.

Exclusions

The previous error regarding the Novel object 'Key' issue did not occur in the collection of greyscale rating data. As before, 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, 208 data points were excluded for reaction times that were too fast or too slow according to this criterion.

Ratings/Effect of Expectation Index (ϵ)

Consistent with the previous (full-color) experiment, the ratings of observers varied systematically with the mechanical material properties of Familiar and Novel objects, and are largely consistent with the Full-Color Rating data from the previous experiment. Figure 27, which shows average ratings for all attributes, again illustrates that observers were able to perform the task sensibly. In general, the data for ratings between the full-color and greyscale experiments are remarkably consistent with one another. The differences between the ratings of Full-Color and Greyscale rating data lie largely in the 'Wobbling' condition, as there is greater variability for Wobbling items for all four rating attributes between the Full-Color and Greyscale Rating data. Familiar Expected items, for ratings of Liquidity, tended to be rated as slightly less liquid than their colored counterparts. Error bars for the individual objects tended to group more closely to the averages of the bars in the greyscale data, whereas they varied more significantly in the Full-Color rating data.

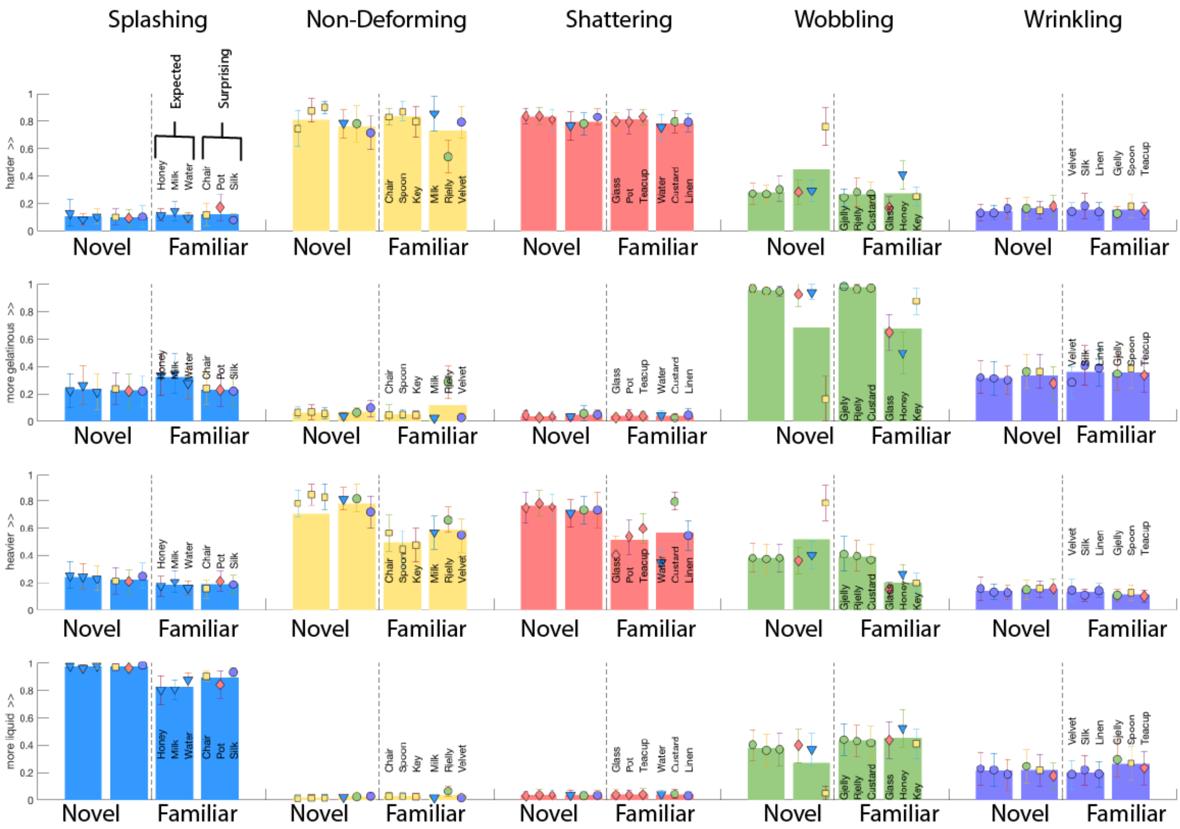


Figure 27. **Greyscaled Rating Data for all Attributes (Expected/Surprising).** Ratings are consistent with their deformations: Splashing items are rated as very liquid, and non-deforming/shattering items as very hard. Wobbling items vary more significantly in their ratings of Wobbliness. In general, the data for ratings between the full-color and greyscale experiments are remarkably consistent with one another. Error bars are one standard error of the mean.

Referring to Figure 27, the jelly, spoon and teacup that wrinkled surprisingly (Figure 27, purple bars) were rated as only slightly less hard than the linen, silk, and velvet curtains that wrinkled expectedly. While this is the opposite of the finding in the Full-Color rating data, prior knowledge about hard items still led to increased ratings of hardness compared to their soft curtain counterparts, despite all of these objects wrinkling. Thus, prior object knowledge about hardness influenced ratings of hardness in the direction of the expectation (if only slightly, for Greyscale ratings). Remarkably consistent for greyscale data, a similar interpretation can be made for hardness ratings of Shattering objects: Even in the absence of typical color, the clay pot, wine glass, and teacup that Shattered expectedly were rated as harder on average than the custard, linen curtain, and water that Shattered surprisingly.

In contrast to the Full-Color ratings, greyscale ratings of the jelly, spoon and teacup that wrinkled unexpectedly (purple bars) were rated as equally gelatinous as the linen, silk, and velvet curtains that wrinkled expectedly. While object knowledge may have influenced gelatinous ratings for the jelly, the spoon and teacup require a different explanation, as they are ordinarily non-gelatinous objects. It is possible that for these rigid items, shape and optics (or the lack thereof, in the absence of typical color, shape and optics priors may have interacted to make the spoon appear more ambiguous, leading to increased ratings of gelatinousness.

Figure 28 plots the Effect of Expectation Index ($\boldsymbol{\epsilon}$) for each attribute and each material class, averaged across subjects. In 11/20 conditions, $\boldsymbol{\epsilon}$ was not significantly greater for Familiar objects than Novel objects, sign test: p=.15. Results from a paired t-test find that on average, $\boldsymbol{\epsilon}$ was significantly greater for Familiar objects than Novel objects, t(18)=8.54, p=.040. Even in the absence of color, judgments of material qualities are not based purely on the observed material mechanics, but are also affected by prior knowledge about the object.

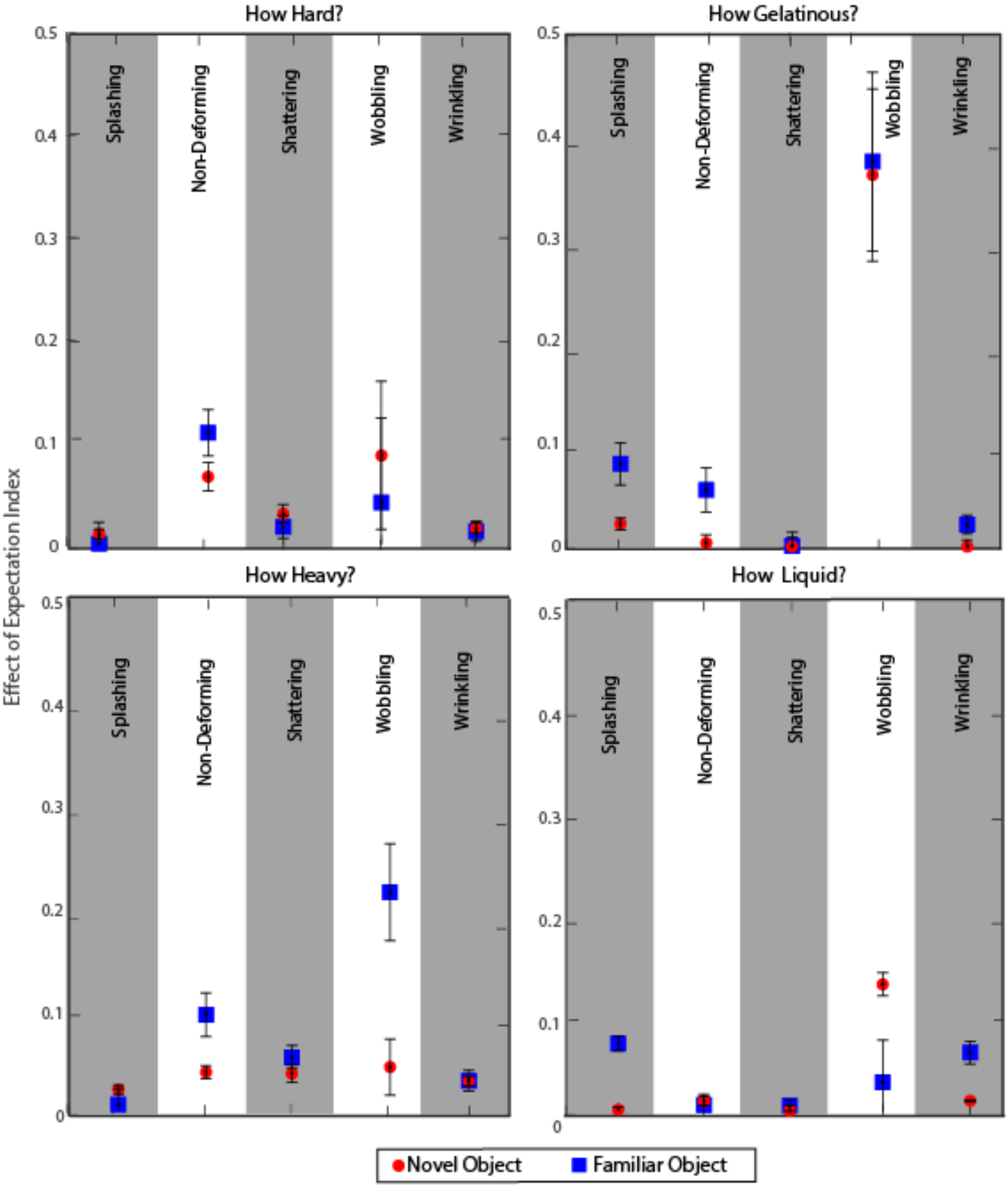


Figure 28. Effect of Expectation Index (ϵ). ϵ was calculated as the absolute difference between ratings in the Expected and Surprising conditions (see analysis section, Equation 2). Error bars are one standard error of the mean.

Predictability index Π

As with the Full-Color ratings, we wanted to verify whether participants really build predictions based on object knowledge (even in the absence of color) by comparing their ratings of objects that behave expectedly, to ratings of the first frames of these movies. If object knowledge, even in the absence of color, leads to strong predictions about material properties, then ratings of the first frame should be highly consistent with Expected conditions for Familiar objects, but not Novel objects. We had a separate group of observers (N=17) rate only the first frames of these movies, and compared these to our previous ratings from the Expected condition.

Figure 29 plots all ratings of the First Frame next to the Expected movie condition for Novel and Familiar objects. Notably (and, consistent with the Full-Color ratings), First Frame ratings of Novel objects differ from the other three conditions (Familiar First Frame, Familiar Expected, Novel Expected), which are all more similar to one another. Novel and Familiar Expected object conditions tended to be rated similarly because objects in these conditions deformed in the same way (e.g. splashed, shattered etc). The corresponding Familiar first frame condition was rated similarly to these two conditions, suggesting that object knowledge (even in the absence of color) was used to build predictions about the material qualities of these objects: observers were able to (correctly) determine that the teacup is hard and the curtain is soft. While this generally holds true, there were some surprising differences, finding more consistency between Novel and Familiar ratings for greyscaled objects (compared to the greater variability between Familiar and Novel in the full-color condition).

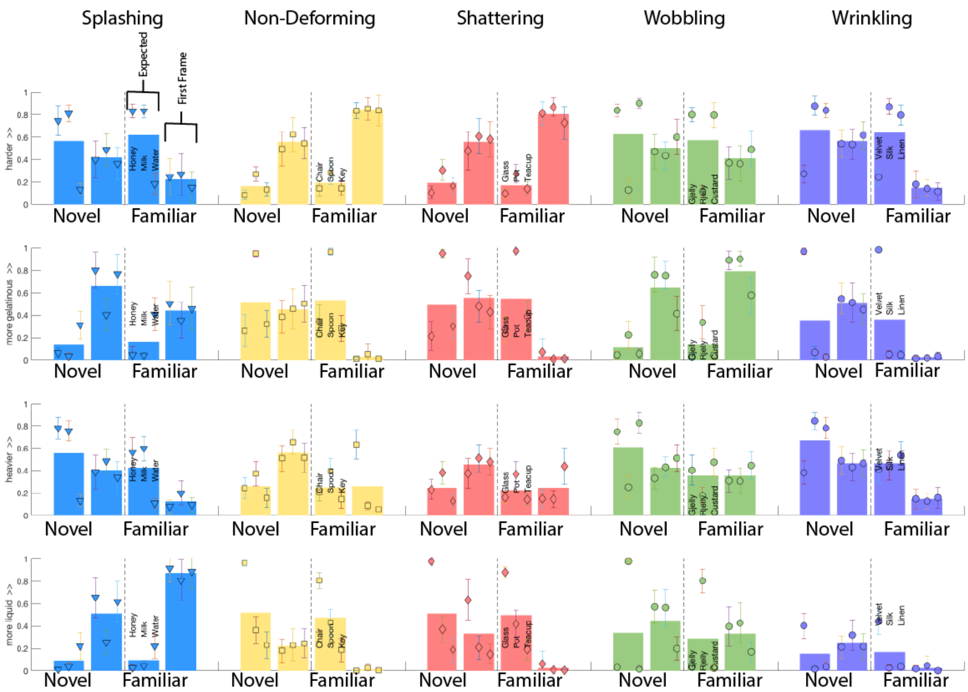


Figure 29. All Ratings for Expected and First Frame Conditions. The Expected condition presents the same data as the previous rating figure, while the First Frame ratings are from a new set of observers. Error bars are one standard error of the mean.

As in the Full-Color analyses, to quantify the Predictability of the Expected conditions, we calculated a predictability index:

$$\boldsymbol{\Pi = 1 - |\,\text{av. ratings First Frame} - \text{av. ratings Expected}\,|} \qquad \textbf{Equation 4}$$

Consistent with the full-color data, we subtract from one so that a higher number indicates the material outcome was more predictable. Like the other indices, we first calculated the average rating of the 3 objects in each material class, then computed the rating difference. Predictability scores are plotted in Figure 30. We performed a binomial sign test, which was used to compute the likelihood of obtaining k or more instances in which $\boldsymbol{\Pi}$ was greater for Familiar versus Novel objects. Figure 30 shows that Familiar objects were more predictable than Novel objects 14/20 times (although not significant, sign test: p=.12). Again, note that the predictability index for Novel objects was not zero, which suggests that observers generate some predictions about the mechanical properties based on the estimated optical properties of these objects, and/or based on associations they have with the general material properties of a bounded convex shape (e.g. a rigidity prior (Ullman, 1979; Grywacz & Hildreth, 1987; but also see Jain & Zaidi 2011).

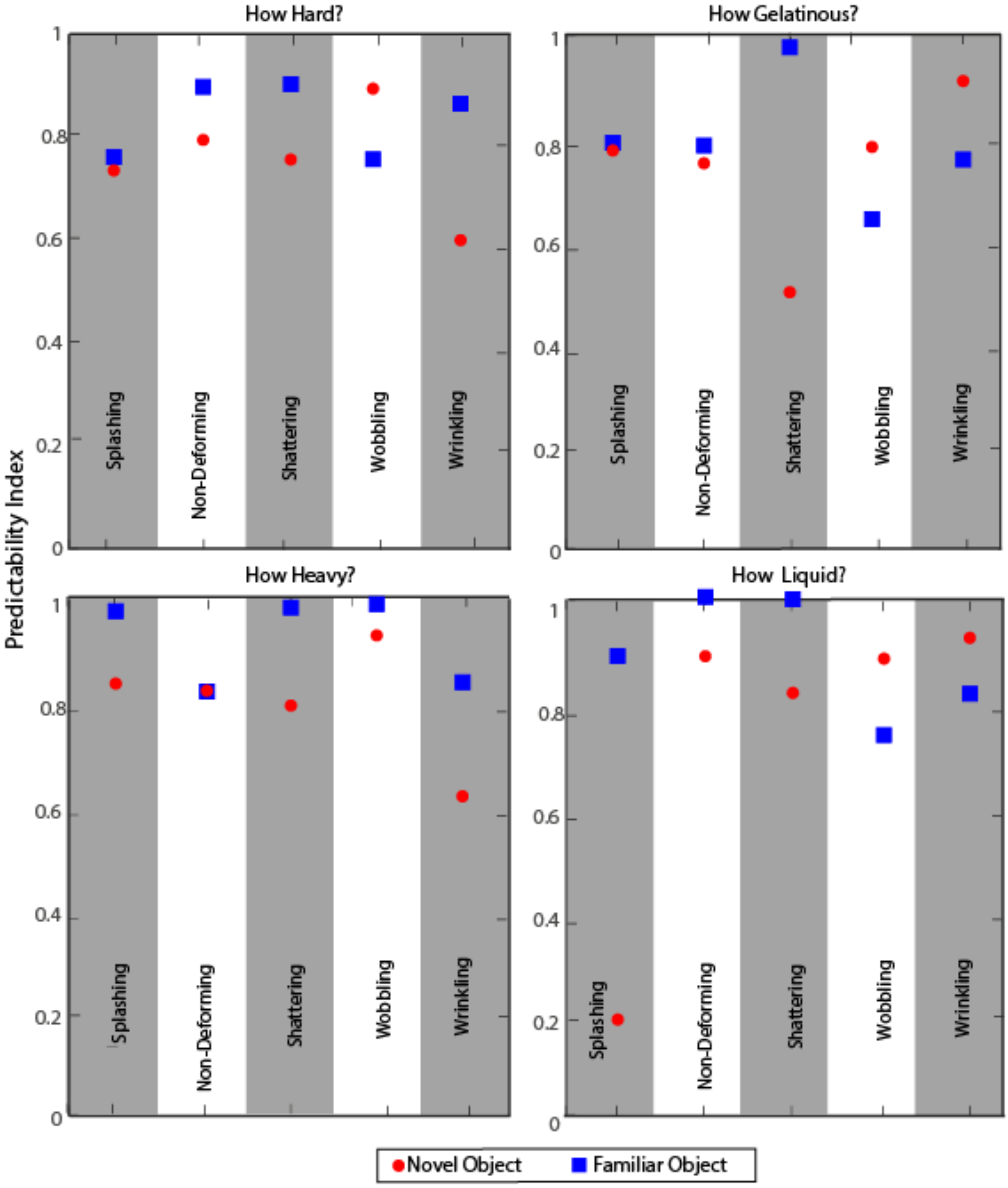


Figure 30. Predictability index Π calculated as one minus the absolute difference between First Frame ratings and Expected condition ratings. There are no error bars because a separate group of observers rated first frames, so difference scores were computed between ratings averaged across observers.

Interobserver reliability

As in the full-color rating analysis, an alternative way to assess how predictable or Surprising the material mechanics were is to look at the consistency in ratings between observers. Again, average correlations between observer ratings are plotted in Figure 31, and were calculated as follows: first, observers' raw ratings for each object type (Familiar, Novel) and each experimental condition (Expected, Surprising, First Frame) were correlated with every other observer, and the average was computed. As in the full-color data, ratings of the first frame of Novel objects again show far less interobserver reliability than the other conditions. This further shows that for Novel objects, no specific expectation (at the group level) was generated about the material qualities of these objects. Conversely, observers were quite consistent when rating the first frame of Familiar objects. The Unexpected condition follows the same pattern as the full-color data (Novel Unexpected objects having a higher correlation than their Familiar Object counterparts), while for greyscaled stimuli, Familiar Objects have a higher correlation than their Novel object counterparts.

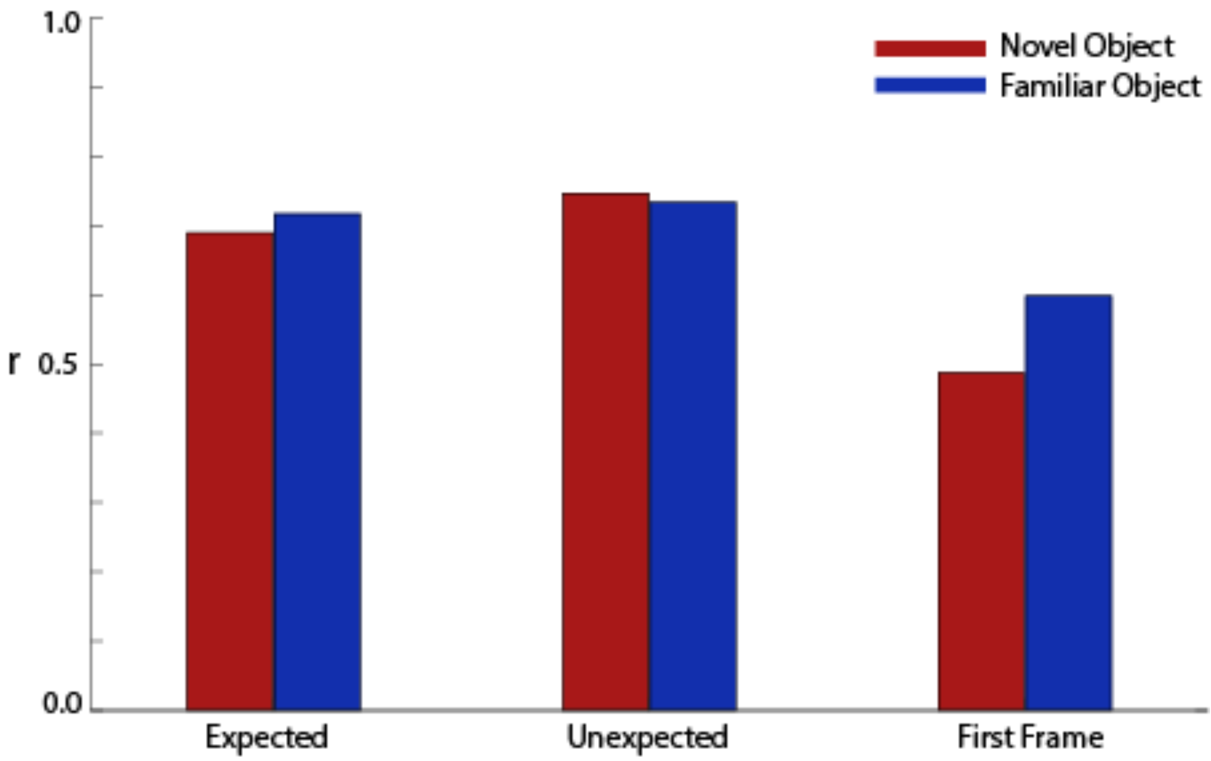


Figure 31. Average interobserver correlations. Ratings of the first frame of Novel objects show far less interobserver reliability than the other conditions. This might indicate that no specific expectation (at the group level) was generated about the material qualities of Novel objects. The Unexpected condition follows the same pattern as the full-color data (Novel Unexpected objects having a higher correlation than their Familiar Object counterparts), while for greyscaled stimuli, Familiar Objects have a higher correlation than their Novel object counterparts.

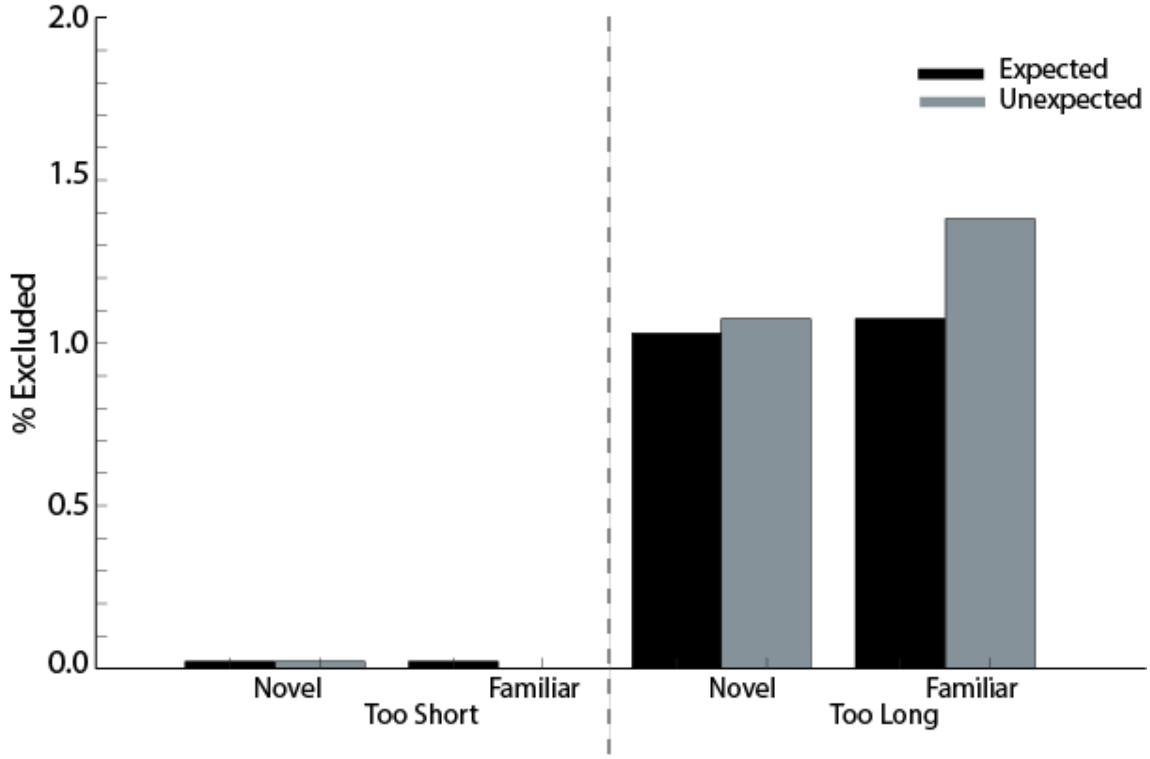


Figure 32. RT Exclusion Histograms. Consistent with the full-color rating data, trials where Familiar objects behave unexpectedly are excluded more than other types of trials. This finding is consistent with our Reaction Time (RT) results, as it suggests that participants continue to observe the Surprising events for a longer amount of time. This is consistent with our previous results/the general finding that object knowledge influences the ratings of objects.

