How we see liquids

Jan Jaap van Assen

How we see liquids

Jan Jacob Reindert van Assen

Department of Psychology

Justus-Liebig-Universität Gießen

This dissertation is submitted for the degree of

Doctor of Philosophy

January 2018

Acknowledgements

I want to thank my PhD supervisor Roland Fleming for his patience, knowledge, great mentoring and creating a great work atmosphere. I have learned many things in Gießen mostly because of Roland. Being part of the Marie Curie ITN PRISM network, which was coordinated by Roland, allowed me to meet great people and many friends on beautiful locations. Because of this research network I worked three months in Madrid and one month in Bordeaux which were both great experiences. He attracts people who have love and passion for their work resulting a great office culture.

Karl Gegenfurtner, the head of the department, deserves credit as well for running the department so successfully. He and Roland attract great speakers for the weekly colloqium talks, provide generous facilities and a nice internationally oriented work environment.

I would like to thank my lab members, Vivian, Filipp, Eugen, Yaniv, Guido, Lina, Kate, and ex-lab members, Patrick and Steve, who had to deal with my personal traits in person. They were always in for a laugh or helping me out with problems. They were a big part of the nice environment at work.

I would like to thank other scientists who helped me with professional or personal advice throughout my PhD, Sylvia Pont, Maarten Wijntjes, Flip Phillips, Katja Doerschner, Dicle Dövencioğlu, Andrea van Doorn and Jan Koenderink.

The people at Next Limit in Madrid deserve some credit as well for providing me with the necessary training and software to generate my stimuli. Especially Ángel Tena and Alex Ribao have been very supportive throughout the years.

Finally I am very grateful for my parents Jaap and Martha, and my sisters Sophie and Chris, who have been extremely supportive throughout my life.

This Club of Vertuoso's, upon a full Night, when some eminent Maggot-monger, for the Satisfaction of the Society, had appointed to demonstrate the Force of Air, by some hermetical Pot gun, to shew the Difference of the Gravity between the Smoak of Tobacco and that of Colts-foot and Bittany, or to try some other such like Experiment, were always compos'd of such an odd Mixture of Mankind, that, like a Society of Ringers at a quarterly Feast, here sat a fat purblind Philosopher next to a talkative Spectacle-maker; yonder a half-witted Whim of Quality, next to a ragged Mathematician; on the other Side a consumptive Astronomer next to a water-gruel Physician; above them, a Transmutator of Metals, next to a Philosopher-Stone-Hunter; at the lower End, a prating Engineer, next to a clumsy-fisted Mason; at the upper End of all, perhaps, an Atheistical Chymist, next to a whimsy-headed Lecturer; and these the learned of the Wise-akers wedg'd here and there with quaint Artificers, and noisy Operators, in all Faculties; some bending beneath the Load of Years and indefatigable Labour, some as thin-jaw'd and heavy-ey'd, with abstemious Living and nocturnal Study as if, like Pharaoh's Lean Kine, they were designed by Heaven to warn the World of a Famine; others looking as wild, and disporting themselves as frenzically, as if the Disappointment of their Projects had made them subject to a Lunacy. When they were thus met, happy was the Man that could find out a new Star in the Firmament; discover a wry Step in the Sun's Progress; assign new Reasons for the Spots of the Moon, or add one Stick to the Bundle of Faggots which have been so long burthensome to the back of her old Companion; or, indeed, impart any crooked Secret to the learned Society, that might puzzle their Brains, and disturb their Rest for a Month afterwards, in consulting upon their Pillows how to straiten the Project, that it might appear upright to the Eye of Reason, and the knotty Difficulty to be rectify'd, as to bring Honour to themselves, and Advantage to the Public.

—NED WARD, The Vertuoso's Club

Abstract

We have great understanding of objects and materials we encounter in everyday life. This helps us to quickly identify what is predator and what is prey, what is eatable and poisonous. Despite large image differences our visual system is able to extract material properties very consistently. Liquids are a category of materials that appear to be particularly challenging, due to their volatile nature. We are able to estimate complex liquid properties such as runniness or sliminess. How are we able to do this? How is it possible that we can perceive that honey is thicker than milk. Or that water in a glass is the same material as water spraying in a fountain. Four studies were conducted to achieve a better understanding of the image information we use to estimate liquid properties.

In study 1 we specifically look at the contributions of optical cues while estimating a range of liquid properties. Using the same liquid shapes, but with different optical appearances, we studied which perceived properties (e.g., runniness) are influenced by optical or mechanical cues.

We can encounter liquids in many different states and contexts. In study 2 we specifically look at the constancy of viscosity perception despite radical changes in shape. How consistently do we actually perceive liquids? We simulated a range of different scenes to learn how sensitive observers are to shape changes when estimating viscosity.

In study 3 we look into specific shape features underlying visual inferences about liquids. By comparing observers' viscosity ratings with perceived shape features, we show how the brain exploits 3D shape and motion cues to infer viscosity across contexts despite dramatic image changes.

In study 4 we estimate the perceived viscosity of an image with neural networks. Machine learning is a powerful tool and facilitates major breakthroughs with difficult visual tasks. Here we trained a neural network specifically designed to mimic human performance while estimating viscosity.

Our results show that the perception of liquids is mainly driven by optical, shape and motion cues. We show great perceptual constancy in rating viscosity across a wide range of scenes. Mid-level features (e.g., spread, pulsing) are an important and reliable source to estimate viscosity consistently across contexts.

Table of contents

List of figures

Chapter 1

Introduction

A digital version of this dissertation is availible for download:
http://www.janjaap.info/dissertation/dissertation.pdf

Humans differentiate themselves from other animal species in various ways. One very important difference is how we interact with our environment. We have great understanding of objects and materials we encounter in everyday tasks. This understanding contributes to our survival, we must know what is predator or prey, when something is eatable or poisonous, when an object glows and is hot to the touch, when a surface is smooth and slippery. Mostly through vision we achieve these property driven affordances prior to any interactions with the actual object. These concepts of intrinsic properties seem to come quite naturally to us but use highly sophisticated and efficient paradigms. This is becoming evidently more clear in this last decade where we more intensively try to reproduce these qualities in machinery.

Different types of information need to be extracted from the retinal image to be able to make material property estimations. Optical cues provide information about the surface properties of objects, or in the case of transparent and translucent materials they provide information about the object's content as well. How we interpret optical surface properties is influenced by shape and illumination conditions. The shape itself can be informative about the causal meaning of the object (e.g., a crushed can or bitten apple, Spröte et al. 2016). There is temporal information, how objects move, change in shape or interact with other objects (Paulun et al., 2017; Schmidt et al., 2017). The environment around an object also provides context and affordances (Oliva and Torralba, 2007). Overall there are many sources of information in an image that allow us to make material property estimations. The hard part is that all these different sources of visual information are not mutually exclusive, making it extra difficult to single out specific processes that could explain how

we actually perceive properties such as roughness, sliminess, softness, runniness, and shininess.

Another impressive feature of our visual system is that we are able to estimate material properties consistently across an immense space of possible depictions. The consistency at which we perceive material properties shows the true power of our visual system where we can identify objects and materials despite radical changes in the retinal image (e.g., spaghetti submerged in a dark, green lit, underwater cave still looks like soft, breakable, flexible, plausibly delicious, spaghetti). The computations involved to achieve perceptual constancy must represent properties with concepts that are invariant across contexts.

To achieve a better understanding on how we use these different types of visual information we need to use controlled environments in which we can parse individual cues, isolate them, and limit contamination of other types of information. By now advances in computer graphics enable us to build these controlled environments with a level of photorealism that is difficult to discern from real images. The problem is that we combine cues using 'weak fusion' processes to come to our final estimation of a material property (Landy et al., 1995; Ernst and Bülthoff, 2004). Individual cues contain errors and more accurate estimates can be acquired by combining the separate cues. The problem is how you assign the contribution of individual cues to differences in estimation performance in an experimental setup. The precision with which we can study different types of visual information increases as technology advances. But being able to reproduce specific cues in controlled environments doesn't automatically mean you will be able to explain material estimates. We need to apply new techniques to be able to derive individual cue contributions and in which proportions they contribute to the final material estimate. This is becoming even more important now that research is moving from relatively low dimensional problems (e.g., color, gloss perception) towards higher dimensional problems such as the perception of nonrigid, breaking materials (Schmid and Doerschner, 2018) or predicting shape transformation processes (Schmidt and Fleming, 2016).

1.1 Classification of materials

Classification of materials is very important because it allows us to assign prior knowledge to that specific material you recognize. This creates expectations that help to interact with the materials and perform planned actions successfully. By identifying material classes such as stone, fabric, metals, liquids we are able to quickly adjust our expectations and therefore improving subsequent interactions with the environment. This quality is essential to eat food, drive a car or successfully fight an armored knight. Classification

helps to narrow down possible characteristics of the materials and to narrow down which cues are most informative to adopt even more detailed expectations. It is a hierarchical framework, where gold is a member of metals but by identifying it as a subclass of metals the feature space of for example shapes is massively reduced. Gold doesn't occur as often in bulky quantities as iron.

The large differences and similarities within classes make classification very difficult. Denim behaves similarly as woven cotton and papyrus, and papyrus and cotton are optically more similar, but we would assign papyrus to a different material class. Therefore we need more specific features to tease materials apart. There are many cases where materials in different categories have common properties. The space to classify materials is not uniform, some features might be very relevant to identify gels but not for metals, for example softness. Materials can also appear in different states such as glowing hot metal, wet fabrics, rotting bananas or dried out mud (Zaidi, 2011). For classification to be successful it needs to be able to group very different images together and split similar images apart.

Despite all these complications in this vast material space we are very able recognizing and classifying materials, even when only presented for a short moment (Sharan et al., 2009, 2013). There must be some lower dimensional construct that enables us to navigate through this space. Fleming et al. 2013 showed that nine subjective ratings, such as roughness and colorfulness, allowed observers to predict material classes with a 90% precision. Furthermore, observers assigned similar semantic labels to material categories. This agreement demonstrates that there are perception driven structures in large feature spaces that could classify a large proportion of the materials with fewer descriptors.