Consistent with the full-color rating data, trials where Familiar objects behave unexpectedly are excluded more than other types of trials. This finding is consistent with our Reaction Time (RT) results, as it suggests that participants continue to observe the Surprising events for a longer amount of time. This is consistent with our previous results/the general finding that object knowledge influences the ratings of objects.

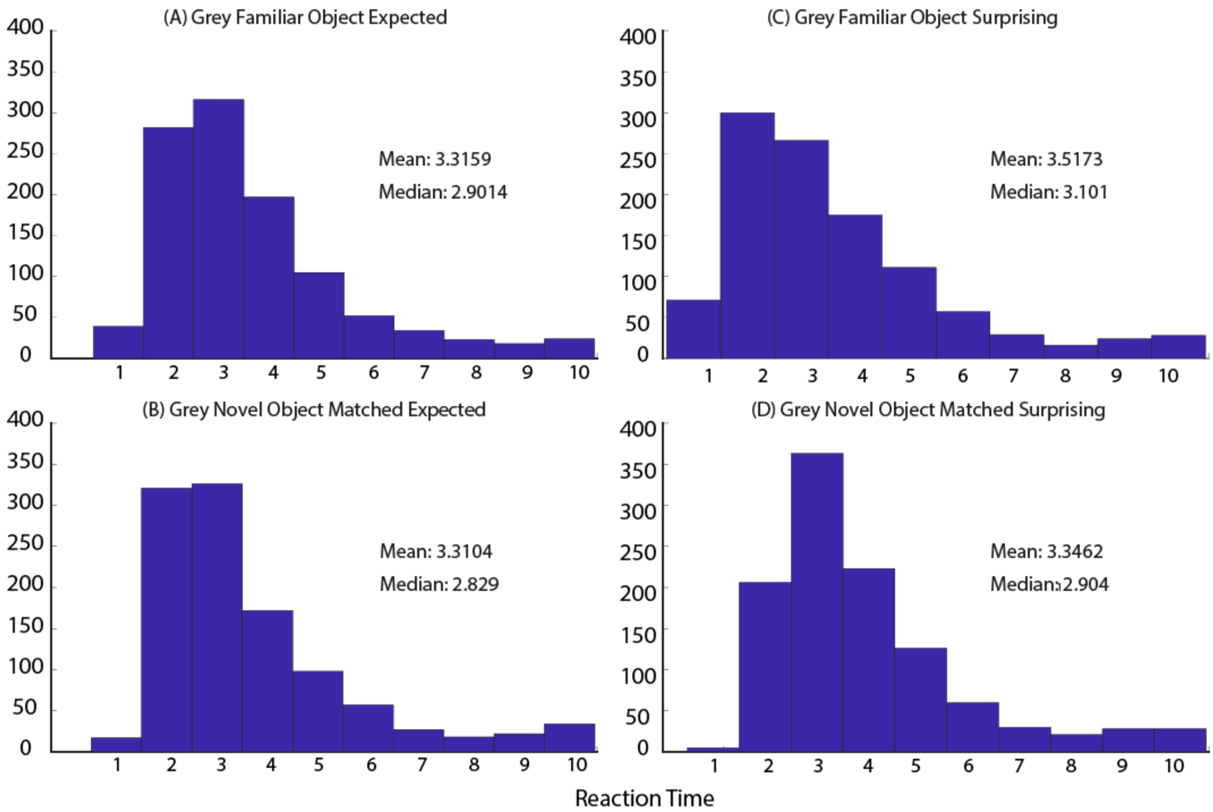


Fig. 33a-d. RT Histograms. Histograms for Reaction Times. Reaction times for greyscaled stimuli of Familiar Objects are slower for Familiar objects that behave surprisingly (C) versus expectedly (A). This difference is not present for Novel objects with matched outcomes (C and D).

Figure 33a-d presents histograms for reaction times of responses to the various conditions. Consistent with the previous Full-Color rating experiment, reaction times for greyscaled stimuli of Familiar Objects are slower for Familiar objects that behave surprisingly (33C) versus expectedly (33A). This difference is not present for Novel objects with matched outcomes (33C and D).

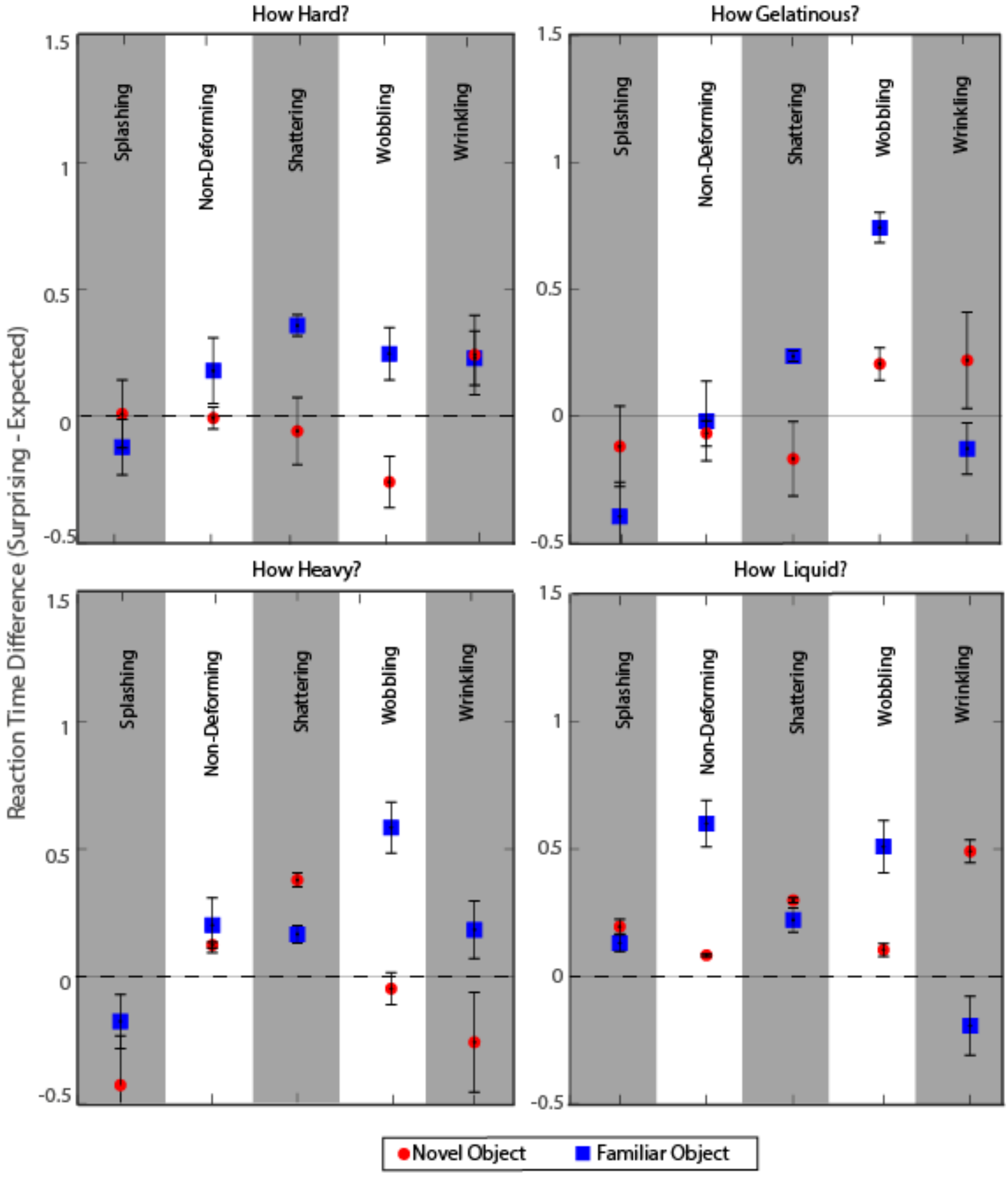


Figure 34. Reaction time difference (τ_D) between Surprising and Expected conditions averaged across subjects. A score above zero indicates Surprising trials took longer, and a score below zero indicates Expected trials took longer. Error bars are one standard error of the mean.

As shown in stimulus Figure 26, a trial started with 3 second period at which the first frame of a movie was shown, followed by 3 replays of the entire clip, each replay lasting for 2 seconds. Thus, the first replay of the movie ended at 5 seconds into a given trial. Figure 34 plots the reaction time difference (τ_D) between Surprising and Expected conditions. τ_D was greater than zero in 15 out of 20 conditions (i.e. Surprising events took longer to rate), sign test: p=.08. Of these conditions, the τ_D effect was stronger for Familiar objects versus Novel objects 11 times, sign test: p=.015. On average (and consistent with the Full-Color Rating experiment RTD data) when judging Familiar objects, observers took longer to rate objects that deformed Surprisingly (versus expectedly). This was not the case on average for Novel objects. Results from paired t-tests corroborate these findings at an individual level. On average, reaction times in the Surprising condition were not significantly longer than in the Expected condition for Familiar objects, t(18)=-3.17, p=.064, but were for Novel objects, t(18)=2.22, p=.035. Furthermore, τ_D was not significantly larger for Familiar versus Novel objects, t(18)=1.49, p=.064.

Consistent with the Full-Color Rating experiment reaction time difference data, the increased reaction time when rating Familiar objects that behave in a Surprising way fits with our hypothesis that – in a Bayesian framework - these trials might involve the correction of a larger prediction error, when compared with trials that are consistent with expectations generated by object (or material) knowledge. For direct comparison ,we plot color and greyscale reaction time differences below, in their separate 'Familiar' and 'Novel' object conditions.

Color vs. Grey Comparisons

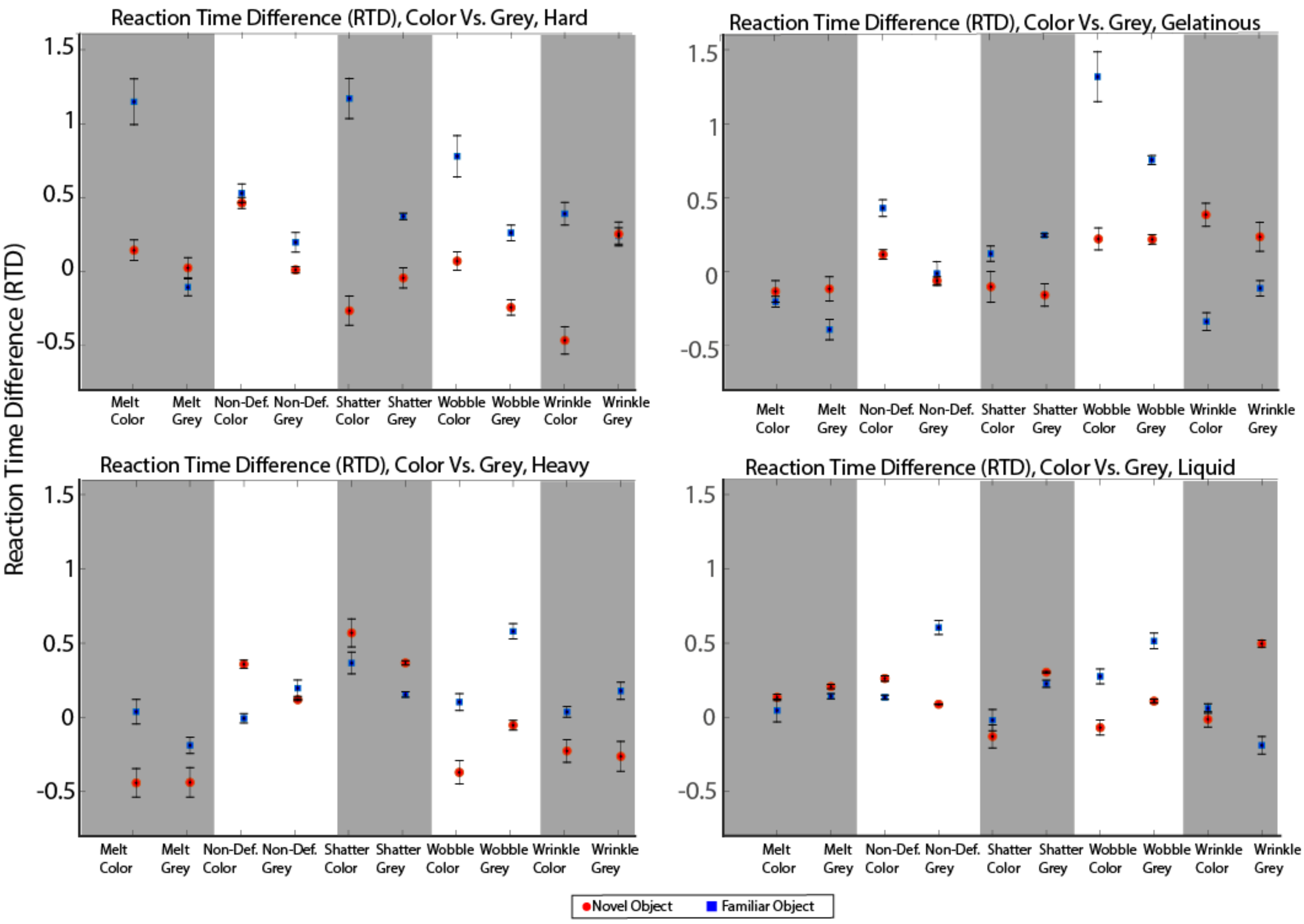


Figure 35. Reaction Time Difference Comparison Plots. Figure 35 plots the reaction time difference (RTD, Surprising minus Expected) for all rating adjectives (Hard, Gelatinous, Heavy, and Liquid) by their deformation methods (Melting, Non-Deforming, Shattering, Wobbling, Wrinkling), comparing the greyscale data from this experiment to that of the full-color data. With certain exceptions, the reaction time differences are rated similarly between color and greyscale.

Figure 35 plots the reaction time difference (RTD, Surprising minus Expected) for all rating adjectives (Hard, Gelatinous, Heavy, and Liquid) by their deformation methods (Melting, Non-Deforming, Shattering, Wobbling, Wrinkling), comparing the greyscale data from this experiment to that of the full-color data from the previous experiment. With certain exceptions, the reaction time differences are rated similarly between color and greyscale. These exceptions tend to occur in the 'Wobble' conditions, whereby the lack of optical properties in the greyscale condition leads to greater ambiguity, and thus, a greater spread of reaction time difference scores.

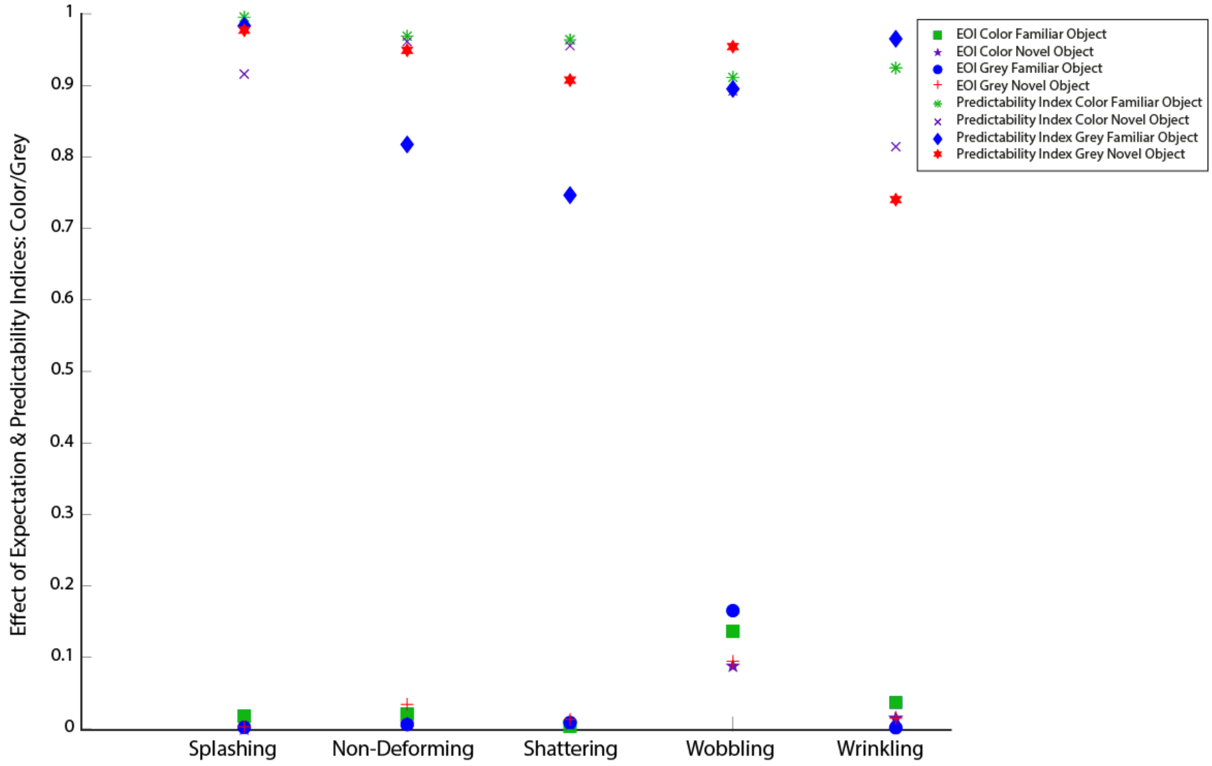


Figure 36. Mean Index Comparisons. Figure 36 plots the average Effect of Expectation Indices for each deformation method (Melting, Non-Deforming, Shattering, Wobbling, Wrinkling) for both Color and Greyscale experiments, in addition to the Predictability Scores for both Color and Greyscale Experiments. Values for each of these scores tended to be rated similarly between color and Greyscale experiments, providing further support for the idea that in the absence of color, observers may access their prior knowledge about typical color of these objects, leading to similar ratings. These values are the most distinct from one another for Wobbling and Wrinkling items, which is consistent with increased levels of ambiguity for these objects in the absence of color.

*Figure 36. Mean Index Comparisons***.** As in the previous figure, we wanted to compare our two mean indices for color and greyscale items in their respective conditions. Figure 36 plots the average Effect of Expectation Indices for each deformation method (Melting, Non-Deforming, Shattering, Wobbling, Wrinkling) for both Color and Greyscale experiments, in addition to the Predictability Scores for both Color and Greyscale Experiments. Values for each of these scores tended to be rated similarly between color and Greyscale experiments, providing further support for the idea that in the absence of color, observers may access their prior knowledge about typical color of these objects, leading to similar ratings. These values are the most distinct from one another for Wobbling and Wrinkling items, which is consistent with increased levels of ambiguity for these objects in the absence of color.

Considering the differences in Predictability scores between color and greyscale (upper portion of figure), colored familiar objects (stars) are more predictable than grey familiar objects (diamond) in 4/5 cases. The remaining instance where the converse is true occurs for Wrinkling items. This is consistent with our results: As discussed previously, the shape of wrinkling items were difficult to rate and differentiate. The predictability scores of novel items between color (x-shape) and greyscale (hexagon) show that colored novel items were more predictable than their greyscale counterparts in 3/5 cases. In the cases where this effect is not found (splashing, wobbling), it is possible that optical properties are largely driving the effect, and in their absence, the shape alone was an ambiguous cue. Considering the Effect of Expectation Index (lower portion of figure), we find a greater difference between expected and surprising ratings for Familiar Objects (square) than for Novel Objects (circle) in 3/5 cases. For the cases where we find the opposite effect (shattering, wobbling), it is possible that the lack of optical properties caused judgements to become more ambiguous. This is especially interesting in the wobbling cases: Here, we find a greater difference for grey familiar objects between their expected and surprising outcomes than for their colored counterparts.

Discussion

In this experiment, we tested how our results from the previous color experiment differs in the absence of 'typical' color. Ratings and reaction times were found to be similar for full-color and greyscale stimuli, contradicting our hypothesis that, in the absence of color, ratings and reaction times would differ greatly as a function of the decrease in available optical input. These findings are consistent with the idea that prior knowledge (e.g. regarding typical color/optical properties) influences perception. In the absence of 'typical' color information, the observer must make rapid decisions about an object (under uncertainty) based predominantly on shape cues and object priors.

While we average over the three objects in each deformation condition (as the differences in our findings may differ for each object), here it is interesting to consider the rationale of how each object category differs as a function of the removal of optical properties and typical color. With full color and optical property information available, it is easy to tell that certain items (e.g. clay pot or teacup) will shatter upon impact, as their materials are fragile. However, in greyscale, such information regarding typical color and optical qualities is removed. This means that for certain objects, the 'Expected' deformation has been made more ambiguous: What was once a (more obviously) shattering clay pot and teacup could now just as easily be made of plastic (as if it were a child's toy) and therefore would not be expected to break.

In contrast, the expected motion for other items in this data set remain comparable based on shape, even in the *absence* of typical color: Items that are shaped like jellies will wobble, while liquids would be expected to splash. Rigid objects may or may not break: A glass chair would shatter, while a metal chair would not. A greyscaled image of a spoon would still be unlikely to break, given that spoons are almost always made of rigid materials (be they plastic, metal, or wood). In this way, prior knowledge about the ‘typical’ optical properties is applied, supporting our finding of similar ratings. These findings would support the findings of Hansen et. al.’s ‘blue banana’, suggesting that priors on object color may affect the ratings provided by observers to our stimuli: In the absence of color, objects are rated similar to their ‘typical’ colors and object properties.

The fact that reaction times are similar between color and greyscaled familiar object stimuli and tend to cluster together (with certain exceptions of familiar objects around a 1 second reaction time difference) provides further support for the finding that these objects are rated similarly under ambiguity. As noted in the Full-color experiment, it is surprising that we find such a reaction time effect, given that our response method leaves a significant amount of time (relative to the setup of the experiment), we have still found the presence of such a reaction time effect.

Conclusion

Consistent with the previous Full-Color rating experiment, this work shows that (even in the absence of color), prior knowledge of kinematic properties of materials can be activated by the familiar shape of an object, but also may be influenced by the prior for the ‘typical’ optical qualities of a surface. Our results imply that perceived material qualities are not only determined by the retinal stimulation, but instead can also be susceptible to cognitive influences, such as expectations and memory.

CHAPTER 3: Control Experiments

Experiment 3.1: 'Questions' Experiment

In this control experiment, we investigated whether our stimuli were identified by the participants as we intended via a free-response naming task. Using a set of three questions, participants provided their responses in writing. We find that overall, participants identified the objects as we intended, and largely agreed with one another.

Methods

Stimuli

Participants were shown the static first frame of each of the Familiar Objects and Novel Objects, (and separately, Familiar Object and Novel Object Surprising motion). They were then asked to provide written responses to identify each stimulus. Using the responses collected, we sorted the responses into four main groups: 'Hard Items', 'Liquid Items', 'Soft-Deforming Elastic Items', 'Soft-Deforming Non-Elastic Items', and 'Cloth/Fabric Items'. Stimuli: Stimuli used in this experiment were rendered at a lower quality than the previous experiments, and were presented with mid-gloss wooden floor and green background.

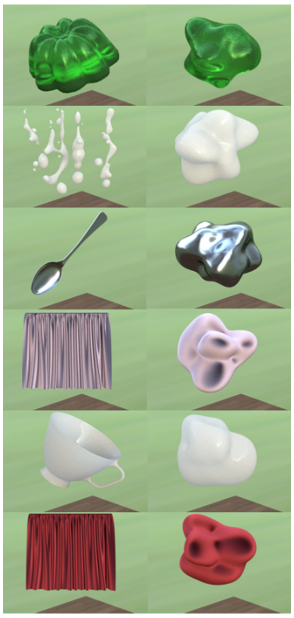


Figure 37. Exemplars of Stimuli for 'Questions' Experiment. A selection of stimuli from those used in Experiment 3. These stimuli used the same objects and animations as those described previously, but provided a different background against which the objects were presented. The mid-gloss wooden floor remained in the scene for these stimuli. As in the previous stimulus set, lighting remained constant (governed by 'Campus' HDR probe). *Animations contributed by A. Schmid/Figure created by L. Alley.*

Apparatus

The experiment was coded in MATLAB 2015a (MathWorks, Natick, MA) using Psychophysics Toolbox (version 3.8.5), 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

In Experiment 3, participants always completed the static 'First Frame' of the experiment first (equivalent to First Frame experiments discussed previously). In doing so, their responses were not affected by the observation of motion. Unlike the previous experiments, 'Novel Objects' were always shown as the first block of the 'First Frame' experiment. In doing so, participants were able to judge the optical qualities (material) of the Novel Object objects, without prior interference from the familiar object shapes. Participants viewed each stimulus in succession, and for each trial provided written responses to the question:

"What material might this object be made of?"

"On a scale of 1 – 10, how sure are you of the above response?"

"If this object were to fall, how might the material deform?"

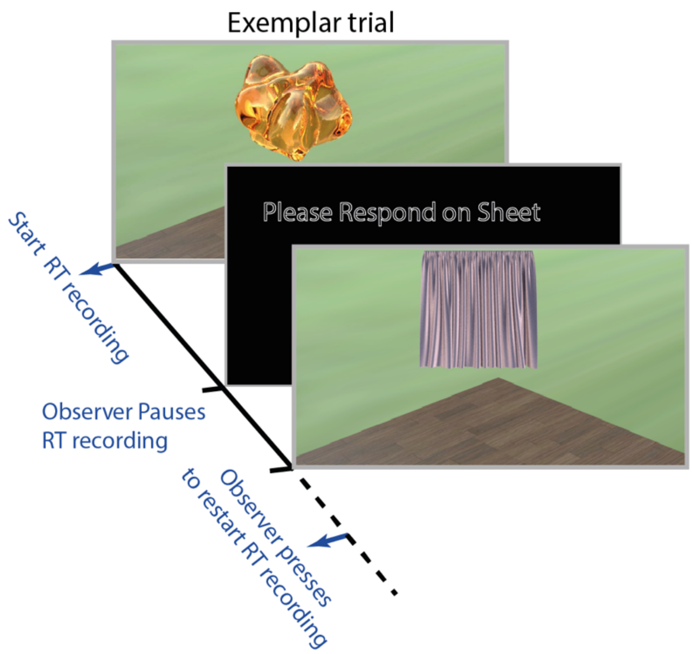


Figure 38. Trial Diagram. Participants provided their responses in writing in this free-naming task. Participants viewed each stimulus in succession, pausing after each to provide their responses in writing. In this way, we are able to ascertain desired amount of time take to observe each stimulus.

Participants provided their responses in writing in this free-naming task. Participants viewed each stimulus in succession, pausing after each to provide their responses in writing. In this way, we are able to ascertain desired amount of time take to observe each stimulus.

Participants

12 naive participants (mean age 25.8; 8 female) participated in both the ‘First Frame’ and ‘Full Motion’ conditions of Experiment 3; 11 were right-handed. All participants were native speakers of German language, and had self-reported normal or corrected-to-normal vision. All participants provided written informed consent, and were reimbursed at a rate of €8 per hour.

Analysis

All responses were provided in German and translated to English for analysis by native speakers of German language. Based on responses given by participants, the data were organized into five categories: ‘Hard Items’, ‘Liquid Items’, ‘Soft Deformable Elastic Items’, ‘Soft Deformable Non-Elastic Items’, and ‘Cloth/Fabric Items’.

Results

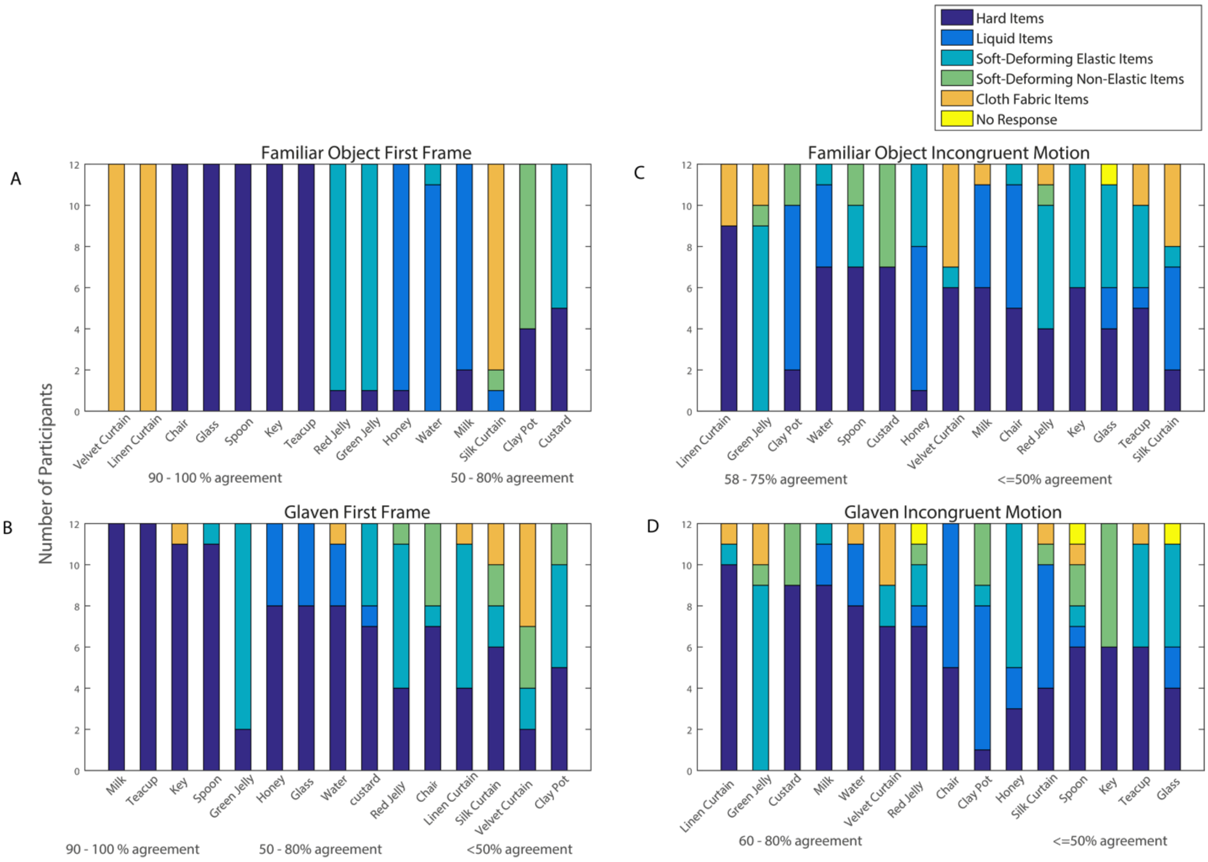


Figure 39 a-d. Results of 'Questions' Experiment. Stacked bar plot of responses broken down by category: 'Hard Items', 'Liquid Items', 'Soft Deformable Elastic Items', 'Soft Deformable Non-Elastic Items', and 'Fabric Items'. These graphs indicate that participants were identifying the objects as we intended, and the results tended to cluster into their own object categories.

Discussion

We presented a series of objects to participants and asked them to freely name the material of which the object the object could be composed. Figure 39a plots responses to the 'Familiar Object First Frame' condition, where judgements were made based on the static first frame of the image alone (a judgement of shape and optical properties). Figure 39b plots responses to the 'Novel Object First Frame' condition, where critical information regarding shape was removed, and judgements were made on material alone. Taken together, this suggests that when information regarding object shape was present, responses regarding the identity of the object and material were largely in agreement, while when information regarding shape was removed, responses regarding the identity and material of the object varied to a greater degree. Figure 39c plots responses given by the participants to the 'Familiar Object Surprising Motion' condition, where recognizable objects were shown to deform in a manner that violated their expectations. Figure 39d plots responses given by the participants to the 'Novel Object Surprising Motion' condition, where both the shape and the motion shown were unfamiliar. From the results in Figures 39c-d, we see two main patterns of responses. Because participants were not given explicit instruction as to on which frame they should base their decision, certain participants gave a response based on their impression of the intact object at the start of the video, while others provided their responses based on the motion outcome observed (i.e. because the chair was shown to Surprisingly melt, many participants gave responses in the 'liquid' category). For the Novel Objects, we find a similar pattern of results, suggesting that for certain items, the optical qualities were salient enough to allow the participant to identify the intended material of the object, irrespective of the novel shape of the Novel Object. We indeed find that most Familiar Objects were identified as we intended, and that this recognition ability, in many cases, does not extend to novel control objects (as we would predict).

These graphs indicate that participants were identifying the objects as we intended, and that the results tended to cluster into their own object categories, consistent with previous findings on object and material categorization. Following from this finding, we additionally investigated the rating-based method of response, and asked how our results differ as a function of response input—How will our results change if participants are asked to make rapid, two-alternative-forced-choice (2AFC) responses?

Experiment 3.2 Full-Color versus Greyscale 2AFC Experiment

In order to investigate these stimuli without using a rating bar, we developed a Two Interval Forced Choice (2 AFC) experiment which allowed us to investigate whether the RT effects we find in the previous experiments are due to the re-positioning of the rating bar, or whether the RT effect we find is an effect of the stimulus presentation. Overall, we find that reaction times tended to be faster for the greyscaled stimuli than for the exact same judgements with full-color stimuli. We suggest that this difference is related to the decrease of informative cues to material identity, requiring additional reliance on the prior.

Methods

Stimuli

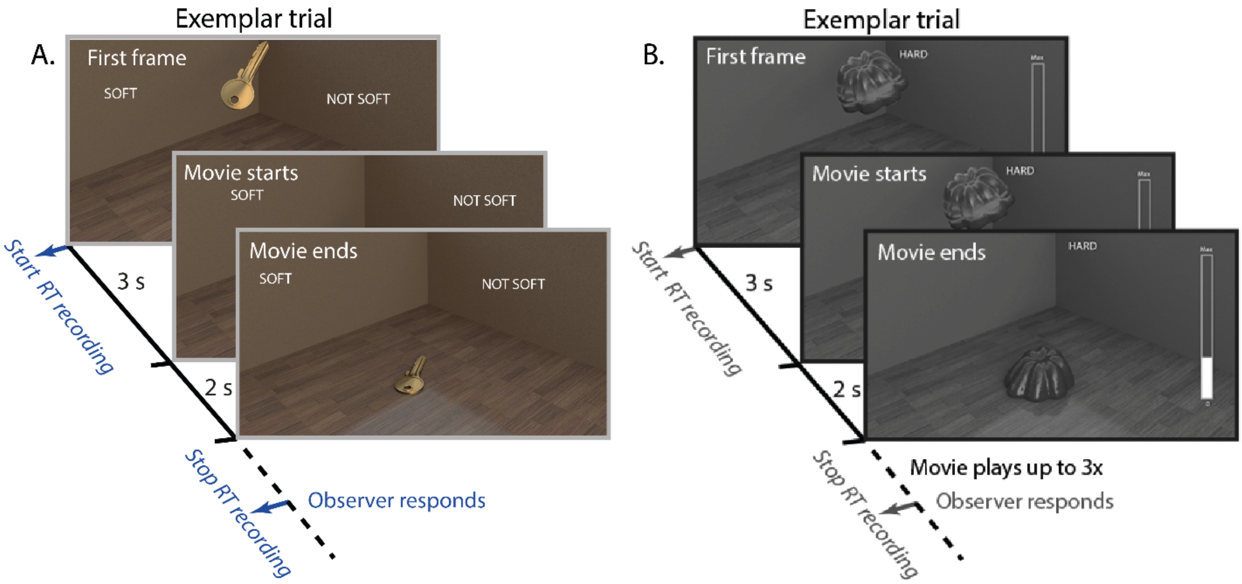


Figure 40 a-b. Color versus Greyscale Stimuli. Figure 40a shows the full-color version of the 2AFC experiment. Observers were asked to perform the 2AFC experiment, using the stimuli described in Chapter 2. Figure 40B: A separate group of observers were asked to perform the same experiment, using greyscaled versions of the stimuli described in Chapter 2. The stimuli were greyscaled using 'rgb2gray' in MATLAB.

Apparatus

The experiment was coded in MATLAB 2015a (MathWorks, Natick, MA) using Psychophysics Toolbox (version 3.8.5), 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

Full-Color Version: Participants were shown instructions on-screen prior to the start of the experiment. Critically, participants were asked to be as fast but as accurate as possible when making their decisions. In each block, an adjective would appear at the top of the screen and be presented consistently, reminding the participant on which adjective they were judging. Participants were shown the object, which would hang still in mid-air for three seconds, after which time the object would fall and impact the ground in a manner that was either expected or unexpected. Using the left and right arrow keys, participants made a response to the following three questions (in blocks):

1. (Is this object) Soft or Not Soft?
2. (Is this object) Liquid or non-Liquid?
3. (Is this object) Rigid or non-Rigid?

Greyscale Version: The Task and Procedure for this experiment followed the same procedure as described above, with the additional modification that the stimuli had been greyscaled to remove all color information, thus forcing participants to make judgements based on shape and motion cues, removing information for color and most surface properties.

Participants

Full Color Version: 12 participants (mean age 28.9; six female) participated in both the 'First Frame' and 'Full Motion' conditions of Experiment 3; 11 were right-handed. All participants were native speakers of German language, and had self-reported normal or corrected-to-normal vision. All participants provided written informed consent, and were reimbursed at a rate of €8 per hour.

Greyscale Version: 12 participants (mean age 31.1; 12 female) participated in both the 'First Frame' and 'Full Motion' conditions of Experiment 3; 11 were right-handed. All participants were native speakers of German language, and had self-reported normal or corrected-to-normal vision. All participants provided written informed consent, and were reimbursed at a rate of €8 per hour.

Analysis

Data for both color and greyscale experiments were analyzed by first excluding data points which fell below the lower cutoff of 1.45 seconds (indicating that the participant could not have seen the point of impact of the object)) and/or above the upper cutoff of 3.45 seconds (two standard deviations of the mean)). As before, we averaged over adjective and object identity such that we consider only the deformations of the object, and not the effect of the question posed.

Results

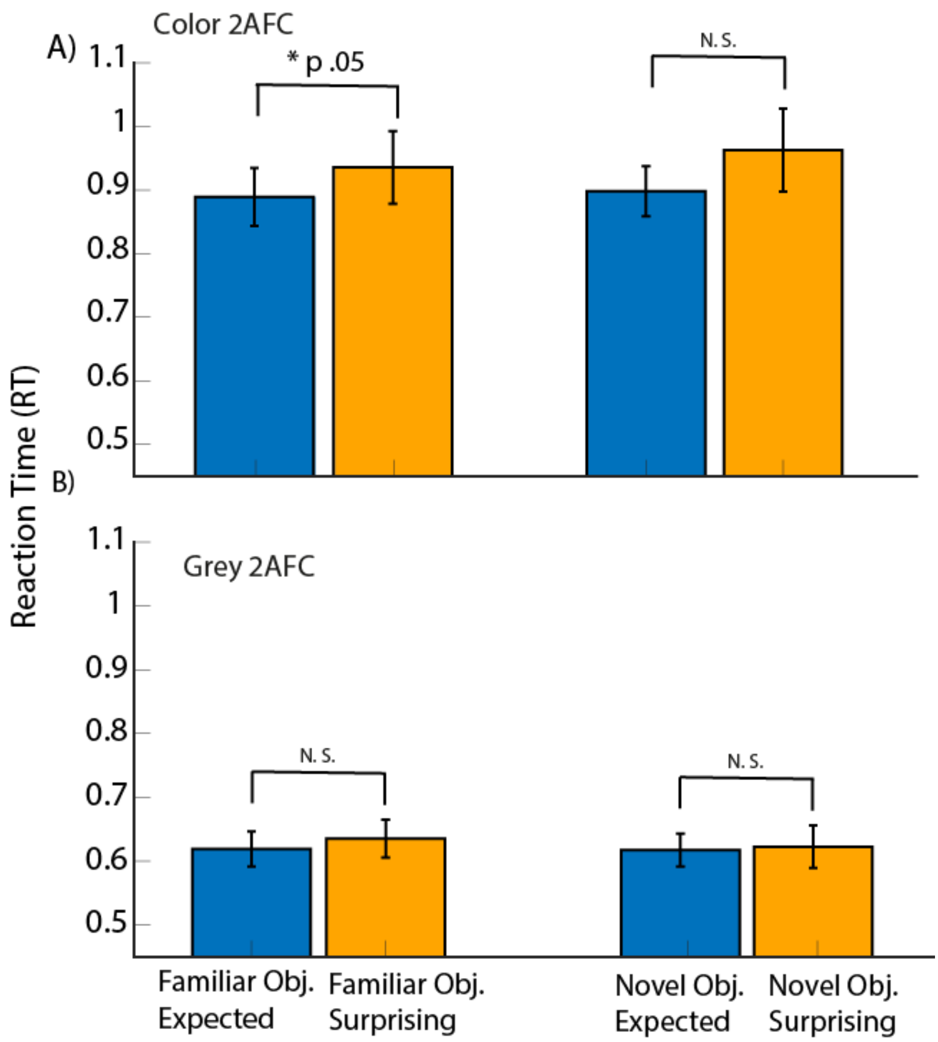


Figure 41 a-b. Comparing 2-AFC Reaction Time Differences for Colored and Greyscale Stimuli. Figure 41a depicts the differences in Reaction Time (RT) between the Expected and Surprising outcomes of Full-Color Familiar and Novel objects. Results show a significant difference between the Expected and Surprising outcomes for the Familiar Objects, but not for the Novel objects, which is consistent with our previous findings. Figure 41b depicts the differences in Reaction Time (RT) between the Expected and Surprising outcomes of greyscaled Familiar and Novel objects. Results show no significant differences between the Expected and Surprising outcomes for either the Familiar or Novel Objects in the absence of color.

For colored stimuli, paired t-tests revealed a significant difference between the reaction time differences for expected and surprising outcomes of Familiar objects ($t(15)=-2.8$, $p=.01$), see Figure 41a. Consistent with previous findings, we did not find significant differences in reaction time between the outcomes of Familiar and Novel objects ($t(15)=-1.06$, $p=.30$).

Surprisingly, for greyscaled stimuli, we find no significant differences in the reaction time differences between Expected and Surprising outcomes of Familiar objects ($t(14)=-3.43$, $p=.40$). We also find no significant differences in the reaction times between Familiar and Novel objects ($t(14)=-3.43$, $p=.73$), see Figure 41b.

For colored stimuli, we used a two-way Analysis of Variance to compare the main effects of Object Type (Familiar Object, Novel Object) and Motion Type (Expected Outcome, Surprising Outcome) on reaction times. No statistically significant effects of Object Type or Motion Type were found.

Discussion

We investigated the effect of color versus greyscale stimuli on the reaction times of Familiar and Novel stimuli, with both Expected and surprising outcomes. We found a significant difference between the reaction times for the Expected and Surprising outcomes of the Familiar Object colored stimuli. This finding is consistent with our previous reaction time difference findings in the rating data.

The fact that we (surprisingly) found no significant differences in reaction time between Familiar and Novel (both colored and greyscaled) objects outcomes may be a result of the fact that in this two-alternative-forced-choice paradigm, the response method is simply too fast to capture differences—in this experiment (in contrast to the rating experiment) rather than manipulate a rating bar and confirm your response, here, the participant merely presses a single key to indicate a (rapid) binary decision, e.g. ‘Is this SOFT or NOT SOFT?’.

Consistent with the greyscaled rating experiment described previously in this chapter, the lack of sufficient color information in greyscaled items leads the observer to make rapid decisions about the object based on shape cues and the influence of expectations and priors regarding typical optical properties, leading to a rapid judgement even under uncertainty. This hypothesis would be consistent with Biederman, 1987, which suggests that shape cues are the dominant cue, given that it is earliest processed.

The differences in our findings may differ for each object, based on their material utility alone. With information available regarding color and optical properties, it is easy to tell that certain items (e.g. clay pot or teacup) will shatter upon impact, as their materials are fragile. However, in greyscale, such information regarding typical color and optical qualities is removed. This means that (also as previously discussed) for certain objects, the 'Expected' deformation has been made more ambiguous: What was once a (more obviously) shattering clay pot and teacup could now just as easily be made of plastic (as if it were a child's toy) and therefore would not be expected to break.

In contrast, the expected motion for other items in this data set remain clear based on shape, even in the absence of typical color. Items like the rigid objects, jellies, and liquids are still clear: Items that are shaped like jellies will wobble, while liquids would be expected to splash. Rigid objects may or may not break: A glass chair would shatter, while a metal chair would not. A greyscaled image of a spoon would still be unlikely to break, given that spoons are almost always made of rigid materials (be they plastic, metal, or wood). It is interesting that greyscaled Familiar and Novel objects are reacted to at the same speed. This supports or hypothesis regarding increased ambiguity as a function of loss of typical color information, even in the presence of familiar shape.

Why do we find a significant difference in reaction times between the Expected and Surprising motions of Familiar Objects for colored stimuli, but not for greyscaled stimuli? We can look to the literature on the expected (typical) color of objects for an explanation. The effect of typical or expected object color on reaction time has been largely documented in color perception literature (see Witzel, 2011, for a review). Following from this finding, we further investigated the method of response, and asked how our results differ as a function of response input—How will our results change if, rather than asking participants to make rapid, two-alternative-forced-choice (2AFC) responses, we ask that they simply detect a change in a single stimulus property, without calling attention to specific properties of the stimulus? Rather than directing attention toward a specific property ('Is this stimulus green?') the following experiment asks the participant to provide their response as soon as they observe an object impacting the floor.

Experiment 3.3: 'Touch' Experiment

In Chapter 2, we demonstrated an effect of reaction time when making rating judgments for material properties. Participants responded faster to objects that deformed as expected, compared to deformations that were unexpected. In these experiments, participants were asked to fill up a rating bar to indicate their response (adjective-based material quality assessment). It is surprising that (given the significant amount of time involved between observing an event and securing one's rating response) we see such a striking difference between the reaction times of expected and unexpected events. To investigate this observation/ identify to what extent our findings are affected by response method and/or adjective judgement, we designed an experimental paradigm to investigate non-rating-based responses to unexpected stimuli.

To eliminate secondary expectations/priors developed over the course of the experiment regarding location and timing (i.e to prevent the participant from habituating to the timing/screen presentation of the falling object), we manipulated the following three experimental parameters: 1. Scene context, 2. On-screen location of video, and 3. initial falling frame. We manipulated scene context by rendering our videos without walls, background, or floor—each object was shown to fall into 'empty' black space, and deform upon impact with a floor that was invisible (the physics of the object interacting with the floor -- and deforming -- remained intact). Manipulation of video location included dividing the screen into 100 possible locations, and selecting a screen location at random on each trial for the video to be presented. In this way, the participant is unable to habituate to/develop a prior for a consistent location on screen. Manipulation of initial falling frame included starting each video at a random start frame prior to impact (Frame 1, 3, 5, 7, or 9). In this way, the participant is unable to habituate to (develop a prior for) a consistent and repetitive timing of button-presses (which may affect reaction time results).

Here, we use a simple 'press-on-impact' task in conjunction with a violation-of-expectation paradigm to investigate whether an effect of reaction time exists when observing unexpected material deformations, without considering the relationship between a deformation and a rating adjective. As before, participants watched familiar objects fall, impact the ground, and deform in a way that was either Expected or Surprising.

In each block, the Expected outcomes of all objects were shown in a random order. This was then followed by a single, randomly-selected (oddball) 'Unexpected Outcome' trial. This process was repeated until all Unexpected outcomes of all objects were shown once, resulting in 15 blocks of 16 trials each (240 total trials). We again hypothesized that (absent a rating adjective (and, consistent with our previous findings) a familiar object that deforms on impact as you would *not* expect (e.g. a shattering drop of honey) would result in a longer reaction time when making a response, compared to judgments where the object/substance deformed as expected (e.g. when the honey drips).

Methods

Stimuli

Stimuli used in this experiment were adapted from those used in Chapter 2. For specific stimuli parameters, we refer the reader to the aforementioned chapter. The background of these stimuli were modified to remove scene context. The background and floor of all scenes were removed, leaving the object to impact an unseen floor. As (from our previous findings) we did not expect Novel Objects to elicit expectations, this set of stimuli was not rendered for this experiment.

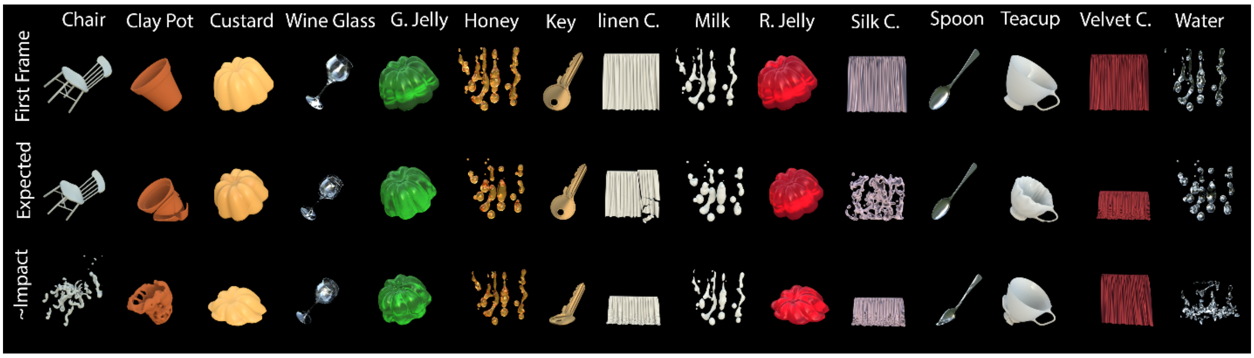


Figure 42. Exemplars of stimuli used in this experiment. These stimuli and their animations are identical to those described in Chapter 2, with the exception that the context of background and floor have been removed. This removal prevents awareness of the location of the floor, where participants must rely on their observation of the shape. Row 1 depicts the first frame of all objects. Row 2 shows the 'Expected' outcome of all objects (near the frame of Impact). Row 3 shows the 'Unexpected' outcome of all objects (near the frame of Impact). It is around these frames that the participants were instructed to press a key (indicating frame of impact).

Apparatus

The experiment was coded in MATLAB 2015a (MathWorks, Natick, MA) using Psychophysics Toolbox (version 3.8.5), 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

An instruction screen was presented prior to the start of the experiment. Participants were instructed to press the spacebar as soon as they observed the object impacting the floor. As in previous experiments, participants were reminded to be as accurate but as fast as possible when making their response.

On every trial, an object appeared (see Figure 43) and remained still for a randomized duration between 0 - 3 seconds to allow the participant to identify the object and to potentially activate corresponding expectations. This duration was randomized in order to prevent the participants from developing an expectation as to when the object would begin to fall. We additionally randomized the frame at which the video would start (1, 3, 5, 7, or 9), as an additional control to prevent the participant from habituating to the falling speed of the object. As a third additional expectation-based control, the on-screen location where the stimulus would appear was randomized for each trial (for each participant), resulting in 100 possible locations. On each trial, the object fell and impacted the floor, deforming either as expected, or in a manner that was unexpected. The participants indicated their response by pressing a key as soon as they observed the object impacting the floor. The next trial would begin immediately after the keypress. Participants were allowed to take breaks at set points during the experiment. The order of all stimuli in all blocks was randomized for all participants.

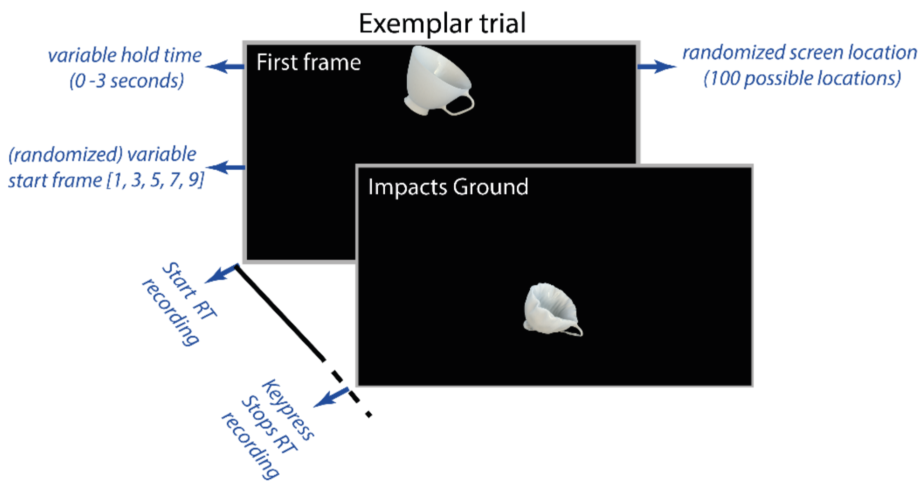


Figure 43. Exemplar of a single trial. The first frame was shown for between 0 - 3 seconds. The object then fell and deformed in a manner that was either Expected or Unexpected. The participants pressed a key as soon as they observed the object impact the floor. The location of the object on screen varied (100 possible locations), to prevent the participants from habituating to the location and timing of the experiment. For this purpose, the size of the videos presented on-screen was scaled down to approximately 6 degrees of visual angle (dva).

Participants

14 naïve participants from JLU Giessen (mean age 24; 9 female) participated in the experiment; 13 were right-handed. All participants had self-reported normal or corrected-to-normal vision. The experiment followed the guidelines set forth by the Declaration of Helsinki. All participants provided written informed consent and were reimbursed at a rate of **€**8/hour.

Analysis

As each object was held static for a variable duration of between 0 - 3 seconds (and the start frame varied on each trial), we subtracted the total amount of time prior to impact (per trial) from the total reaction time value. In this way, the analyses that follow consider time at and after impact only (as our main goal was to investigate whether or not a time difference existed between trial types for this time frame). Cases where the point of impact was not seen (8 cases, .0024 % of data), as well as cases where participants responded significantly after the point of impact (1.5 seconds, two standard deviations above the mean), found in 67 cases (.019 % of data) were excluded from our analysis.

Results

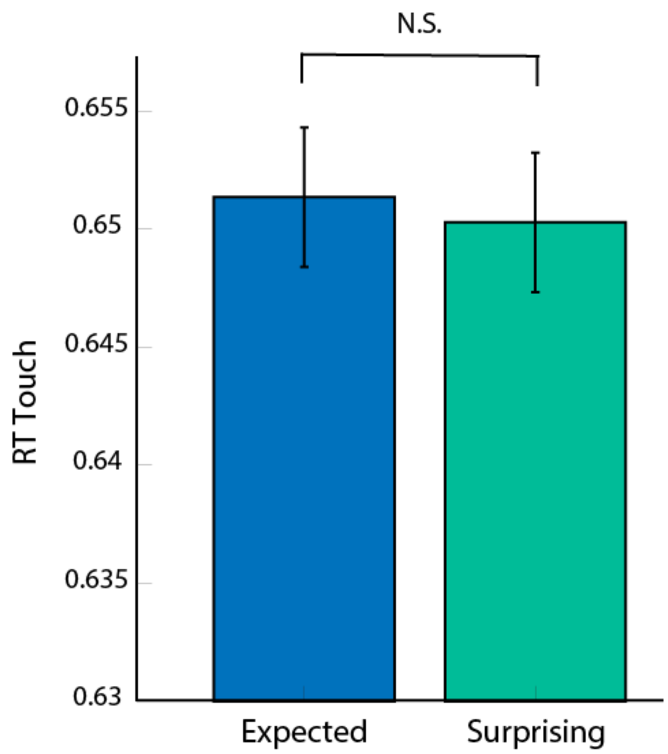


Figure 44. Reaction Time (RT) Difference between Expected and Unexpected Outcomes. No significant differences were found between Expected and Surprising trials.

As shown in Figure 44, paired t-tests revealed no significant differences between the reaction times for expected and surprising outcomes ($t(12)=7.38$, $p=.94$). As this experiment contained Familiar object stimuli only, we cannot compare the difference between Familiar and Novel objects. We would anticipate Novel object results to be consistent with previous experiments which suggest less 'Surprise' reaction time to these objects.

Discussion

After finding no significant difference between the reaction times for Expected and Surprising outcomes, we sought to consider differences in task which may affect our findings. To efficiently perform this task (which, in essence required participants to 'look for' the (invisible) 'floor', rather than attending to the identity of the object directly), participants likely attended to the shape cue as the most dominant cue—as soon as a shape boundary began to deform, the 'floor' has been found, and the button could be pressed. Rather than identifying object type (wine glass, custard, spoon, etc.) the participant simply must attend to the deformation of shape outlines. If true, we would expect that this finding holds for all shapes (familiar or unfamiliar) to an extent, meaning that our findings would hold true for line drawings, as well as complex familiar shapes such as these. Broadly, one of the reasons we may not find a significant difference is due to the design of the experiment, as the response is rapid and arguably far more perceptual based on shape, rather than cognitive (does not require consideration as the adjectives do). Some knowledge of (expected) shape may play a role.

In all previous experiments, the floor was located at a constant, known location, and did not vary. In this experiment, although the object location was varied at a number of positions on screen (and time to impact varied), our results do not suggest that these modifications induced any significant differences against habituation to falling time. If participants had habituated to timepoints or location, we would have observed patterns of similar reaction times between consecutive trials, which we do not find in these results. As the total time prior to impact varies between trials (impact could occur as early as two frames after the first frame), future analyses could consider the differences in reaction time as a function of presented duration (2 – 72 frames) prior to impact. Additional further reaction time analyses could consider whether a delay in reaction time was present for the trial *after* the 'Surprising' trial. If so, this may relate to a delay in processing still present and affecting future trials, given that the 'Surprise' event is then also temporally surprising. As the design of this experiment guarantees that (by the end of the experiment), each participant will have seen the Unexpected outcome of each object identity once, we have eliminated the previous confound of developing a prior over the entire experiment for Unexpected trials (see Supplementary Figures 6 and 7 for comparison), and thus, achieving faster reaction times for multiple presentations of the same Unexpected outcome for each object.

Following from this finding, future experiments which may investigate this task in relation to our stimuli more specifically would involve asking the participant not *when* the object impacts the floor, but rather a question relating to e.g. material mechanics: 'Will this object break (or not)?'in this way, the participant will have to consider the mechanics of the object material, but not in a way as detailed the relationship to an attribute (Shininess, Wobbliness, Furriness, etc.), and not as simplified as locating the floor. In this way, attention to the object identity is also maintained (rather than attention to the scene).

CHAPTER 4: Eye Tracking

Introduction

The second segment of this thesis sought to ask where people look when making judgements about material—especially that of familiar objects. Which areas of the object are most critical for making material property judgements? How do fixations differ between objects, or for different regions of a similar object or scene? What role does motion play in where we fixate? These are the broad questions that still remain unanswered when investigating the relationship between eye movements and material perception.

Eye Tracking a Non-Deforming Set of 'Surprise' Stimuli: A Preliminary Investigation

After obtaining quite striking results from the experiments on our set of 'Surprising' stimuli, the logical next step is to ask *where* participants are fixating when making these rating judgements, and specifically *which* aspects of the stimuli are most important for each of these judgements.

The stimuli used in this preliminary investigation consisted of a subset taken from those described in Chapter 2 (Chair, Green Jelly, Spoon, Velvet Curtain). We refer the reader to that section for specific parameters. We chose a set of these stimuli, specifically focusing on those objects that did not break apart into multiple pieces. As participants could choose to fixate on individual shards of clay pot or glass, we wanted to make our initial observations with more simplified motions. We hypothesized that optical properties that have been previously described in studies of material perception (gloss, specular highlights, areas of high contrast, etc.) would garner the most attention and provide the most useful information for this task.

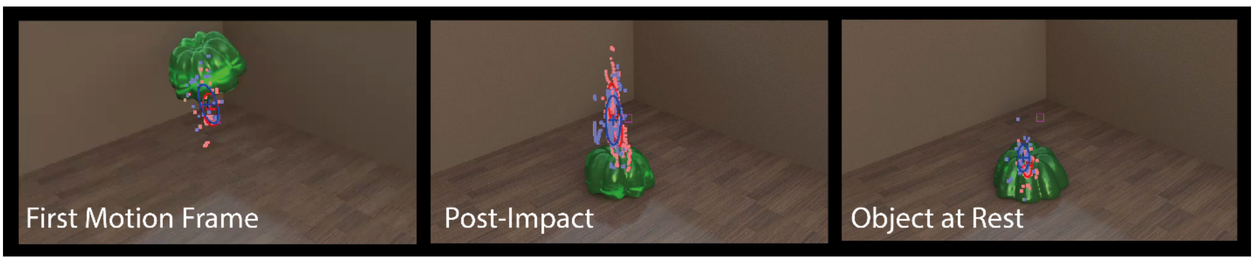


Figure 45. Analyses of Preliminary Eye Tracking Investigation. Shown is one example of the fixations plotted on critical frames (First, Impact, End) of one trial for one participant. Red points indicate fixation locations for judgements of softness, while blue points indicate fixation locations for judgements of lightness. Red and Blue ellipses indicate the centroid of all fixation positions for that task and trial. *Figure created by L. Alley.*

As a first step, we plotted the individual fixations for each task (per frame) for a set of three observers and visually observed these videos of these fixations for any striking patterns in the data. Fixations for softness are plotted as red dots, while fixations for judgements of lightness are plotted as blue dots. We find that for this setup (by simply visually observing fixation patterns), participants tended to exhibit a central fixation bias-- whereby they fixated the object and followed it down to (and slightly anticipated) the floor (cf. Alley et. al., (2011)). Whether this is due to the size of the object (in dva), the speed of identifying the object, or the lack of necessity to consider the object in-depth, will require further investigation.

Given this observation, we aimed to develop a more simplified set of stimuli that would allow us to carefully investigate the perception of material in motion.

Experiment 4.1: “Pilot Blobs”

As the previous set of stimuli generally resulted in a central fixation bias, we generated a new, simplified set of stimuli. These stimuli were created to investigate the perception of motion, material, and shape by poking the object in a specific location, preventing a central fixation bias. Rendering these stimuli with matte greyscale optical properties also allows us to remove the effect of optical properties that affect the perception of material and focus only on material motion (degree of wobble).

Methods

Stimuli

Stimuli were 11 greyscale blobs, rendered in Blender 2.77a (Stichting Blender Foundation, Amsterdam, NL). These blobs varied in their perceptual degrees of stiffness, and were illuminated with 11 different albedos. For each stimulus, lighting remained constant on the blobs, and the blobs were illuminated from a single location (above the object from the left). Figure 46 shows the set of all stimuli.

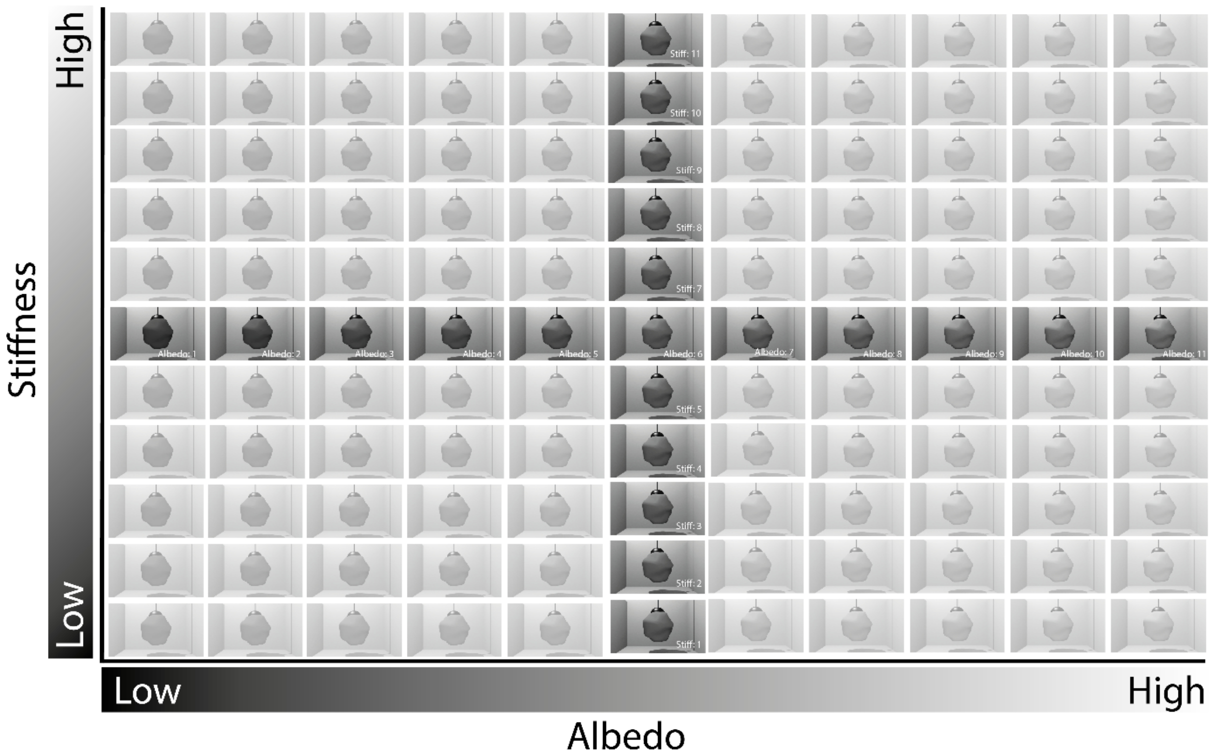


Figure 46. Pilot Blobs Stimuli. A. The set of all stimuli for Experiment 4.1. Each video would play once, followed by the comparison video. The participants responded with the left or right mouse button. *Stimuli provided by M. Toscani.*

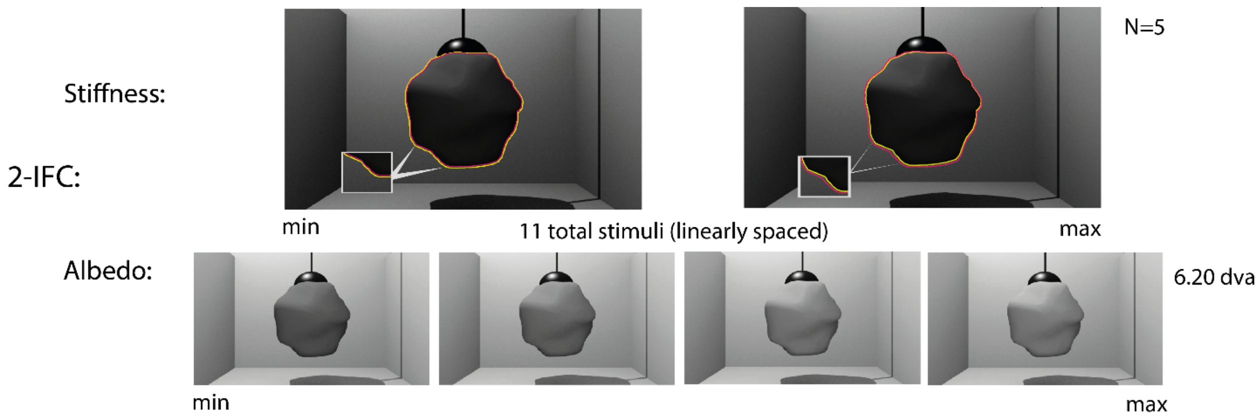


Figure 47. The range of all stimuli: Pilot Blobs. Each video would play once, followed by the comparison video. The participants responded with the left or right mouse button. *Stimuli provided by M. Toscani.*

Apparatus

Stimuli were presented on an LCD monitor (Cambridge Research Systems Ltd, Rochester, UK). Participants placed their heads to a chin rest to ensure that a constant viewing distance (48 cm) was maintained. Screen resolution was 1920 × 1080 pixels (2.743 pixels/ mm; 22.976 pixels/degree of visual angle (dva). Gaze position on screen was monitored online at 500 Hz with an EyeLink II system (SR Research, Mississauga, Ontario, Canada). Drift correction was performed by the experimenter on every trial. If the observer's gaze was not detected within 1° from the instructed fixation position, the trial was repeated.

Task and Procedure

Participants were asked to perform a two-Interval-Forced-Choice (2IFC) experiment via a QUEST staircase, containing 1000 trials in two blocks (500 trials per block). Participants always performed the stiffness block first, followed by the lightness block. Participants knew beforehand which attribute they would be asked to judge.

On each trial, a single video played for the full two-second duration. The eye tracker would then drift check, and the comparison video would play for the full two-second duration. A screen would then show, asking the participant which was stiffer (in the first block) or lighter (second block), and the participant would respond using the left or right mouse button. The participant saw all possible pairs (combinations) of all stimuli. Figure 48 shows a sample trial.

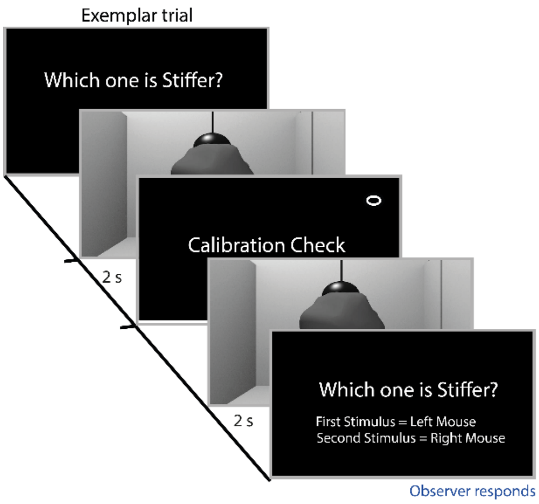


Figure 48. Trial Diagram. Example of a single trial. Each video would play once, followed a drift check, and the video to be compared. The participants responded with the left or right mouse button, indicating which of the videos contained objects which were stiffer/lighter than the comparison.

Participants

A group of 5 naïve participants from JLU Giessen participated in the experiment. All participants had self-reported normal or corrected-to-normal vision. The experiment followed the guidelines set forth by the Declaration of Helsinki. All participants provided written informed consent and were reimbursed at a rate of **€**8/hour.

Analysis

Heatmaps: Fixations were aggregated over all participants and plotted on the average luminance value background image. If we consider the center of the image (horizontally) to be zero degrees (see inset of Figure 48C), the X and Y fixation positions are plotted. Heat maps indicate fixation durations at location, with red regions indicating the longest durations. For the purpose of analyses, the 'reference stimuli' was taken to be the center stiffness and lightness values (see center of Figure 46) for graphic layout.

Psychometric Functions: For each task, data were averaged over all participants. We modeled the probability of the comparison to be reported as stiffer (task 1) or lighter (task 2) than the reference, as a function of the rendered stiffness/lightness values. Using the 'Psignifit 4' (Schütt, Harmeling, Macke, & Wichmann, 2016) MATLAB toolbox, we fit a psychometric function to the observers' responses. The slope of the psychometric function is a measure of the just noticeable difference (JND) for the stiffness and albedo differences at which participants perform stiffness and lightness discriminations. We computed a three-way repeated measures analysis of variance (ANOVA) on the fixation locations to investigate overall differences in vertical position between our three conditions (Fixations (1-3), Motion (Static/Dynamic), and Task (Stiffness/Lightness)).

Results

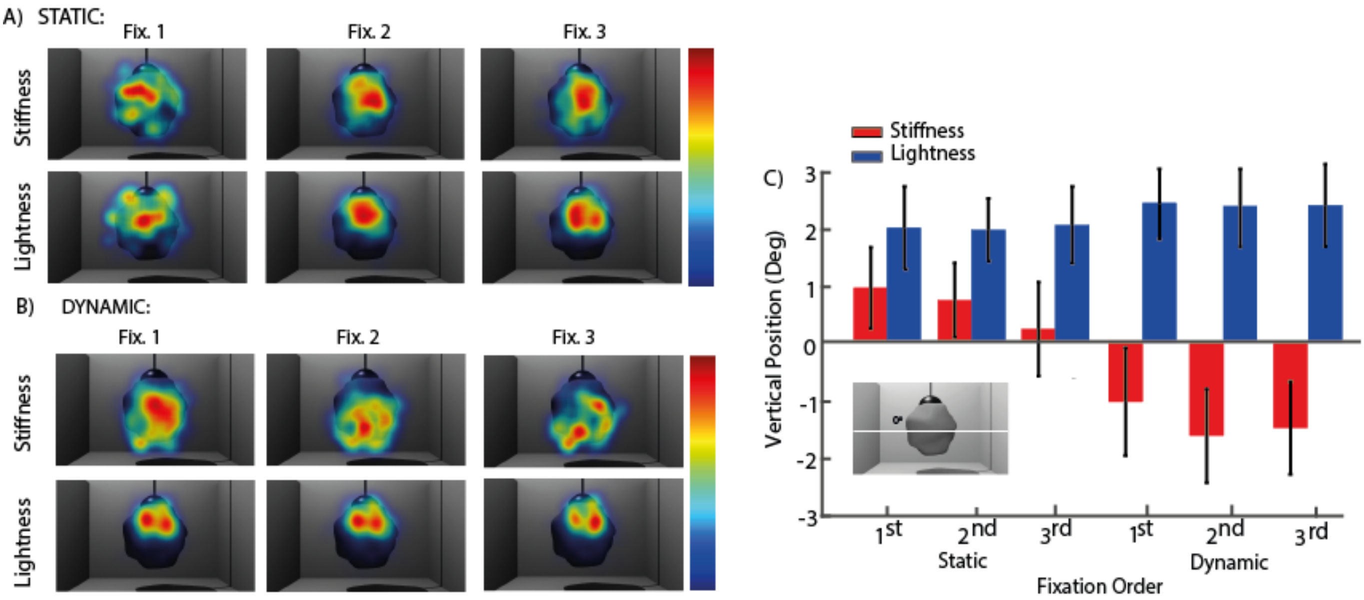


Figure 49 a-c. Pilot Blob Eye Tracking Results. Heatmaps of human fixation patterns (aggregate data over all participants and trials) for static portion of video. **49B.** Heatmaps of human fixation patterns (aggregate data over all participants and trials) for dynamic portion of video. Red areas indicate greatest fixation locations/durations. **49C.** Bar graph of fixation locations comparing static and dynamic timepoints. *Heatmap images contributed by M. Toscani.*

Figure 49a plots heatmaps for the average fixation location and duration over all participants (on the average luminance stimulus exemplar) during the Static phase (prior to object wobbling). 49b plots fixation location and duration averages during the Dynamic phase of the stimulus. Red areas indicate regions of longest fixation duration. We plot average fixation duration for the first three fixations on the average luminance exemplar. Results show that fixations are drawn to the most informative regions of the object for material property judgements in dynamic scenes. For both stiffness and lightness tasks, fixations begin in the same region on the object, and quickly separate as a function of task: For lightness judgements, fixations move upward toward the region of greatest lightness; while for judgements of stiffness, fixations move downward, toward the region of greatest motion.

Figure 49a demonstrates that when the object is not moving (irrespective of the task they are performing) participants are relatively agnostic about where to look, and therefore fixate the top of the object. Figure 49b shows that as soon as the object begins to move, for judgements of stiffness, fixations immediately move down toward the region of highest motion energy. Critically, when the participant is judging lightness and the object begins to wobble, fixations remain at or above the center of the object, as this is the region which is most informative for such a judgement. Figure 49c presents these findings in graphical format: Plotting the vertical position on the Y-axis against fixation-by-condition on the X-axis, we find that fixation location remains generally in the same location for both stiffness and lightness tasks during the static portion of the animation, and as soon as the object begins to move ('dynamic' segment), fixations clearly move in opposite directions as a function of task.

A three-way Analysis of Variance was conducted which investigated the influence of three independent variables (task, interval, and fixation number) on the vertical fixation positions on each stimulus. Task contained two levels (Stiffness and Lightness), Interval contained two levels (Static Segment and Dynamic Segment), and Fixation contained 3 levels (First, Second, or Third Fixation). A statistically-significant interaction between Task (Stiffness/Lightness) and Fixation Number (1-3) was found, $F(2,5099)= 5.73$, $p=.003$, suggesting that the first through third fixation locations vary as a result of the task being performed. A statistically-significant interaction between Task and Interval (Static/Dynamic): $F(1,5099)=6.03$, $p=.014$ was also uncovered, suggesting that fixation locations vary as a function of the task the participant was performing, and whether or not the object had begun to move. This experiment utilizes the method of constant stimuli and a Two-Interval Forced Choice task (2IFC) to determine the psychometric functions for stiffness and lightness discrimination (averaged over participants). Figure 50a-b plots psychometric functions as the percentage of correct responses against the physical stiffness/lightness parameters of the rendered stimuli.

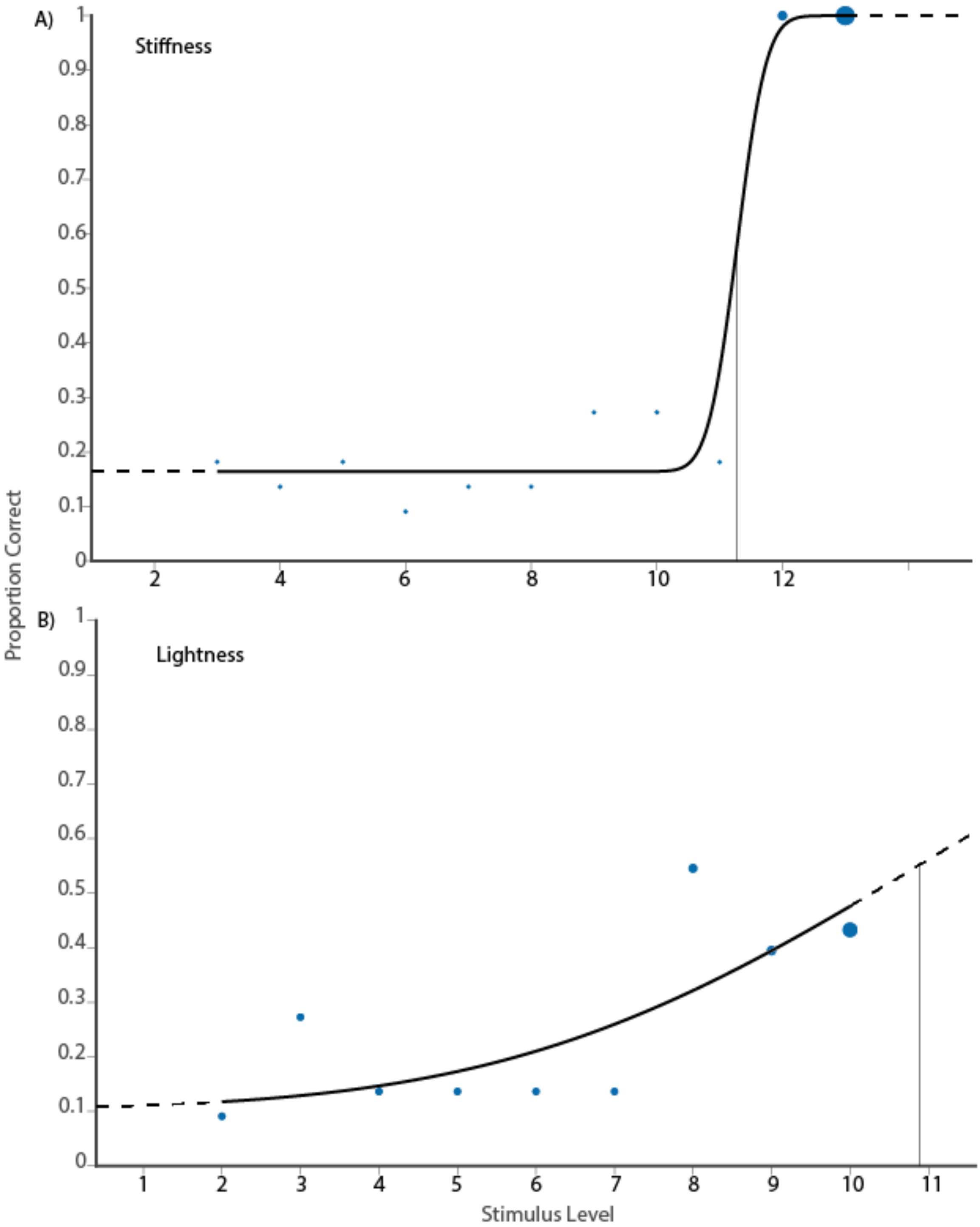


Figure 50 a-b. Psychometric Function Results for Stiffness and Lightness Judgements of Pilot Blob Stimuli. Figure 50a presents the psychometric function averaged over all observers for judgements of stiffness. The y-axis denotes the proportion of correct responses (stimulus comparison judged stiffer than the reference value); the x-axis denotes rendered differences in stiffness (stimulus intensities) between the comparison stimulus and the reference. Stiffness levels were visually selected to be perceptually distinct from one another during rendering. Dark blue dots indicate the measured probabilities, and the black solid line represents the fitted psychometric function for the stiffness condition. Figure 50b presents the psychometric function averaged over all observers for judgements of lightness. The y-axis denotes the proportion of correct responses (comparisons lighter than reference); the x-axis denotes albedo difference (stimulus intensities) between the comparison and the reference. Dark blue dots indicate the measured probabilities, and the black solid line indicates the fitted psychometric function for the lightness task.

Figure 50a-b depicts the psychometric functions for the stiffness and lightness tasks of the Pilot Blob stimuli, plotting proportion correct as a function of increasing stiffness/lightness levels. Aggregating data over all participants for the stiffness task, we find that as the levels of stiffness increased, the percentage of correct judgements did not substantially increase until the stimulus with the greatest amount of wobble, indicating that participants were not able to properly discriminate levels of stiffness relative to the comparison stimulus (center stimulus 3). We describe potential rationales for this finding in Discussion. For the lightness task (Figure 55b), we find that the percentage of correct responses increases in a semi-linear fashion with an increase in the degree of lightness (see Figure 51, center row for a visual comparison). Generally, as the lightness levels of the stimuli increased, the proportion of correct responses increased, indicating that observers were able to better discriminate lightness levels relative to the comparison stimulus (center stimulus 3).

Discussion

This experiment investigated how fixations on a dynamic, wobbling object differs in location as a function of the material property judgement task being performed. Using tasks of stiffness and lightness judgements, we investigated how fixations are drawn toward the most informative regions of a scene for the task at hand, and how additional information *irrelevant* to such task is essentially ignored during that time frame.

While our task is simple, our results provide an interesting basis for investigating how fixations are directed toward different regions of the visual scene to make different types of judgements, including material property judgements from surface and from kinematic properties. We show that fixations differ as a function of task, and additionally differ as a function of whether the object in a visual scene is static or in motion. Although this set of stimuli does not utilize familiar objects, it has many parallels to real-world scenes, which allow us to generalize our findings. As in nearly all real-world scenes, illumination comes from above. Objects that are held from cords or sticks (as in the present set of stimuli) and then bumped (consider a chandelier swaying back and forth after it is hit by a flying object) behave in these characteristic ways. Consider holding a water bottle and shaking it back and forth to observe its contents. The motion of the water in the bottle is the focus of attention, rather than the outer shape of the container. Fixations on a region of an object is always drawn to the judgement to be made, and indeed, this is demonstrated in our findings: Under limited resources, the visual system allocates fixations toward the most relevant information in a scene.

As participants freely fixate visual scenes to navigate the world, we allowed participants to freely fixate our scenes. Consistent with the experiments described by Toscani 2013a/2015, if we were to require participants to fixate a given point, we would expect to find a decrease in performance. Our results indicate for judgements of lightness, the brightest portions of an object in a scene are particularly relevant for lightness discrimination tasks, whether the object is static or in motion. For stiffness discrimination tasks, the dynamic regions of the object are the most relevant for stiffness discrimination, and are thus fixated to the exclusion of all other information. Further motion energy analyses will be necessary to investigate whether fixations are indeed directed to the region of greatest motion energy.

Effect of Expectation/Developing a prior. As in the previous experiments described in Chapter 2, it is possible that a prior develops over the course of the experiment, as participants 'learn' that the wobble will occur in the bottom left corner (where the object is consistently 'poked' from behind. Given that participants perform a large number of trials in this experiment, with multiple (repeated) presentations of exemplars, it is possible that participants 'learn' (have repeated practice with) the range of stimuli in our experiment (particularly those at the ends of the spectrum of stimuli), and as such, later trials reflect effects of learning. Additional analyses relative to the course of the experiment (as in the studies described in Chapter 2) will be necessary to investigate the effect of these priors over the course of the experiment.

Stimulus Limitations: While our experimental task was designed to elicit distinct stiffness and lightness judgments from our observers (c.f. Toscani et al., 2013a), psychometric functions indicate that participants were not able to properly discriminate differences in stiffness levels of our stimuli, leading to floor and ceiling effects. Looking at the stimuli, the 'wobble' only occurs the bottom left corner, as it is consistently poked in the same location from behind. The resulting wobble is small, short in duration (non-repeating), and difficult to detect at the highest stiffness levels. These confounds almost certainly influence our results. During the rendering of the stimuli, levels of stiffness were chosen to be perceptually distinct, and laid out on a linear scale. Such psychometric functions confirm that the levels for stiffness were not chosen effectively.

For these reasons, we develop an alternative set of stimuli in the following experiment, which attempts to address these concerns by eliminating such confounds, and allowing for a wider range of analyses.

Experiment 4.2: "Punching Bag"

This experiment sought to create a new set of stimuli that parametrically varied perceived stiffness and albedo, which would allow attention to be drawn to a particular region of the stimulus, depending on the task being performed. This stimulus included colored walls and floor for context, and the scene was illuminated from the upper left.

Methods

Stimuli

Stimuli consisted of 48 greyscale punching bags, rendered in Blender 2.77a (Stichting Blender Foundation, Amsterdam, NL). These blobs varied in their perceptual degrees of stiffness, and were illuminated under six different lighting conditions (albedos). For each stimulus, lighting remained constant on the blobs, and the blobs were illuminated from a single location (above the object from the left). These stimuli also included colored walls and a floor for additional context.

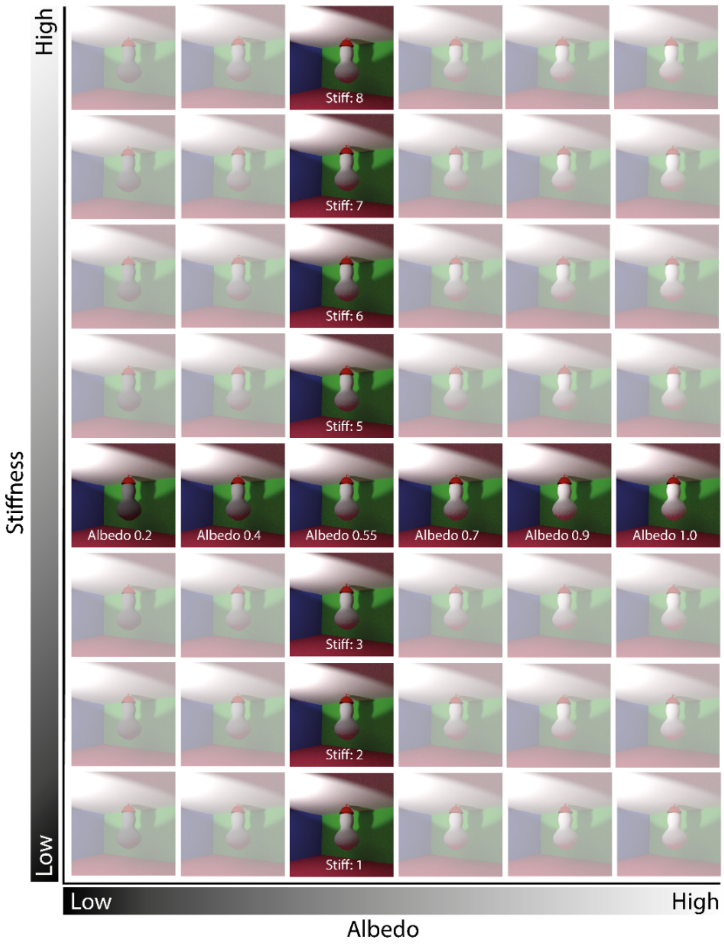


Figure 51. Stimuli for 'Punching Bag' Experiment. Here we display all combinations of stiffness (varied in eight levels) and albedo (varied in six levels) for this experiment. Observers saw all combinations of albedo and stiffness for the stimuli over the duration of the experiment. *3D geometry contributed by R. Ennis.*

Apparatus

The apparati used in this experiment were identical to those described in the previous experiment. We refer the reader to that section.

Task and Procedure

Participants were asked to perform a two-Interval-Forced-Choice (2IFC) experiment, containing 1000 trials in two blocks (500 trials per block). Participants always performed the stiffness block first, followed by the lightness block. Participants knew beforehand which attribute they would be asked to judge.

On each trial, the first frame of a video was presented for one second. Following this display, the video then played for the full two-second duration. The eye tracker would then drift check, and the comparison video would play for the full two-second duration. A screen would then show, asking the participant which was stiffer (in the first block) or lighter (second block), and the participant would respond using the left or right mouse button. The participant saw all possible pairs (combinations) of all stimuli. Figure 52 shows a sample trial.

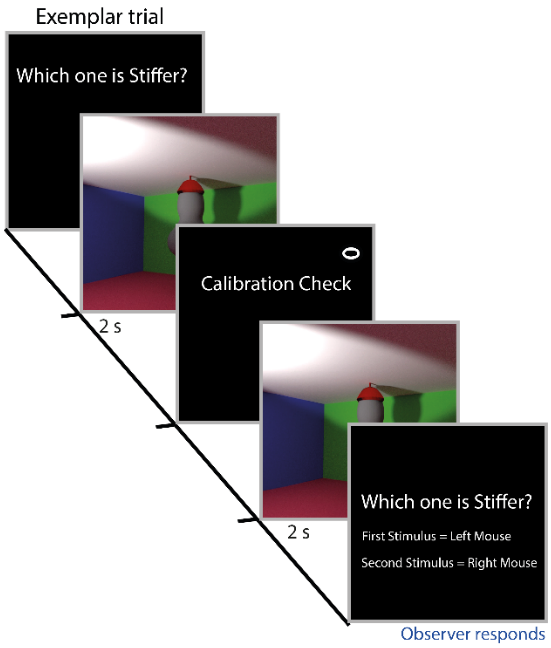


Figure 52. Trial Diagram for 'Punching Bag' Experiment. Example of a single trial. Each video would play once, followed by the comparison video. The participants responded with the left or right mouse button to indicate (in the firs block) in which video the object was stiffer, and (in the second block) which video was lighter.

Participants

A series of naïve participants from JLU Giessen participated in the experiment. All participants had self-reported normal or corrected-to-normal vision. The experiment followed the guidelines set forth by the Declaration of Helsinki. All participants provided written informed consent and were reimbursed at a rate of **€**8/hour.

Analysis

Heatmaps: As in the previous experiment, if we consider the center of the image (horizontally) to be zero degrees (see equivalent inset of Figure 49C), the X and Y fixation positions are plotted relative to their fixation durations. In figure 53b, trials which have fewer than three fixations were excluded from analysis, and data were averaged over fixations. Higher bars indicate a position lower on-screen. Error bars are 1 Standard Error of the Mean.

Psychometric Functions: As in the previous 'Pilot Blobs' experiment, for each task, data were averaged over all participants. We modeled the probability of the comparison to be reported as stiffer (task 1) or lighter (task 2) than the reference, as a function of the rendered stiffness/lightness values. The 'Psignifit 4' (Schütt, Harmeling, Macke, & Wichmann, 2016) MATLAB toolbox allowed us to fit a psychometric function to the observers' responses. The slope of the psychometric function is a measure of the 'just noticeable difference' (JND) for stiffness and lightness levels at which the participants perform stiffness/lightness discriminations.

Classifier: Our analysis used a linear classifier to investigate whether using only the X-Y positions of the fixations would allow us to differentiate the task being performed (stiffness versus lightness). This 'leave-one-out' linear classifier was trained on the X and Y fixation positions, of all trials of all participants, leaving one out at a time. In this way, the linear classifier is trained on the X – Y fixation positions a participant made for *all remaining* stimuli. The trained classification function was then used to predict the task being performed of the excluded trial. We computed average classifier accuracy (percentage of correctly classified trials (relative to all trials)). The linear classifier was trained using the 'classify' function from the MATLAB Statistics and Machine Learning toolbox (MathWorks, Inc.).

Results

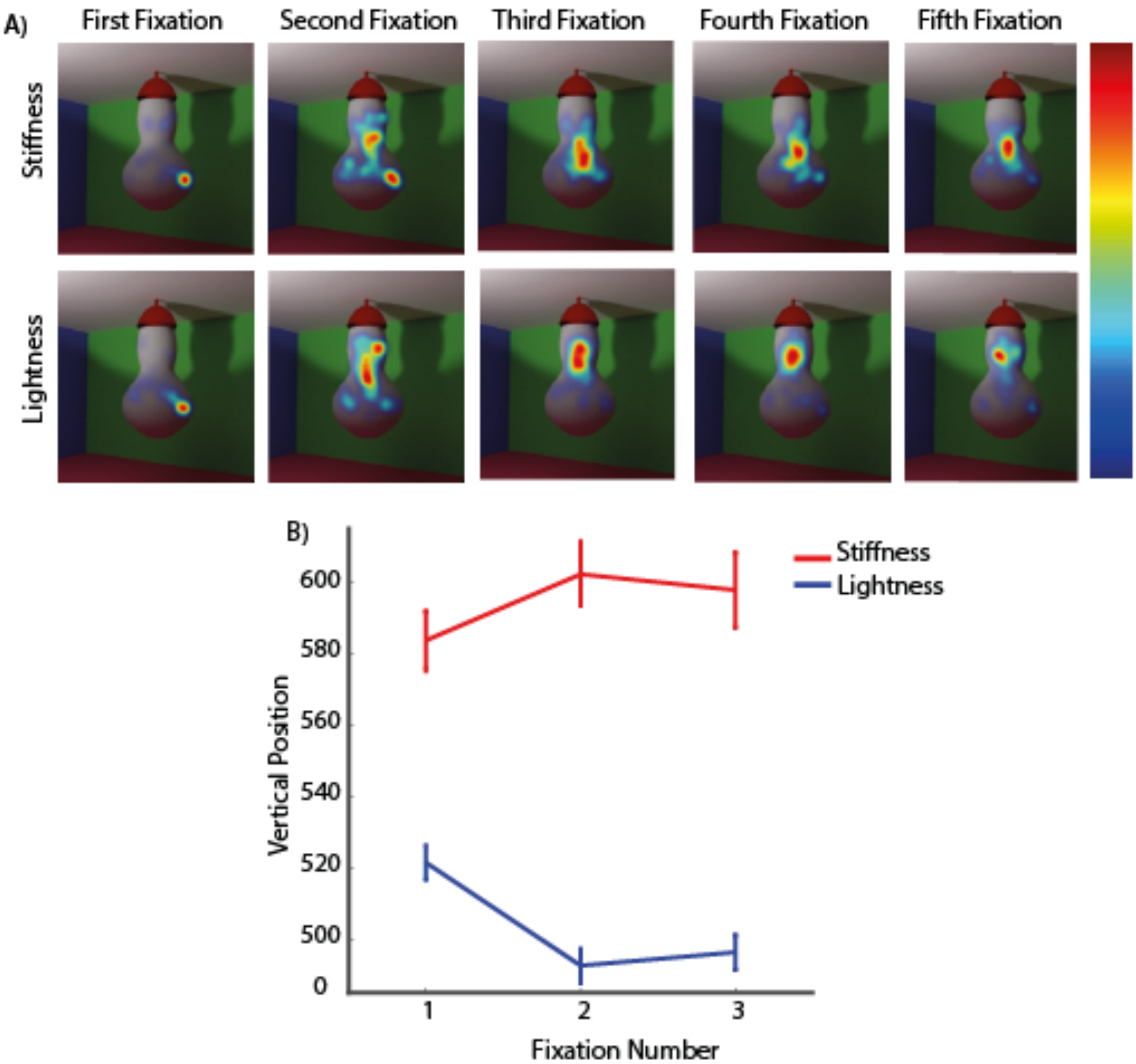


Figure 53 a-b. Results of 'Punching Bag' Experiment. 53a) Heatmaps of human fixation patterns (aggregate data over all participants and trials) for dynamic portion of video. Red areas indicate greatest fixation locations/durations. 53b) Plot of fixations on vertical axis as a function of task and fixation. Errorbars represent 1 Standard Error of the Mean. *Heatmap images contributed by M. Toscani.*

As in the previous experiment, results show that fixations are drawn to the most informative regions of the object for material property judgements in dynamic scenes. Figure 53a plots heatmaps for the average fixation location and duration over all participants (on the average luminance stimulus exemplar). Red areas indicate regions of longest fixation duration. We plot average fixation duration for the first five fixations. Consistent with the findings of the previous experiment, for both stiffness and lightness tasks, fixations begin in the same region on the object, and quickly separate as a function of task: For lightness judgements, fixations move upward toward the region of greatest lightness; while for judgements of stiffness, fixations move toward the region of greatest motion. Figure 53b plots the average vertical position averaged over all observers for the first through third fixations. In this figure, higher bars represent fixations lower on screen. The locations of the initial fixations are in distinct locations on the object, and separate further apart for the fixations that follow (upward for lightness, downward for stiffness), further providing support for the hypothesis that fixations differ as a function of the material judgement task that the participant has been asked to perform.

Given this finding of a distinct separation in fixations based on task, we asked if, by using fixations (equivalent to Inverse-Yarbus) a linear classifier is able to predict the task being performed by the participant.

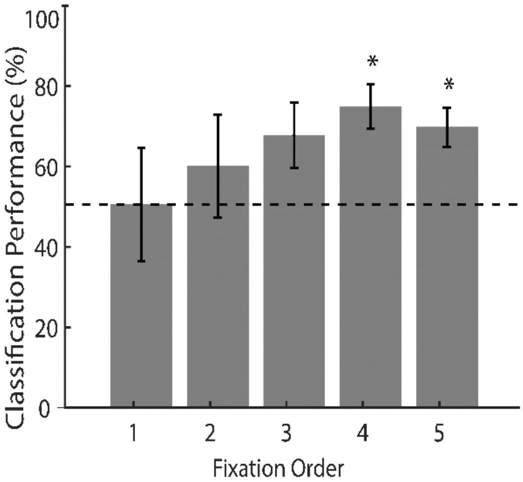


Figure 54. Classification Performance as a function of fixation order. Here we plot classification performance (as percentage) as a function of fixation order. Results suggest that a linear classifier is able to reliably distinguish, by the fourth fixation, the task being performed by the participant (stiffness versus lightness). Errorbars represent 1 Standard Error of the Mean. *Classification Analysis/Figure contributed by M. Toscani.*

Figure 54 depicts the classification performance from the first to the fifth fixation. By the third fixation, classification performance is above 50%. Our results demonstrate that a linear classifier is able to reliably predict which task is being performed by the participant by the fourth fixation. Correspondingly, the heat maps of Figure 53 show fixations graphically—by the third fixation, fixations have tended to cluster together and have settled on a particular location on the object. This suggests that (for our stimuli at a minimum), in contrast to the findings of Greene (2012), a linear classifier is able to use initial fixations to classify the task being performed by the participant.

Psychometric Functions

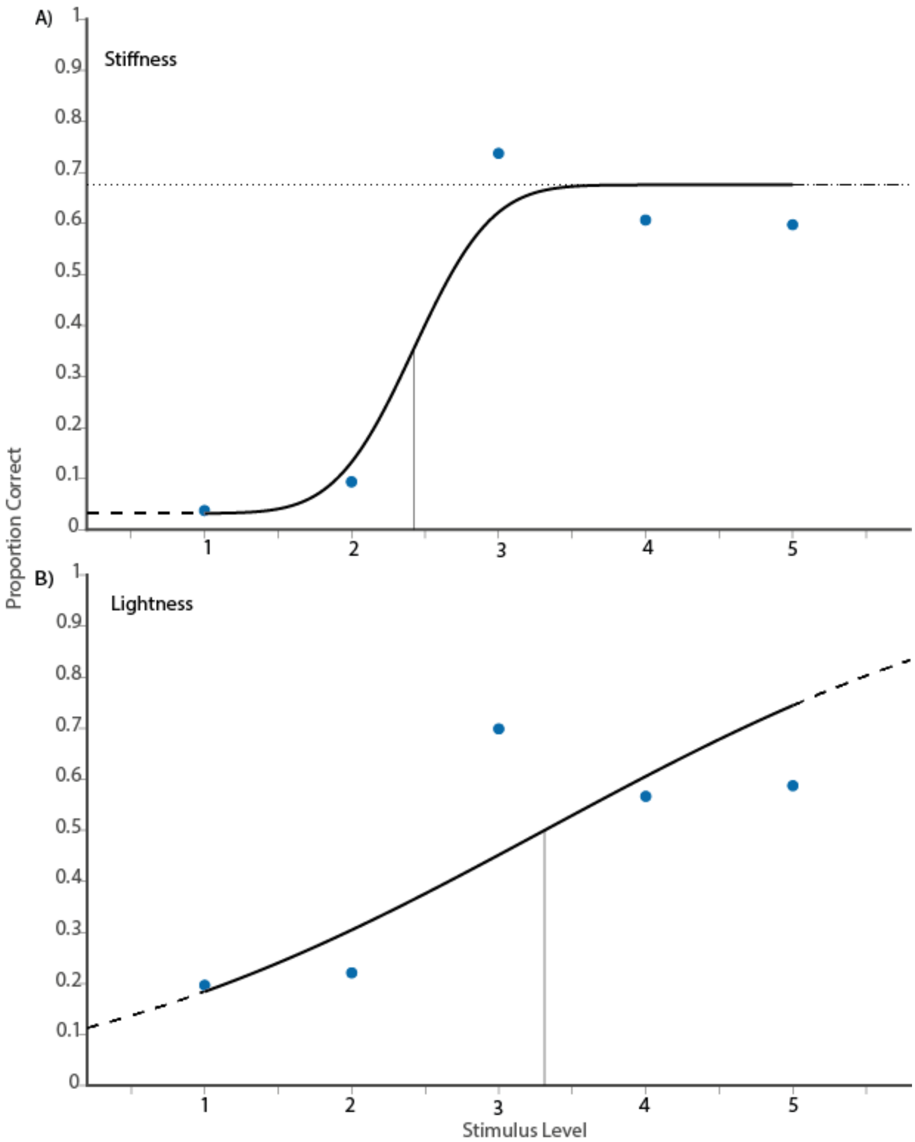


Figure 55 a-b. Psychometric Function Results for Stiffness and Lightness Judgements of Punching Bag Stimuli. Figure 55a presents the psychometric function averaged over all observers for judgements of stiffness. The y-axis denotes the probability of judging the comparisons as stiffer than the reference (center stiffness level, 3); the x-axis denotes perceptual differences in stiffness (stimulus intensities) between the comparison and the reference. Stiffness levels were visually selected to be perceptually distinct from one another. Black dots indicate the measured probabilities, and the black solid line represents the fitted psychometric function for the stiffness condition. Error bars represent the standard error of the mean. Figure 55b presents the psychometric function averaged over all observers for judgements of lightness. The y-axis denotes the probability of judging the comparisons as lighter than the reference (center lightness level, 3); the x-axis denotes albedo difference (stimulus intensities) between the comparison and the reference. Black dots indicate the measured probabilities, and the black solid line represents the fitted psychometric function for the lightness condition.

Figure 55a-b depicts the psychometric functions for the stiffness and lightness tasks of the Punching Bag stimuli, plotting proportion correct as a function of increasing stiffness levels. Aggregating data over all participants for the stiffness task, we find that as the levels of stiffness increased, the percentage of correct judgements increased, indicating that participants were able to properly discriminate levels of stiffness relative to the comparison stimulus (center stimulus 3). For the lightness task (Figure 55b), we find that the percentage of correct responses increases nearly linearly with an increase in the degree of lightness (see Figure 51, center row for a visual comparison). Generally, as the lightness levels of the stimuli increased, the proportion of correct responses increased, indicating that observers were able to properly discriminate lightness levels relative to the comparison stimulus (center stimulus 3).

Discussion

Consistent with the results of the previous experiment, this experiment investigated how fixation location differs as a function of the material property judgement task being performed. Using tasks of stiffness and lightness judgements, we investigated how fixations are drawn toward the most informative regions of a scene for the task at hand, and how additional information *irrelevant* to such task is essentially ignored during that time frame. The results of this experiment replicate the previous findings, and again suggest that participants are using an efficient sampling strategy to perform each of the tasks, sampling only that information which is relevant for the task at hand.

In contrast to the previous experiment, the psychometric functions resulting from this set of stimuli (for lightness and stiffness judgements) suggest that the kinematic rendering parameters were more appropriately chosen to be more clearly perceptually distinct between stiffness levels. Our findings over both experiments suggest that, even with the change in shape, participants are able to more clearly distinguish kinematic stiffness and luminance levels for the 'Punching Bag' set of stimuli. Given this finding of distinct fixation regions based on task and the updated shape of the stimuli, this stimulus set would lend itself well to investigations on the effect of object motion in the periphery (see Chapter 5 for more discussion on this topic). Also in contrast to the previous experiment, this experiment does not contain a static portion of the video for comparison— as soon as the object was presented, it began to wobble. This could also present a confound in our results, in that the visual system must focus on the object presented on screen, before directing fixations to a particular region of the stimulus. Even without this comparison interval, it is surprising that fixations so rapidly move in opposite directions as a function of the task to be performed (within the first 3 fixations).

The classification accuracy of the linear classifier results again alludes to the fact that participants are utilizing efficient strategies when making material property judgements about the dynamic objects in our scenes. Given that the fixations are significantly distinct between tasks, the linear classifier is able to accurately predict results by the fourth fixation. This is in line with the findings from Land and Hayhoe (2001), indicating that fixations can provide insight into the task being performed by the participant, as they are guided toward information relevant only for the task at hand. It remains to be seen to what extent these findings can generalize to other objects and material perception conditions. As the fixations for the two tasks are sufficiently spread apart in screen location (those at the top of the screen for lightness judgements, and the bottom of the object for stiffness judgements), this approach works well for this set of stimuli, as our stimuli were rendered matte. For stimuli which are smaller, moving more substantially, or have features (e.g. specular highlights which are closer together in distance, as in our falling 'Surprise' stimuli) more analyses will be necessary.

What would we expect to find if this experiment was conducted in color (rather than greyscale matte)? Hamel et. al. (2015) investigated this question, with particular attention to the influence of color on the exploration of videos. In a passive-viewing condition (task-free), they conducted an experiment to investigate the influence of color on eye movements in natural scenes. Participants watched color and greyscale versions of videos, while gaze was tracked. Eye positions considering mean amplitude of saccades and mean duration of fixations were analyzed and compared to the predictions of a saliency model. The authors concluded that color has a 'slight' influence on eye movements, given that the number of 'attractive' regions was greater for colored than for greyscale stimuli. Although this experiment was conducted using videos with more complex content (relative to) rendered objects, these findings are consistent with the findings of this and previous experiments discussed in this thesis. As most luminance and motion experiments have been conducted in the greyscale domain for simplicity of stimulus, more work will be necessary to investigate the interaction of color in guiding eye movements.

CHAPTER 5: General 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. 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 shown 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. 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. 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. 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, 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. 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. Here, we use an experimental paradigm analogous to cue conflict to create gross violations of expectations about materials. We show that the qualities of "surprising" materials are perceived different to expected ones that behave the same, and that surprise leads to increases in processing time of the stimuli. 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.

Knowledge affects (material) perception

There are many demonstrations showing that knowledge affects how we perceive the world: from detecting the Dalmatian amidst black and white blotches or an animal in the scene (Thorpe, Fize, & Marlot, 1996), deciding on the identity of a blob by Torralba and Oliva (2007), to being a greeble or bird expert (Gauthier, James, Curby, & Tarr, 2003). Object knowledge does not just facilitate categorical judgments, it also affects the estimation of visual properties such as color or motion (Hansen et al., 2006; Olkonnen and Allred, 2014; Scocchia et. al. 2013). How exactly knowledge alters and facilitates neural processes in visual perception is a topic of ongoing research (e.g. Rahman & Sommer, 2008; Gauthier, Skudlarski, Gore, & Anderson, 2000; or Kveraga et al., 2007).

Knowledge about materials entails several dimensions and can include taxonomic relations: gold is a metal, metals are elements with physical properties, metals are usually malleable and ductile, as well as perceptual regularities: gold looks yellowish, often has a very shiny, polished and smooth appearance, feels cool to the touch etc. Our experimental results suggest that identifying a material (i.e. knowing what it is) not only co-activates its typical optical qualities, but also elicits strong predictions about the typical mechanical properties and resulting material 'behaviors'. For example, liquids are not only translucent or transparent, they also tend to run down, splash, or ooze. Importantly, we seem to have quite specific ideas of what running down, splashing, or oozing should look like (e.g. Dovencioglu et al., 2018 probed such 'ideas' explicitly), supposedly because we have ample visual (but also haptic) experiences with liquids, and thus opportunities to learn the regularities (statistical or other) associated with a specific material category. We will follow up this thought in the sections below. The surprising outcome in our experiments is that these specific ideas, or 'priors', about 'material behaviors appear to interfere with the bottom-up processing of visual information, leading to differences in ratings of material properties between Expected and Surprising conditions.

While we have probed existing priors (built over a lifetime) on the mechanical properties of familiar objects, recent work also suggests that observers can fairly quickly learn about these properties when watching materials 'behave' (Bates et al., 2018) and use this knowledge to predict material kinematics under novel circumstances.

Shape and optics factors. The shape manipulation in our experiment is different from that of the experiments by Paulun et al. (2015 & 2017), Schmid et al. (2017), or van Assen et al. (2016 & 2018): They systematically varied one parameter (e.g. viscosity or softness) in their simulations of soft materials and then derived shape statistics that best captured the changes in shape across these variations and that predict observers' judgements, while we change shape categorically in the Familiar object condition. Returning to the idea of direct and indirect routes in material perception (Paulun et al. 2017), one might be tempted – at first sight - to classify their findings as belonging more to the direct (visual estimation) and ours to the indirect (association) route. However, it is probably more likely that the visual system strongly relies on previously formed associations between shape and mechanical material category in all of these experiments: observers probably have had as much experience with how liquids of various viscosities deform as they have had with shattering plates and wobbling jellies. Thus, there is nothing inherent in the shape that signals softness or viscosity per se, but instead it is (multisensory) experience with materials that enables us to judge mechanical (and potentially even some optical) properties like wobbliness or stiffness (etc.). Thus, any kind of material judgement relies on preceded perceptual learning. This may sound like a truism; however, it could explain previous results by Schmidt et al (2017), who found only weak or no correlations between perceived shape deformations and softness. In their study, they used novel objects for which the postulated shape-material associations (prior perceptual learning) was (by definition) entirely absent. Alternatively, these shapes might have 'reminded' observers of objects that are typically non-soft. Whether the former or the latter applies, if association is the key, then sufficient perceptual training with this specific stimulus set should yield the expected deformation- softness correlation in Schmidt et al. (2017).

One more prediction yielded by the idea of learning-necessary-for-material perception is that the effect of expectation index $\boldsymbol{\epsilon}$ and the reaction time difference τ_D in our experiment should decrease over the course of the experiment, since observers saw the same Surprising stimuli in each of the 4 blocks (1 block per rating question). As observers progress through the experiment, both ϵ and τ_D decrease, though this trend is somewhat clearer for τ_D. Interestingly, in both Schmidt et al. (2017) and Paulun et al (2017), optical properties predicted perceived stiffness (or softness) quite strongly. They argue that the unfamiliarity of the shape, the invisibility of the transformation effector (i.e. object or force that impacted the stimulus) and the fact that it was a static scene all rendered the shape cue less reliable than the optical cues (Schmidt et al., 2017) when estimating softness. Conversely, the fact that the changes in optical material properties were categorical (e.g. wood, marble, latex, velvet, wax), with each category strongly associated with a specific material and a corresponding degree of stiffness, could explain their results.

Consistent with the above discussed literature, we found that not only familiar shape of an object elicited strong predictions about the material mechanics, but also the optical properties alone yielded expectations about how an object should behave when dropped, as evident by the non-zero effect of expectation indices in the Novel object condition. We can also see this effect present in the first frame experiment: for example, the velvet Novel object is rated *harder* than the one donning the key material.

A Bayesian account

We believe that the results of this study fit well within the Bayesian framework, which offers an account of how prior knowledge is integrated with sensory input. Our experiment constitutes a situation not unlike classical cue-conflict experiments (e.g. Ernst & Banks, 2002, Knill 2007), where the sensory cues may be in conflict with one another and/or the prior belief. While we do not aim to model our results formally, we still believe that this analogy is useful in interpreting our findings. We will first focus on the rating differences that we found in the Expected and Surprising conditions. Here, objects that physically deformed, i.e. Wrinkled, splashed, etc, in the same way were not rated the same way. In fact, in many cases, ratings were 'pulled' towards the expected material property, not the signaled one. For example, a wrinkling spoon or teacup was rated *harder* than any of the wrinkling cloths. Here, prior knowledge about spoons and teacups being *hard* seems to have led to increased ratings of *hardness* compared to their soft curtain counterparts, despite all of these objects wrinkling. This outcome would be best explained by a so-called down-weighing of the cues (Knill, 2007) to *hardness*, which suggests that the visual system entertains multiple priors (strong and weaker ones) about the state of the world, and that, depending on the sensory input, it adjusts the weights of these priors. This implies that in a cue-conflict situation, the unlikely interpretation of the input does not simply get vetoed down (Landy et al. 1995), but that instead it would factor into the final percept – just as we observed it in our results.

When faced with violated expectations, as in our Surprising condition, we suggested that the visual system needs to update the generative model in order to minimize the prediction error i.e. the error between the expected state and 'measured' state of the world, for future tasks. Because this updating is a reiterative process, we reasoned that it would take observers longer to perform perceptual tasks when judging material attributes of surprisingly deforming objects. This is exactly what we found. Interestingly, Itti and Baldi (2009) defined surprise as the difference between the posterior and prior beliefs about the world, thus one could argue that the bigger the surprise in a condition is, the longer the reiterative error correction might take.

One might criticize that reaction times for ratings were much longer than times measured in classic reaction times studies (e.g see a review by Eckstein (2011) on visual search). However, it is not all that uncommon to consider reaction times of 2 seconds and longer, as in categorical color perception (Boynton & Olson, 1990, or Okazawa, et al., 2011). In defense of the reaction time effects, we would also like to point out that we treated reaction time data as conservatively as classic reaction time studies, e.g. by removing data points that were two standard deviations above the mean. Nevertheless, explanations for rating and reaction time differences other than the Bayesian account may be possible and we will consider those next.

Alternative explanations for rating differences

High level factors. It is quite striking that the same material deformations (material mechanics) were rated differently in the Expected and Surprising conditions. Assuming that low level differences do not explain the rating differences (see below), could the results be explained in other ways than by a Bayesian framework? Toscani et al. (2013) showed that depending on the task, observers pointed their gaze at specific points at the stimulus, e.g. near the brightest regions on an object for lightness judgements. One possibility could thus be that the task, object knowledge, and expectations about the material 'behavior' guided eye movements of observers also in our experiment to specific locations on the stimulus. While in Expected conditions fixation patterns might have been effective with respect to the task, e.g. observers correctly anticipated how the object would deform (or shatter, splash, Wobble etc.), it is possible that in the Surprising conditions the 'wrong' expectation guided eye movements to the 'wrong' locations on the object, which in turn lead to a different sampling of information and ultimately influenced their judgements of material qualities. We are currently investigating this possibility directly.

Low level factors. We equated the material mechanics across Expected and Surprising conditions, however, we are still comparing rather different stimuli: they not only differ in object identity (see Table 1) but also in several low-level image properties – due to the various shapes, sizes and optical properties of objects. While Novel objects were rendered with identical optical properties as the Familiar object, their overall appearance e.g. their shape and size was more homogenous than that of the Familiar objects. Perhaps, for example, a wobbling key is rated as less wobbly because there is simply less 'substance' to Wobble compared to the custard. However, this logic does not always apply: for example, the teacup is about the same (retinal) size as any of the curtains, yet, it is rated as '*harder*' in the Surprising condition. Similarly, the red jelly is rated as substantially wobblier than any of the Expected non-deforming objects (spoon, key, chair), and it is not clear how exactly (low-level) shape differences could explain this. Instead it is more likely that in the latter case, the optical properties and the shape of the object contributed to this differential rating of objects in Expected and Surprising conditions. Thus, we are fairly confident in ruling out that the different levels of shape and size heterogeneity in Familiar and Novel object conditions confound the rating differences between Expected and Surprising trials in the two conditions. However, to be certain, one should in future experiments, create novel objects with matched variability in shapes across Expected and Surprising conditions.

We suggest that the increase in reaction time for ratings of Familiar objects in the Surprising condition might be best explained by the fact that these stimuli require slightly longer processing times. However, alternative explanations are possible. For example, it could be that observers, upon recognizing the object, immediately set the rating slider to a specific position, and then – when confronted with the Surprising outcome – have to correct their initial estimate. The fact that more data points in the Familiar object condition were excluded as being too short in reaction time might support this idea. However, such a strategy would not only have caused observers to be overall slower on Surprising Familiar object trials, but also to be faster on Expected Familiar object trials (compared to Novel objects). That, however, was not the case. All three conditions (Expected Familiar object, Expected and Surprising Novel objects) had similar mean RT estimates; only the Surprising Familiar RTs were longer. It is also not clear why correcting a premature slider positioning would take longer than adjusting the slider position from a random position (with respect to the current stimulus).

Considerations and Limitations of this approach

We do not consider the effect of rating attribute in our data analysis, i.e. whether we are asking the participants to rate *heaviness*, *hardness* etc. However, there might be interactions between the question that is asked and the visual stimulus. For example, asking how *liquid* a shattering wineglass is may not be as difficult as asking the same question for a wobbling wine glass. Sometimes the judged property may simply not be relevant/ apply to a stimulus, or may have nothing to do with the expected material mechanics. It is plausible that if we repeated this experiment asking observers how glossy an object looked that we would not find any effects. The relevance of the rating attribute could be the subject of future experiments.

Fixations on Kinematic Objects

We conducted two experiments to investigate the effect of task on fixation location when making material property judgements. The findings of our eye tracking experiments suggest that not only are participants likely using an efficient sampling strategy for the task at hand, but that this sampling strategy is largely the same across observers. The setup of this experiment allows us to probe a number of factors that generalize to where on an object people fixate when making judgements about surface properties of objects, or make inferences about properties of an object (e.g. mass) when passively viewing objects which interact with other objects. Our studies are consistent with past literature that suggests that eye movements are goal-directed for the task at hand. This is akin to making perceptual judgments about real-world items like visual and haptic impressions doneness of steak (bouncing back when indented), or the squishiness of an avocado. A passive observer can gain a significant amount of information just by observing the degree of deformation of the object. Here, we find that observers indeed optimize their fixations to perform material perception tasks such as these. As these stimuli were all rendered with matte optical properties, additional analyses will be necessary to investigate if and how our findings generalize to optical/surface property judgements.

Observer Consistency in Dynamic Scene Fixations/Central Fixation Bias. Previous research (Smith and Mital, 2013) has found that the way participants fixate static scenes is distinct from the way dynamic scenes are fixated. Because static scenes are unaffected by temporal demands, participants who are asked to perform a given task can choose to fixate the scene in a number of different ways (while still achieving the same goal). This makes applying a linear classifier substantially more difficult (Greene, et. al. 2012). In contrast, research has found that the temporal demands that a dynamic scene presents places limitations on the observer, and observers have been found to be remarkable consistent in their fixations. This property makes studying fixations in dynamic scenes particularly interesting, and lends itself well to a linear classifier approach. In tandem with this finding, Dorr et. al., (2010) found similar eye movements across observers for dynamic scenes, and that repeated presentations of the same stimulus yielded more coherent eye movements than independent stimulus presentations across observers. It is notable that even with a tendency toward a central fixation bias we do not find one.

While these analyses are a first step into investigating where participants look when making material property judgements involving different tasks, there are a number of analyses that we have yet to consider, and could be the topic of future research.

Pupil Dilation: In future analyses, one area that remains to be investigated in our study is that of using pupil dilation to clarify our findings. Pupil dilation has long been used as a measure of surprise (c.f. Lavin et. al. (2014)). Consistent with existing literature, if observers watched our stimuli wobble (or fall) unexpectedly, the pupil should dilate in response to the unexpected motion/outcome. In the case of our eye tracking stimuli, such a response would indicate that our wobbling behavior was not expected for the given shape or material. Given that our eye tracking stimuli consist of unfamiliar shapes, this is an interesting avenue for investigation.

'Light-From-Above' Prior: Further work will be required to more carefully investigate the effect of the 'light from above' prior in our scenes. In the experiments we describe, the scene is illuminated from the top left. If the scene were to be illuminated from an alternate location (generally violating our existing 'light from above' prior), we would expect, consistent with our current results, that fixations move toward regions of greatest luminance for lightness judgements.

Effect of vision in the periphery. As this stimuli did not move a significant amount and always occurred in the same location (and as such, did not require significant saccades to an alternate location on-screen), it is also surprising that we see such a distinction based on task. They are required to look in much the same place, and yet, fixations still differ as a function of task. One aspect of the stimuli that has yet to be investigated is the effect of vison in the periphery. One way to approach this question is by an online blocking out of that which is centrally fixated and asking participants to perform the same task. If participants are able to sufficiently perform the task in the absence of central fixation, vision in the periphery must play a role.

Motion Energy Analyses: To investigate whether participants are actually guiding their fixations to the region of greatest motion, motion energy analyses will be necessary. Without such analyses, we can only speculate as to why people fixate on the regions of greatest motion.

CONCLUSION

This work shows that kinematic properties of materials can be activated by the familiar shape of an object, but also by the optical qualities of a surface. Our results imply that perceived material qualities are not only determined by the retinal stimulation, but instead can also be susceptible to cognitive influences, such as expectations and memory. When making judgements regarding the material of objects, participants optimize their fixations for the task at hand, and ultimately are successful at navigating the world based on these judgments.

References

Abdel Rahman R, Sommer W. (2008). Seeing what we know and understand: how knowledge shapes perception. Psychon Bull Rev.(6):1055-63. doi: 10.3758/PBR.15.6.1055. PubMed PMID: 19001567.

Adelson, E. H. (2001). Human Vision and Electronic Imaging VI. In *Proceedings of the. SPIE* (Vol. 4299). Retrieved from http://persci.mit.edu/pub_pdfs/adelson_spie_01.pdf

Aliaga, C., O'Sullivan, C., Gutierrez, D., & Tamstorf, R. (2015). Sackcloth or silk? *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception - SAP '15*, 41–46. https://doi.org/10.1145/2804408.2804412

Alley, L. M., Schmid, A. C*., & Doerschner, K. (2019). Expectations affect the perception of material properties. BiorXiv. doi: 10.1101/744458 *Denotes shared first authorship

Alley, L. M., Toscani, M., Ennis, R. J., & Doerschner, K. (2019). Initial fixations differ for brightness and stiffness judgements. Journal of Vision, 19(10), 148a. https://doi.org/10.1167/19.10.148a

Alley, L., Schmid, A., & Doerschner, K. (2017). Neo's Spoon and Newton's Apples: Prediction of rigid and non-rigid deformations of materials. Journal of Vision, 17(10), 224. https://doi.org/10.1167/17.10.224

Alley, L., Rathakrishnan, V., Harman, C., Kourtev, H., Kugel, A., Haladjian, H., Pylyshyn, Z. (2011). Tracking objects and tracking our eyes during disrupted viewing. *Journal of Vision*, *11*(11), 277–277. doi: 10.1167/11.11.277

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, *21*(24), R978–R983. https://doi.org/10.1016/J.CUB.2011.11.022

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Ballard, D., Hayhoe, M., Pelz J.(1995), Memory representations in natural tasks Journal of Cognitive Neuroscience, 7, pp. 66-80

Bar M. Visual objects in context. (2004). Nat Rev Neurosci.;5(8):617-29. Review. PubMed PMID: 15263892.

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. W. (2015). Humans predict liquid dynamics using probabilistic simulation. *Cognitive Science Society*, 172–177.

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. (2019). Modeling human intuitions about liquid flow with particle-based simulation. *PLOS Computational Biology*, *15*(7), e1007210. https://doi.org/10.1371/journal.pcbi.1007210

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332. https://doi.org/10.1073/pnas.1306572110

Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Perception & Psychophysics*, *30*(4), 407–410. https://doi.org/10.3758/BF03206160

Bergmann Tiest, W. M., & Kappers, A. M. L. (2007). Haptic and visual perception of roughness. *Acta Psychologica*, *124*(2), 177–189. https://doi.org/10.1016/j.actpsy.2006.03.002

Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2005). Remote effects of highlights on gloss perception. *Perception*, *34*(5), 565–575. https://doi.org/10.1068/p5401

Bi, W., Jin, P., Nienborg, H., & Xiao, B. (2018). Estimating mechanical properties of cloth from videos using dense motion trajectories: Human psychophysics and machine learning. *Journal of Vision*, *18*(5), 12. https://doi.org/10.1167/18.5.12

Bian, Z., & Andersen, G. J. (2013). Aging and the perception of egocentric distance. Psychology and Aging, 28(3), 813–825. https://doi.org/10.1037/a0030991

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

Birmingham E, Bischof WF, Kingstone A. (2009). Saliency does not account for fixations to eyes within social scenes. Vision Res.;49(24):2992-3000. doi: 10.1016/j.visres.2009.09.014. Epub 2009 Sep 24. PubMed PMID: 19782100.

Birmingham E, Bischof WF, Kingstone A. (2008). Social attention and real-world scenes: the roles of action, competition and social content. Q J Exp Psychol (Hove). ;61(7):986-98. PubMed PMID: 18938281.

Blake A, Bülthoff HH, Sheinberg D. (1993). Shape from texture: ideal observers andhuman psychophysics. Vision Res. 1993 Aug;33(12):1723-37. PubMed PMID: 8236859.

Blake A, Bülthoff H. Shape from specularities: computation and psychophysics. (1991). Philos Trans R Soc Lond B Biol Sci. Feb 28;331(1260):237-52. Erratum in: Philos Trans R Soc Lond Biol 1992 Dec 29;338(1286):417. PubMed PMID: 1674154.

Blake A, Bülthoff H. (1990). Does the brain know the physics of specular reflection? Nature. Jan 11;343(6254):165-8. PubMed PMID: 2296307.

Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. Vision Research, 30(9), 1311–1317. https://doi.org/10.1016/0042-6989(90)90005-6

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897X00357

Carrasco M. Visual attention: the past 25 years. (2011). Vision Res. 1;51(13):1484-525. doi: 10.1016/j.visres.2011.04.012. Epub. Review. PubMed PMID: 21549742; PubMed Central PMCID: PMC3390154.

Cholewiak, S. A., Kim, S.-H., Ringstad, P., Wilder, J., & Singh, M. (2009). Weebles may wobble, but conical frustums fall down: Investigating perceived 3-D object stability.

Community, B. O. (2018). Blender - a 3D modelling and rendering package. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

Davidoff, J. B., & Ostergaard, A. L. (1988). The role of colour in categorical judgements. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 40(3-A), 533–544. https://doi.org/10.1080/02724988843000069

Debevec, P. (1998). Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography (SIGGRAPH '98). ACM, 189--198.

De Loof E, Van Opstal F, Verguts T. (2016). Predictive information speeds up visual awareness in an individuation task by modulating threshold setting, not processing efficiency. Vision Res.;121:104-112. doi: 10.1016/j.visres.2016.03.002. Epub. PubMed PMID: 26975499.

Doerschner, K., Kersten, D., & Schrater, P. R. (2011). Rapid classification of specular and diffuse reflection from image velocities. *Pattern Recognition*, *44*(9), 1874–1884. https://doi.org/10.1016/j.patcog.2010.09.007

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology*, *21*(23), 2010–2016. https://doi.org/10.1016/j.cub.2011.10.036

Dorr M, Martinetz T, Gegenfurtner KR, Barth E. (2010). Variability of eye movements when viewing dynamic natural scenes. J Vis.;10(10):28. doi: 10.1167/10.10.28. PubMed PMID: 20884493.

Dövencioglu DN, van Doorn A, Koenderink J, Doerschner K. (2018). Seeing through transparent layers. J Vis. 2018;18(9):25. doi: 10.1167/18.9.25. PubMed PMID: 30267077.

Draschkow, D., & Võ, M. L.-H. (2016). Of “what” and “where” in a natural search task: Active object handling supports object location memory beyond the object’s identity. Attention, Perception, & Psychophysics, 78(6), 1574–1584. https://doi.org/10.3758/s13414-016-1111-x

Draschkow, D., & Võ, M. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-16739-x

Eckstein, M. P. (2011). Visual search: A retrospective. Journal of Vision, 11(5), Article 14. https://doi.org/10.1167/11.5.14

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Federmeier KD, Kutas M. (1999). Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. Brain Res Cogn Brain Res. Oct 25;8(3):373-92. PubMed PMID: 10556614.

Findlay JM. (1981). Spatial and temporal factors in the predictive generation of saccadic eye movements. Vision Res.;21(3):347-54. PubMed PMID: 7269312.

Fleming RW, Dror RO, Adelson EH. (2003). Real-world illumination and the perception of surface reflectance properties. J Vis.;3(5):347-68. PubMed PMID:12875632

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*(9), 798–820. https://doi.org/10.1167/4.9.10

Fleming RW, Wiebel C, Gegenfurtner K. (2013). Perceptual qualities and material classes. J Vis. 2013 Jul 11;13(8). pii: 9. doi: 10.1167/13.8.9. PubMed PMID:23847302.

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, Vol. 94, pp. 62–75. https://doi.org/10.1016/j.visres.2013.11.004

Fleming, R. W., Nishida, S., & Gegenfurtner, K. R. (2015). Perception of material properties. *Vision Research*, *115*, 157–162. https://doi.org/10.1016/J.VISRES.2015.08.006

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In *The Origins of Object Knowledge* (pp. 135–164). https://doi.org/10.1093/acprof:oso/9780199216895.003.0006

Foulsham T, Underwood G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J Vis. Feb 20;8(2):6.1-17. doi: 10.1167/8.2.6. PubMed PMID: 18318632.

Gallivan, J. P., Adam McLean, D., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural mechanisms of human tool use. *ELife*, *2013*(2), 1–29. https://doi.org/10.7554/eLife.00425

Gao, T., & Scholl, B. J. (2010). Are objects required for object-files? Roles of segmentation and spatiotemporal continuity in computing object persistence. *Visual Cognition*, *18*(1), 82–109. https://doi.org/10.1080/13506280802614966

Gauthier I, James TW, Curby KM, Tarr MJ. (2003). The influence of conceptual knowledge on visual discrimination. Cogn Neuropsychol. May 1;20(3):507-23. doi: 10.1080/02643290244000275. PubMed PMID: 20957582.

Gauthier I, Skudlarski P, Gore JC, Anderson AW. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci. ;3(2):191-7. PubMed PMID: 10649576.

Gibson EJ, Owsley CJ, Walker A, Megaw-Nyce J. Development of the perception of invariants: substance and shape.(1979). Perception.;8(6):609-19. PubMed PMI 530802.

Greene MR, Liu T, Wolfe JM. (2012). Reconsidering Yarbus: a failure to predict observers' task from eye movement patterns. Vision Res. Jun 1;62:1-8. doi:10.1016/j.visres.2012.03.019. Epub 2012 Apr 2. PubMed PMID: 22487718; PubMed Central PMCID: PMC3526937.

Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations. Journal of the Optical Society of America A, 4(3), 503. https://doi.org/10.1364/josaa.4.000503

Haji-Abolhassani A, Clark JJ. An inverse Yarbus process: predicting observers' task from eye movement patterns. Vision Res. 2014 Oct;103:127-42. doi: 10.1016/j.visres.2014.08.014. Epub 2014 Aug 28. PubMed PMID: 25175112.

Hamel, S., Houzet, D., Pellerin, D., & Guyader, N. (2015). does color influence eye movements while exploring videos.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. https://doi.org/10.1038/nn1794

Hartung B, Schrater PR, Bülthoff HH, Kersten D, Franz VH. (2005). Is prior knowledge of object geometry used in visually guided reaching? J Vis. 10;5(6):504-14. PubMed PMID: 16097863.

Hillis, J. H., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630. https://doi.org/10.1126/science.1075396

Ho YX, Landy MS, Maloney LT. How direction of illumination affects visually perceived surface roughness. (2006). J Vis;6(5):634-48. PubMed PMID: 16881794; PubMed Central PMCID: PMC2761220.

Hurlbert, A.C., Cumming, B.G., & Parker, A. (1991). RECOGNITION AND PERCEPTUAL USE OF SPECULAR REFLECTIONS.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Itti, L., & Baldi, P. (2000). *Bayesian surprise attracts human attention Laurent*. *72*(2), 181–204. https://doi.org/10.1038/nature13314.

Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention (2000). Vision Res.;40(10-12):1489-506. PubMed PMID: 10788654.

Itti et al., 1998. Itti, L., Koch C., Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 20 (11), pp. 1254-1259

Jain, A., & Zaidi, Q. (2011). Discerning nonrigid 3D shapes from motion cues. Proceedings of the National Academy of Sciences, 108(4), 1663–1668. https://doi.org/10.1073/pnas.1016211108

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. Cognitive Psychology, 24(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-o

Kaiser, D., & Cichy, R. M. (2018). Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *Journal of Neurophysiology*, *120*(2), 848–853. https://doi.org/10.1152/jn.00229.2018

Kamprani, Katerina. "The Uncomfortable." Theuncomfortable.com, www.theuncomfortable.com/.

Kawabe, T., & Nishida, S. (2016). Seeing jelly. Proceedings of the ACM Symposium on Applied Perception - SAP '16. Presented at the the ACM Symposium. https://doi.org/10.1145/2931002.2931008

Kawabe, T., Maruya, K., & Nishida, S. (2015). Perceptual transparency from image deformation. *Proceedings of the National Academy of Sciences*, *112*(33), E4620–E4627. https://doi.org/10.1073/pnas.1500913112

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision*, *11*(9), 1–19. https://doi.org/10.1167/11.9.4

Kim, J., Tan, K., & Chowdhury, N. S. (2016). Image statistics and the fine lines of material perception. *I-Perception*, *7*(4), 1–11. https://doi.org/10.1177/2041669516658047

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, *7*(8), 13. https://doi.org/10.1167/7.8.13

Koenderink, J. J., & Van Doorn, A. J. (1982). The shape of smooth objects and the way contours end. *Perception*, *11*(2), 129–137. https://doi.org/10.1068/p110129

Komatsu, H., & Goda, N. (2018, November 10). Neural Mechanisms of Material Perception: Quest on Shitsukan. *Neuroscience*, Vol. 392, pp. 329–347. https://doi.org/10.1016/j.neuroscience.2018.09.001

Kubricht, J., Jiang, C., Zhu, Y., Zhu, S.-C., Terzopoulos, D., & Lu, H. (2016). Probabilistic Simulation Predicts Human Performance on Viscous Fluid-Pouring Problem. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 1805–1810. Retrieved from https://www.seas.upenn.edu/~cffjiang/research/cogsci16/nips16.pdf

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, Vol. 65, pp. 145–168. https://doi.org/10.1016/j.bandc.2007.06.007

Land MF, Hayhoe M. (2001). In what ways do eye movements contribute to everyday activities? Vision Res.;41(25-26):3559-65. Review. PubMed PMID: 11718795.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7892735

Lavín, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. Frontiers in Behavioral Neuroscience, 7. https://doi.org/10.3389/fnbeh.2013.00218

Long, B., Konkle, T., Cohen, M.A., & Alvarez, G.A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. Journal of Experimental Psychology: General.

Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: Achievements and challenges. *Journal of Vision*, *10*(9), 19–19. https://doi.org/10.1167/10.9.19

Marr D, Nishihara HK. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B Biol Sci. Feb 23;200(1140):269-94. PubMed PMID: 24223.

Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*(20), 1909–1913. https://doi.org/10.1016/j.cub.2012.08.009

Marlow, P. J., & Anderson, B. L. (2016). Motion and texture shape cues modulate perceived material properties. Journal of Vision, 16(1), Article 5. https://doi.org/10.1167/16.1.5

MathWorks, Inc. (1996). MATLAB : the language of technical computing : computation, visualization, programming : installation guide for UNIX version 5. Natwick :Math Works Inc.

Motoyoshi I, Matoba H. (2011). Variability in constancy of the perceived surface reflectance across different illumination statistics. Vision Res. 2012 Jan 15;53(1):30-9. doi: 10.1016/j.visres.2011.11.010. Epub PubMed PMID: 22138530.

Murata A, Gallese V, Luppino G, Kaseda M, Sakata H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. J Neurophysiol. 2000 May;83(5):2580-601. PubMed PMID:10805659.

Muryy A. A. Welchman A. E. Blake R. W. Fleming R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. Proceedings of the National Academy of Sciences, USA, 110 (6), 2413–2418.

Nayar, S. K., Fang, X.-S., & Boult, T. (1997). International Journal of Computer Vision, 21(3), 163–186. https://doi.org/10.1023/a:1007937815113

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A*, *15*(12), 2951. https://doi.org/10.1364/josaa.15.002951

Okazawa, G., Koida, K., & Komatsu, H. (2011). Categorical properties of the color term “GOLD.” Journal of Vision, 11(8), 4–4. https://doi.org/10.1167/11.8.4

Oliva A, Torralba A. The role of context in object recognition.(2007). Trends Cogn Sci. ;11(12):520-7. Epub. Review. PubMed PMID: 18024143.

Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PloS one*, *9*(1), e86488. https://doi.org/10.1371/journal.pone.0086488

Panichello, M. F., Cheung, O. S., & Bar, M. (2013, January 21). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, Vol. 3, p. 620. https://doi.org/10.3389/fpsyg.2012.00620

Pappas, T. N. (2002). Human vision and electronic imaging. *Journal of Electronic Imaging*, *10*(1), 10. https://doi.org/10.1117/1.1336802

Pappas, T. N. (2002). Human vision and electronic imaging. *Journal of Electronic Imaging*, *10*(1), 10. https://doi.org/10.1117/1.1336802

Parkhurst D, Law K, Niebur E. (2002). Modeling the role of salience in the allocation of overt visual attention. Vision Res. Jan;42(1):107-23. PubMed PMID:11804636.

Paulun, V. C., Schmidt, F., van Assen, J. J. R., & Fleming, R. W. (2017). Shape, motion, and optical cues to stiffness of elastic objects. *Journal of Vision*, *17*(1), 20. https://doi.org/10.1167/17.1.20

Peters RJ, Iyer A, Itti L, Koch C. (2005). Components of bottom-up gaze allocation in natural images. Vision Res. Aug;45(18):2397-416. PubMed PMID: 15935435.

Pinna, B., & Deiana, K. (2015). Material properties from contours: New insights on object perception. *Vision Research*, *115*, 280–301. https://doi.org/10.1016/j.visres.2015.03.014

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. Journal of Vision, 15(8), 13. https://doi.org/10.1167/15.8.13

Phillips, F. (2004). Creating Noisy Stimuli. Perception, 33(7), 837–854. https://doi.org/10.1068/p5141

Porta, F. Preposterous (2016). [Video file]. Retrieved from https://www.florentporta.com/project/preposterous

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32(1), 3–25. https://doi.org/10.1080/00335558008248231

Price CJ, Humphreys GW. The effects of surface detail on object categorization and naming.(1989).QJ Exp Psychol A.;41(4):797-827. PubMed PMID: 2587799.

Ramachandran, V. S. (1988). Perception of shape from shading. Nature, 331(6152), 163–166. https://doi.org/10.1038/331163a0

Sakano Y, Ando H. Effects of head motion and stereo viewing on perceived glossiness.(2010). J Vis.;10(9):15. doi: 10.1167/10.9.15. PubMed PMID: 21106677.

Schütz AC, Braun DI, Gegenfurtner KR. (2011). Eye movements and perception: a selective review. J Vis. Sep 14;11(5). pii: 9. doi: 10.1167/11.5.9. Review. PubMed PMID: 21917784.

Scocchia L, Valsecchi M, Gegenfurtner KR, Triesch J. Visual working memory contents bias ambiguous structure from motion perception (2013). PLoS One;8(3):e59217. doi: 10.1371/journal.pone.0059217. Epub. PubMed PMID: 23527141; PubMed Central PMCID: PMC3602104.

Schmid, A. C., & Anderson, B. L. (2017). Perceptual dimensions underlying lightness perception in homogeneous center-surround displays. Journal of Vision, 17(2), 6. https://doi.org/10.1167/17.2.6

Schmid, A. C., & Doerschner, K. (2018). Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *Journal of Vision*, *18*(1), 14. https://doi.org/10.1167/18.1.14

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vision Research, 122, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: a selective review. Journal of Vision, 11(5):9, 1-30

Schmidt, F., Hegele, M., & Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, *17*(11), 10. https://doi.org/10.1167/17.11.10

Seuss, D. (1949). Bartholomew and the Oobleck. New York: Random House.

Seuss, D. (1957). The Lorax. New York: Random House.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2008). Eye movements for shape and material perception. Journal of Vision, 8(6), 219–219. https://doi.org/10.1167/8.6.219a

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of Vision*, *14*(9), 1–24. https://doi.org/10.1167/14.9.12

Smith TJ, Mital PK.(2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. J Vis. Jul 17;13(8). pii: 16.doi: 10.1167/13.8.16. PubMed PMID: 23863509.

Soechting, J. F., Juveli, J. Z., & Rao, H. M. (2009). Models for the Extrapolation of Target Motion for Manual Interception. *Journal of Neurophysiology*, *102*(3), 1491–1502. https://doi.org/10.1152/jn.00398.2009

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, Vol. 15, pp. 745–756. https://doi.org/10.1038/nrn3838

Tatler BW, Wade NJ, Kwan H, Findlay JM, Velichkovsky BM. (2010). Yarbus, eye movements, and vision. Iperception. 2010;1(1):7-27. doi: 10.1068/i0382. Epub Jul 12. PubMed PMID: 23396904; PubMed Central PMCID: PMC3563050.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system (1996). Nature;381(6582):520-2. PubMed PMID: 8632824.

Todd JT, Norman JF, Mingolla E. Lightness constancy in the presence of specular highlights (2004). Psychol Sci.;15(1):33-9. PubMed PMID: 14717829.

Toscani M, Valsecchi M, Gegenfurtner KR. (2013). Selection of visual information for lightness judgements by eye movements. Philos Trans R Soc Lond B Biol Sci.Sep 9;368(1628):20130056. doi: 10.1098/rstb.2013.0056. Print 2013 Oct 19. PubMed PMID: 24018718; PubMed Central PMCID: PMC3758199.

Toscani M, Valsecchi M, Gegenfurtner KR. Optimal sampling of visual information for lightness judgments (2013). Proc Natl Acad Sci U S A. 2;110(27):11163-8. doi: 10.1073/pnas.1216954110. Epub. PubMed PMID: 23776251; PubMed Central PMCID: PMC3704015.

Toscani, M., Zdravkovic, S., & Gegenfurtner, K. R. (2016). Lightness perception for surfaces moving through different illumination levels. Journal of Vision, 16(15), 21. https://doi.org/10.1167/16.15.21

Triesch J, Ballard DH, Hayhoe MM, Sullivan BT. (2003). What you see is what you need. J Vis.;3(1):86-94. PubMed PMID: 12678628.

Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objets: recognition, awareness, and modularity. Current Opinion in Neurobiology, 8(2), 218–226. https://doi.org/10.1016/s0959-4388(98)80143-8

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends in Cognitive Sciences*, *21*(9), 649–665. https://doi.org/10.1016/j.tics.2017.05.012

Underwood G, Foulsham T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. Q J Exp Psychol (Hove). Nov;59(11):1931-49. PubMed PMID: 16987782.

Urgen, B. M., & Boyaci, H. (2019). Supplementary Material for When expectations are not met : unraveling the computational mechanisms underlying the effect of expectation on perceptual thresholds. *BioArxiv*, *28*, 2–3. https://doi.org/10.1016/j.cub.2017.12.037

Urgen, B. M., & Boyaci, H. (2019). When expectations are not met: unraveling the computational mechanisms underlying the effect of expectation on perceptual thresholds. *BioRxiv*, 545244. https://doi.org/10.1101/545244

van Assen, J. J. R., Barla, P., & Fleming, R. W. (2018). Visual Features in the Perception of Liquids. *Current Biology*, *28*(3), 452-458.e4. https://doi.org/10.1016/j.cub.2017.12.037

van Assen, J. J. R., & Fleming, R. W. (2016). Influence of optical material properties on the perception of liquids. *Journal of Vision*, *16*(15), 12. https://doi.org/10.1167/16.15.12

Wendt G, Faul F, Mausfeld R. Highlight disparity contributes to the authenticity and strength of perceived glossiness(2008). J Vis.;8(1):14.1-10. doi: 10.1167/8.1.14. PubMed PMID: 18318617.

Wiebel CB, Toscani M, Gegenfurtner KR. Statistical correlates of perceived gloss in natural images (2015). Vision Res. Oct;115(Pt B):175-87. doi:10.1016/j.visres.2015.04.010. Epub 2015 Apr 29. PubMed PMID: 25937518.

Witzel C, Valkova H, Hansen T, Gegenfurtner KR(2011). Object knowledge modulates colour appearance. Iperception. 2011;2(1):13-49. doi: 10.1068/i0396. Epub. PubMed PMID: 23145224; PubMed Central PMCID: PMC3485772.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. Journal of Experimental Psychology: Human Perception and Performance, 19(4), 899–911. https://doi.org/10.1037/0096-1523.19.4.899

Yarbus A. L., (1967) Eye Movements and Vision (New York: Plenum Press)

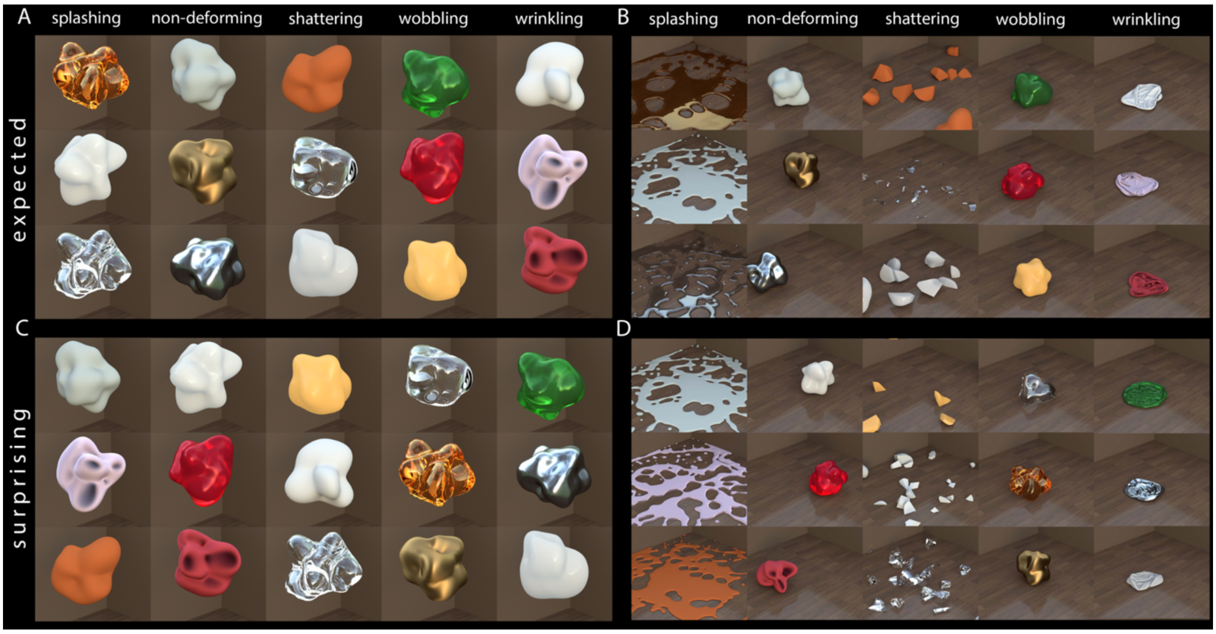
Yilmaz, O., & Doerschner, K. (2014). Detection and localization of specular surfaces using image motion cues. Machine Vision and Applications, 25(5), 1333–1349. https://doi.org/10.1007/s00138-014-0610-9

Zanganeh Momtaz H, Daliri MR. (2016). Predicting the eye fixation locations in the gray scale images in the visual scenes with different semantic contents. Cogn. Neurodyn. Feb;10(1):31-47. doi: 10.1007/s11571-015-9357-x. Epub.

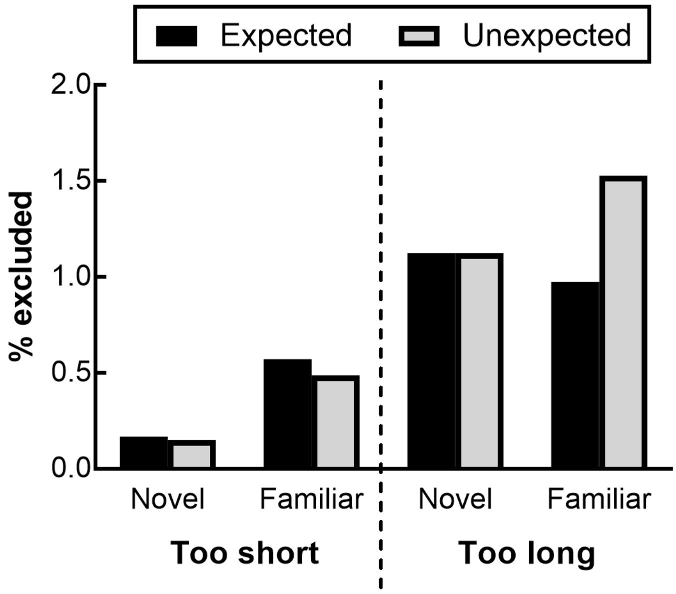
Zelinsky G.J., Sheinberg DL. (1997).Eye movements during parallel-serial visual search. J Exp. Psychol Hum Percept Perform. Feb;23(1):244-62. PubMed PMID: 9090154

Zenodo. https://doi.org/ 10.5281/zenodo.3378916

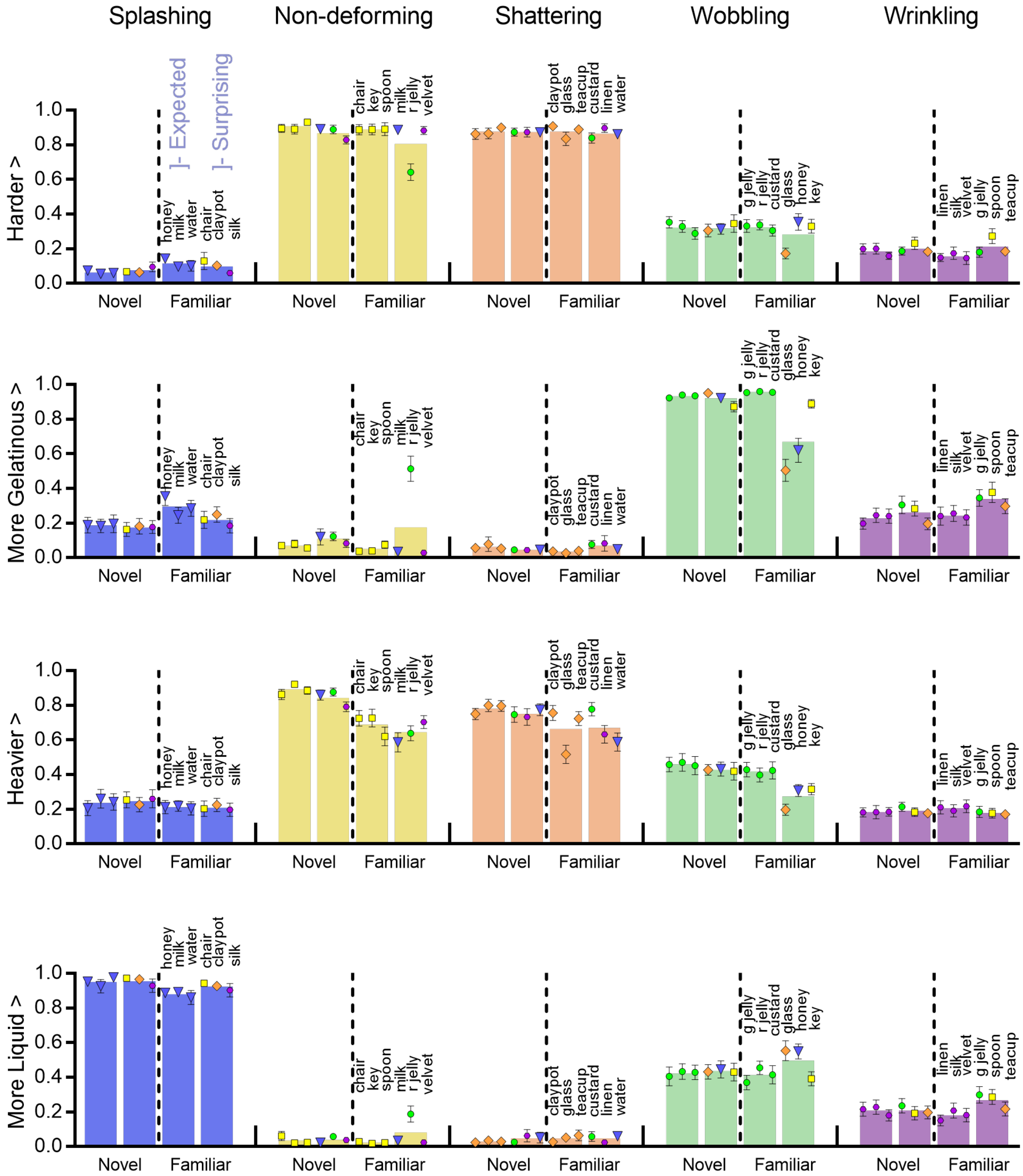
Supplementary Figures



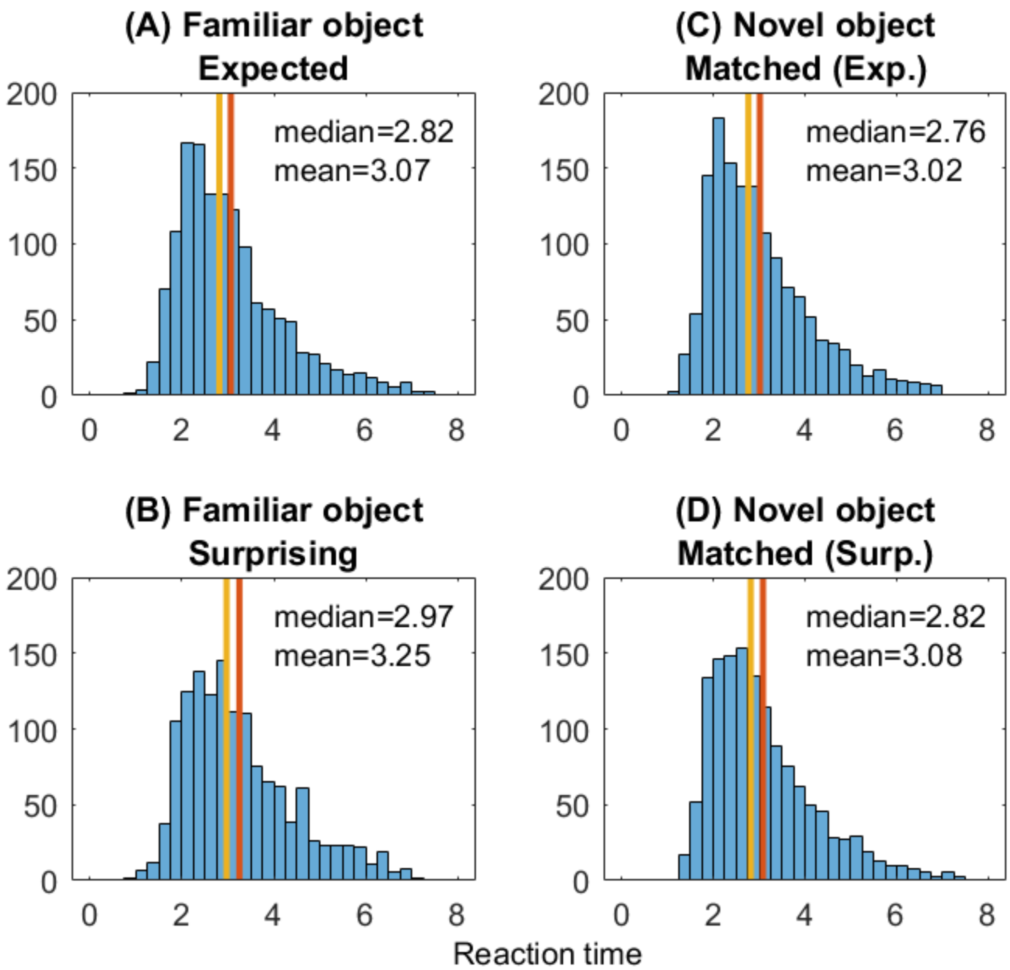
Supplementary Figure 1. Novel Objects inherit their optical and mechanical properties from the corresponding familiar objects (See Figure 3). Panel A shows Novel Objects grouped according to their material mechanics. Panel B depicts corresponding last frames of animations that show how a given Novel Objects falls to the ground. Panel C shows which Novel Objects deformed how in the surprise condition, and Panel D, the corresponding last frames of the animation. Even though control objects do not have a familiar shape, their shape and their 'inherited' optical material properties may nevertheless elicit some expectations. E.g. a bounded solid object may tend to be judged as solid and hard, bronze-colored object as heavier or red translucent smoothly curved objects as softer.



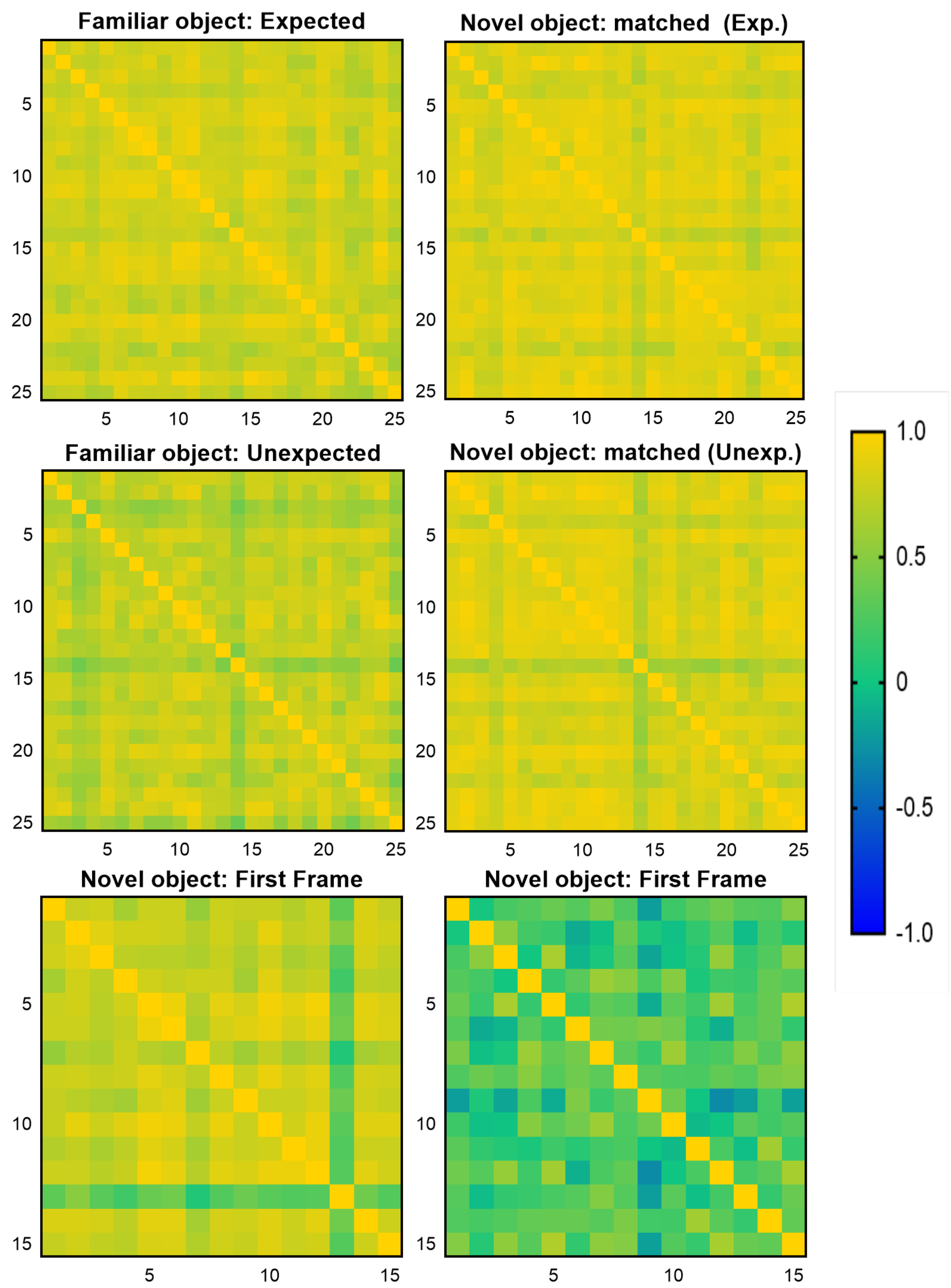
Supplementary Figure 2. RT exclusions. Trials where Familiar objects that behave unexpectedly are excluded more than other types of trials. It is also interesting that more Familiar objects were excluded for too short RTs – suggests that observers anticipated the rating during the 3 second static hold. Supports the idea of object knowledge influencing the ratings.



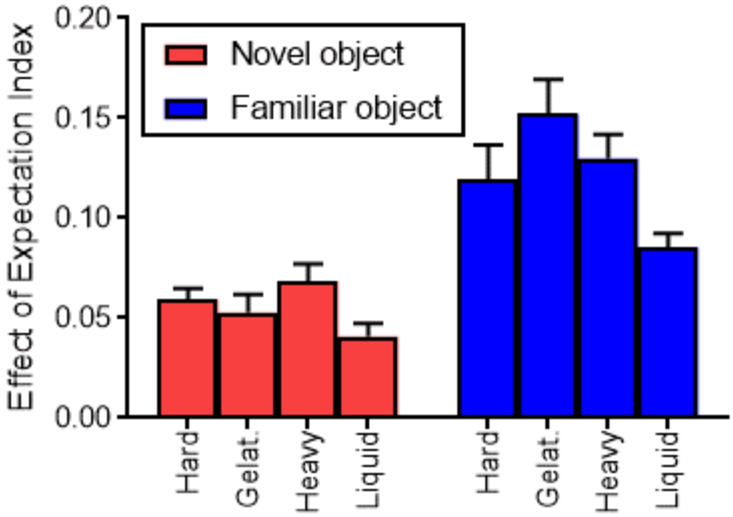
Supplementary Figure 3. Ratings for all attributes. Ratings of hard and gelatinous are discussed in the main manuscript. Error bars are one standard error of the mean. In general, ratings of heavy and liquid also make sense: splashing objects (blue bars) are rated as highly liquid, non-deforming objects (yellow bars) and shattering objects (orange bars) are rated not at all liquid, and wobbling objects (green bars) are rated in between. The directionality of some ratings is not straight forward on first view, when comparing expected and surprising conditions, e.g. for splashing objects (blue bars). We might have predicted the splashing honey, milk, and water to be rated as less hard and more liquid than the chair, clay pot, and silk curtain that splash surprisingly – object knowledge about chairs, clay pots and silk curtains should “pull up” hardness ratings and “pull down” liquid ratings. However, we see the opposite pattern. Ratings of how gelatinous these objects appear may shed some light on this: splashing honey, milk, and water appear more gelatinous than all other splashing objects (even Novel objects). This suggests that the shape of the droplets in the initial frame do not generate compelling impressions of very runny liquids, but rather some thicker, more viscous substances (First Frame ratings in Supplementary Figure 5 confirm this idea). So, in the case of splashing stuff, object knowledge interferes in our Expected condition in a way that we did not anticipate.



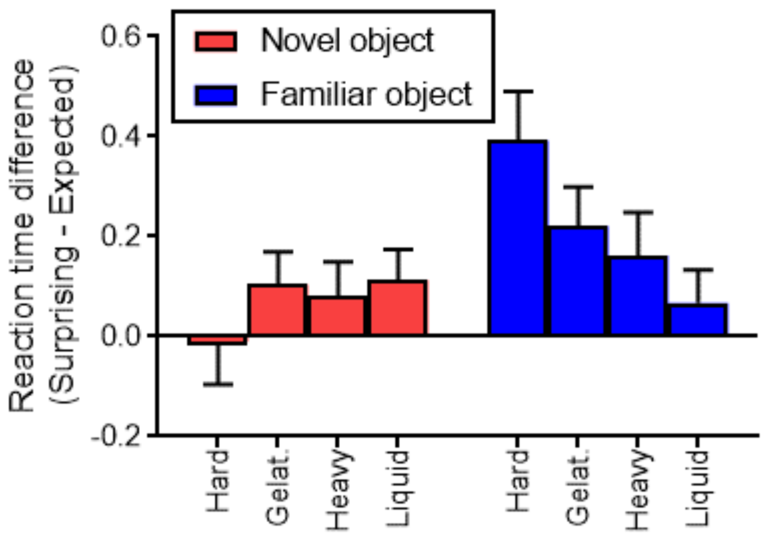
Supplementary Figure 4. Histograms for Reaction Times. Reaction times are slower for Familiar objects that behave surprisingly (B) versus expectedly (A). This difference is not significant for Novel objects with matched outcomes (C and D).



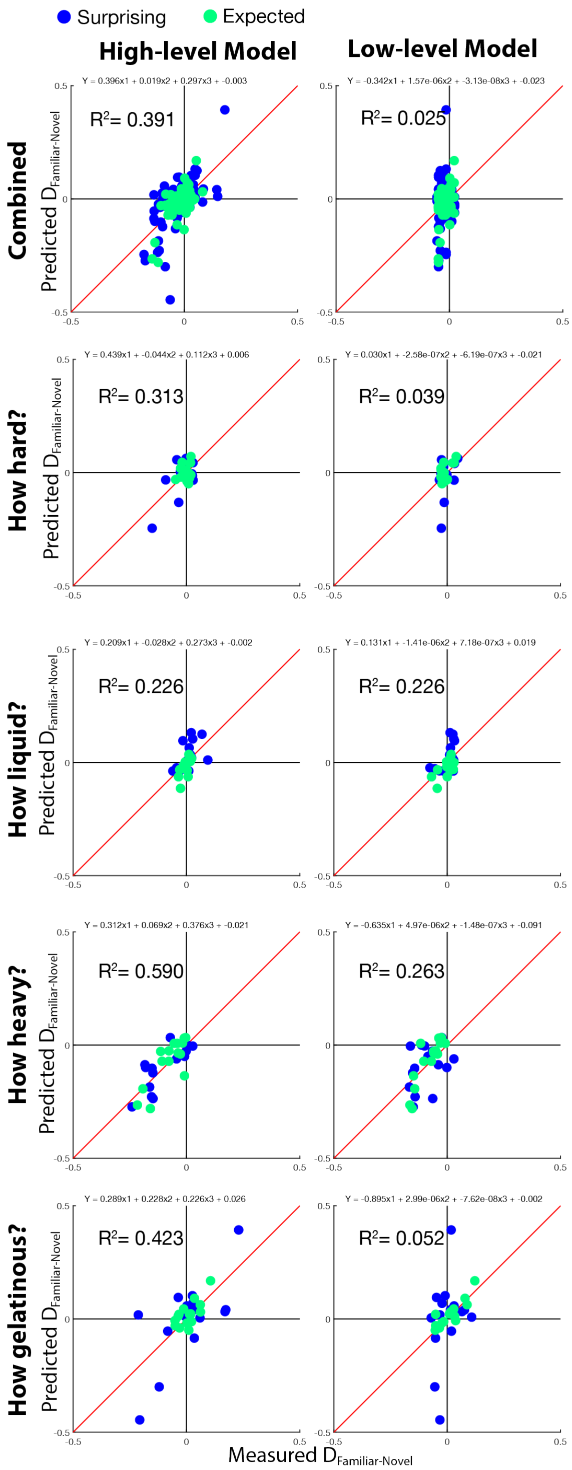
Supplementary Figure 5. Interobserver correlations comparing ratings of each observer with every other observer.



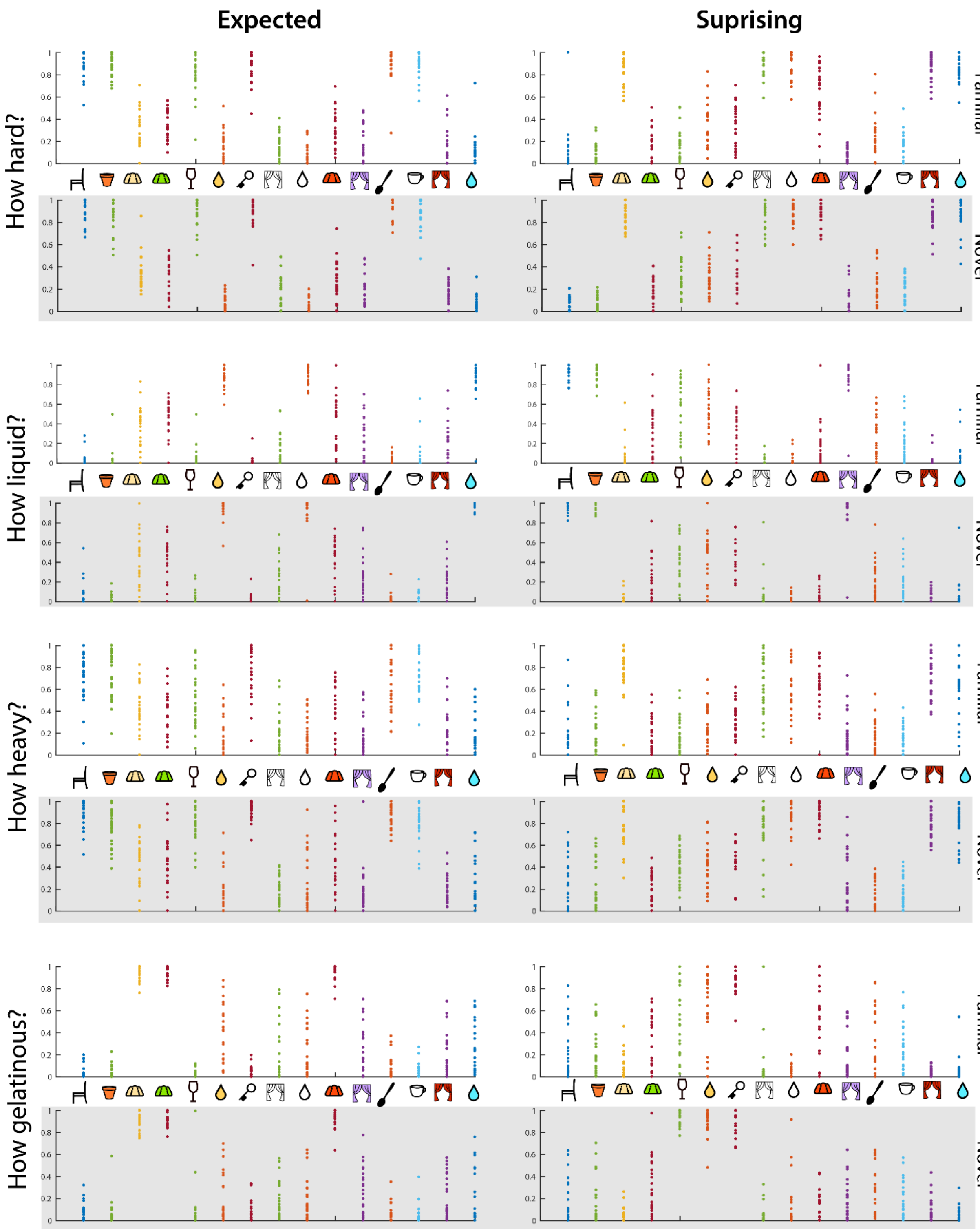
Supplementary Figure 6. Effect of Expectation Index ϵ over time.



Supplementary Figure 7. Reaction Time difference $\tau_D$over time.



Supplementary Figure 8. Low and high level linear regression models predicting the difference between ratings of moving familiar and control objects $D_{familiar\text{-}novel}$, related to Figure 3. We developed two models with the aim to account for the differences we observed in ratings of familiar and novel moving objects. Overall, the high-level model (left side) was more successful in predicting this difference than the low-level model (right side). However, this varied as a function of questions. The high-level model performed best for ratings of gelatinousness and hardness, whereas the low-level model performed as good as the high-level model for ratings of liquidness. This makes sense since ratings of the latter might be dominated by how much a substance physically spreads in the image, whereas the former requires more high-level inference. Low- and high-level models had each exactly 3 predictors.



Supplementary Figure 9. Overall rating distributions, related to Figure 4. Shown all observers ratings for each of the 15 familiar objects and corresponding novel objects, all attributes (hard, liquid, heavy, gelatinous) in Expected (left) and Surprising motion conditions (right). It is quite apparent that mean ratings of approximately 0.5 in any condition are not due to 'bimodal' rating behavior, i.e. some observers giving very low and other observers giving very high ratings. Overall observers tended to agree well in their ratings of these four attributes (also see Figure 4).

			|FF familiar objects - motion novel objects|						Prior Pull= |Moving Familiar - Corresp. Novel Objects|					
			EXPECTED			SURPRISING			EXPECTED			SURPRISING		
DEFORMATION	BLOCK		Sig. diff	p-value	alpha	Sig. diff	p-value	alpha	Sig. diff	p-value	alpha	Sig. diff	p-value	alpha
		chair	0.085	0.003	0.0016	**0.743**	**0.000**	0.0016	0.008	0.367	0.0027	0.061	0.155	0.0027
	hard	clay pot	0.029	0.359	0.0016	**0.827**	**0.000**	0.0016	0.044	0.390	0.0027	0.039	0.109	0.0027
		silk	0.090	0.009	0.0016	0.015	0.625	0.0016	0.026	0.231	0.0027	0.035	0.269	0.0027
		chair	0.003	0.871	0.0016	0.090	0.032	0.0016	0.031	0.090	0.0027	0.056	0.061	0.0027
	gelatinous	clay pot	0.017	0.507	0.0016	0.109	0.023	0.0016	0.019	0.426	0.0027	0.068	0.049	0.0027
melt		silk	**0.171**	**0.001**	0.0016	**0.104**	**0.009**	0.0016	0.011	0.889	0.0027	0.007	0.521	0.0027
		chair	**0.107**	**0.001**	0.0016	**0.502**	**0.000**	0.0016	0.137	0.003	0.0027	0.051	0.404	0.0027
	heavy	clay pot	**0.121**	**0.001**	0.0016	**0.401**	**0.000**	0.0016	0.006	0.881	0.0027	0.002	0.910	0.0027
		silk	0.022	0.593	0.0016	0.100	0.075	0.0016	0.007	0.877	0.0027	0.063	0.104	0.0027
		chair	0.007	0.803	0.0016	**0.903**	**0.000**	0.0016	0.032	0.195	0.0027	0.029	0.095	0.0027
	liquid	clay pot	0.017	0.074	0.0016	**0.958**	**0.000**	0.0016	0.003	0.029	0.0027	0.038	0.007	0.0027
		silk	**0.149**	**0.002**	0.0016	**0.848**	**0.000**	0.0016	0.021	0.545	0.0027	0.025	0.663	0.0027
		milk	**0.072**	**0.000**	0.0095	**0.762**	**0.000**	0.0095	**0.041**	**0.015**	0.0280	0.004	0.794	0.0280
	hard	red Jello	**0.182**	**0.000**	0.0095	**0.782**	**0.000**	0.0095	0.017	0.673	0.0280	**0.247**	**0.000**	0.0280
		velvet	0.026	0.215	0.0095	**0.643**	**0.000**	0.0095	0.011	0.430	0.0280	0.055	0.168	0.0280
		milk	0.087	0.041	0.0095	**0.150**	**0.003**	0.0095	**0.062**	**0.008**	0.0280	0.086	0.069	0.0280
	gelatinous	red Jello	0.033	0.071	0.0095	**0.779**	**0.000**	0.0095	0.020	0.196	0.0280	**0.392**	**0.000**	0.0280
rigid		velvet	**0.202**	**0.000**	0.0095	0.044	0.067	0.0095	0.008	0.881	0.0280	0.055	0.036	0.0280
		milk	0.154	0.010	0.0095	**0.754**	**0.000**	0.0095	0.041	0.289	0.0280	**0.274**	**0.000**	0.0280
	heavy	red Jello	0.098	0.076	0.0095	**0.522**	**0.000**	0.0095	0.027	0.760	0.0280	**0.238**	**0.000**	0.0280
		velvet	**0.095**	**0.001**	0.0095	**0.511**	**0.000**	0.0095	0.031	0.689	0.0280	**0.088**	**0.003**	0.0280
		milk	0.020	0.625	0.0095	**0.881**	**0.000**	0.0095	0.034	0.519	0.0280	0.009	0.344	0.0280
	liquid	red Jello	0.032	0.449	0.0095	**0.339**	**0.000**	0.0095	0.014	0.782	0.0280	**0.130**	**0.009**	0.0280
		velvet	**0.096**	**0.009**	0.0095	**0.046**	**0.001**	0.0095	0.001	0.979	0.0280	0.014	0.187	0.0280
		custard	**0.181**	**0.000**	0.0019	**0.701**	**0.000**	0.0019	0.021	0.841	0.0037	0.034	0.305	0.0037
	hard	linen	0.081	0.009	0.0019	0.755	0.000	0.0019	0.051	0.109	0.0037	0.024	0.343	0.0037
		water	**0.088**	**0.000**	0.0019	0.724	0.000	0.0019	0.042	0.223	0.0037	0.011	0.278	0.0037
		custard	**0.158**	**0.000**	0.0019	0.719	0.000	0.0019	0.030	0.008	0.0037	0.031	0.072	0.0037
	gelatinous	linen	**0.146**	**0.000**	0.0019	0.009	0.621	0.0019	0.043	0.349	0.0037	0.039	0.336	0.0037
shatter		water	0.027	0.587	0.0019	**0.177**	**0.000**	0.0019	0.090	0.032	0.0037	0.003	0.792	0.0037
		custard	0.115	0.011	0.0019	**0.403**	**0.000**	0.0019	0.030	0.356	0.0037	0.031	0.217	0.0037
	heavy	linen	0.041	0.137	0.0019	**0.511**	**0.000**	0.0019	0.029	0.194	0.0037	0.101	0.050	0.0037
		water	0.085	0.090	0.0019	**0.617**	**0.000**	0.0019	0.037	0.086	0.0037	**0.187**	**0.004**	0.0037
		custard	0.055	0.310	0.0019	**0.325**	**0.000**	0.0019	0.036	0.391	0.0037	0.032	0.161	0.0037
	liquid	linen	**0.138**	**0.002**	0.0019	0.014	0.691	0.0019	0.064	0.064	0.0037	0.039	0.268	0.0037
		water	**0.029**	**0.001**	0.0019	**0.895**	**0.000**	0.0019	0.116	0.008	0.0037	0.007	0.750	0.0037
		glass	**0.062**	**0.046**	0.0464	**0.499**	**0.000**	0.0464	0.032	0.810	0.0037	0.133	0.020	0.0037
	hard	honey	**0.058**	**0.000**	0.0464	**0.182**	**0.000**	0.0464	0.070	0.045	0.0037	0.040	0.263	0.0037
		key	0.025	0.386	0.0464	**0.569**	**0.000**	0.0464	0.002	0.698	0.0037	0.015	0.255	0.0037
		glass	0.012	0.784	0.0464	**0.884**	**0.000**	0.0464	0.051	0.208	0.0037	**0.447**	**0.000**	0.0037
	gelatinous	honey	**0.232**	**0.000**	0.0464	**0.501**	**0.000**	0.0464	**0.167**	**0.001**	0.0037	**0.301**	**0.001**	0.0037
wobble		key	0.010	0.673	0.0464	**0.801**	**0.000**	0.0464	0.041	0.093	0.0037	0.017	0.386	0.0037
		glass	**0.443**	**0.000**	0.0464	0.072	0.037	0.0464	**0.282**	**0.000**	0.0037	0.230	0.000	0.0037
	heavy	honey	0.057	0.191	0.0464	**0.281**	**0.000**	0.0464	0.006	0.559	0.0037	**0.124**	**0.004**	0.0037
		key	**0.518**	**0.000**	0.0464	0.018	0.651	0.0464	0.195	0.004	0.0037	0.104	0.141	0.0037
		glass	0.008	0.566	0.0464	**0.404**	**0.000**	0.0464	0.016	0.258	0.0037	0.123	0.067	0.0037
	liquid	honey	0.028	0.157	0.0464	**0.478**	**0.000**	0.0464	0.065	0.050	0.0037	0.103	0.012	0.0037
		key	0.008	0.407	0.0464	**0.418**	**0.000**	0.0464	0.002	0.859	0.0037	0.039	0.579	0.0037
		green Jello	**0.145**	**0.000**	0.0016	0.004	0.885	0.0016	0.009	0.660	0.0274	0.007	0.418	0.0274
	hard	spoon	0.047	0.029	0.0016	**0.651**	**0.000**	0.0016	0.040	0.371	0.0274	0.042	0.609	0.0274
		teacup	0.065	0.034	0.0016	**0.653**	**0.000**	0.0016	0.012	0.791	0.0274	0.002	0.722	0.0274
		green Jello	**0.049**	**0.002**	0.0016	**0.584**	**0.000**	0.0016	0.021	0.031	0.0274	0.039	0.964	0.0274
	gelatinous	spoon	0.008	0.642	0.0116	**0.219**	**0.000**	0.0116	0.019	0.417	0.0274	0.094	0.275	0.0274
wrinkle		teacup	0.026	0.180	0.0016	**0.116**	**0.002**	0.0016	0.013	0.396	0.0274	0.101	0.001	0.0274
		green Jello	0.167	0.003	0.0016	**0.091**	**0.003**	0.0016	0.073	0.221	0.0274	0.028	0.378	0.0274
	heavy	spoon	**0.504**	**0.000**	0.0016	**0.199**	**0.000**	0.0016	**0.266**	**0.000**	0.0274	0.006	0.752	0.0274
		teacup	**0.339**	**0.000**	0.0016	**0.281**	**0.000**	0.0016	0.073	0.041	0.0274	0.006	0.389	0.0274
		green Jello	0.059	0.207	0.0016	**0.257**	**0.000**	0.0016	0.022	0.246	0.0274	0.063	0.132	0.0274
	liquid	spoon	0.010	0.387	0.0016	**0.159**	**0.000**	0.0016	0.001	0.910	0.0274	0.095	0.027	0.0274
		teacup	0.004	0.748	0.0016	**0.171**	**0.000**	0.0016	0.034	0.260	0.0274	0.020	0.659	0.0274

Supplementary Table 1. Statistical results for rating differences; FF familiar objects and motion novel objects and prior pull.

This table highlights significant absolute rating differences between **FF familiar objects and motion novel objects (left blue 2 columns) and moving familiar and corresponding novel objects (prior pull, right gray 2 columns),** arranged by deformation method, rating question, and object identity. Values are FDR-corrected, based on a family-wise alpha of .05, for 24 tests (for each deformation: 12 tests for each of the Expected and Surprising conditions). Columns represent differences, p-values, and FDR alpha values for both Expected and Surprising conditions. Bolded values indicate statistically significant differences.