1.2 Why liquids?

Of all material classes liquids like water, yogurt and molasses are especially interesting because of their highly mutable shapes. Intrinsic properties (e.g. viscosity, density, surface tension) and external forces (e.g. object interactions, gravity) cause liquids to adopt a wide range of different shapes depending on the context. Despite this we are very well able to intuitively estimate properties of the liquids displayed in Figure 1.1. The physics behind liquids are extremely complex and it is safe to assume that we are not estimating liquid properties, such as viscosity, by computing Navier-Stokes equations or simulating molecular interactions (Bridson, 2015). This large spectrum of possible liquid appearances and the complexity of underlying physics poses a similar problem as with material categorization. Next to large space of possible shapes there is a mixture of other

informative cues that influence liquid estimates. Optical material properties, motion, and interactions with the environment provide different degrees of information. Therefore liquids are interesting to study since we must use some essential visual information that allows us to navigate through this dense liquid space. The space to describe all possible liquids has countless dimensions and some very efficient reduction of information must take place to be able to estimate properties accurately, and accurate we are (Kawabe et al., 2015; Paulun et al., 2015; Van Assen and Fleming, 2016; Van Assen et al., 2018). It is a different yet complex feature space that can provide an alternative perspective on similar problems that arrise with material categorization.

Figure 1.1: Liquids simulated with different viscosities and optical materials

1.2.1 Optical cues

Optical cues provide information about the optical material appearance. The visual appearance of a surface depends on three factors, (1) surface reflectance properties, (2) object geometry, (3) illumination conditions. These factors are often specified using the bidirectional reflectance distribution function (BRDF, Nicodemus 1965). Light can also be transferred through the object medium in two different ways, transparent materials transport light without being scattered (e.g., water, gases, high grade glass) and with translucent materials light scatters which results in faster absorption. Figure 1.2 shows the schematic representation of these concepts. These type of materials require different descriptive functions such as BTDF (Bidirectional transmittance distribution function, Bartell et al. 1981) or BSSRDF (Bidirectional scattering-surface reflectance distribution function, Jensen et al. 2001).

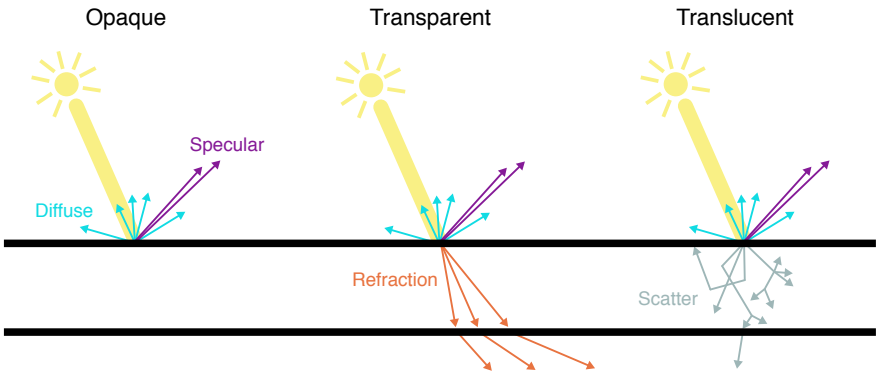


Figure 1.2: Differences between opaque, transparent and translucent materials

There is a large body of optical cue driven research. How we perceive and misperceive light (Koenderink et al., 2004; Ostrovsky et al., 2005; Pont and Koenderink, 2007), shape and surface textures (Landy and Graham, 2004; Dong and Chantler, 2005; Ho et al., 2008; Emrith et al., 2010; Liu et al., 2015), surface reflectance (Vangorp et al., 2007; Wijntjes and Pont, 2010; Marlow et al., 2012), more specifically gloss (Beck and Prazdny, 1981; Nishida and Shinya, 1998; Fleming et al., 2003; Motoyoshi et al., 2007; Kim et al., 2012; Van Assen et al., 2016, see Chadwick and Kentridge, 2015 for a recent review), and colour (see Foster, 2011, for a review). Most of this research is fixated on opaque materials while transparent (Fleming et al., 2011; Faul and Ekroll, 2012; Schlüter and Faul, 2014) and translucent (Fleming et al., 2004; Fleming and Bülthoff, 2005; Xiao et al., 2014) liquid appearances are not uncommon either. Above mentioned research has shown that despite making some mistakes in general we are very good at estimating optical material properties. Especially liquids we encounter in a wide range of optical appearances where opaque (paint, mercury, chocolate), transparent (water, high grade glass), and translucent materials (milk, honey, gels) are not uncommon. This large range of optical cues could provide critical information in properly estimating liquid properties.

1.2.2 Mechanical cues

Another group of cues are especially informative about the mechanics of liquids. Both shape and motion provide information that can be descriptive of the complicated underlying physics.

Shape cues

Shape can be very informative about the causal origins of an object which tells us more about possible material properties (Biederman and Gerhardstein, 1993; Gilchrist et al., 1999; Riesenhuber and Poggio, 2000). Causal origins are constant across contexts, they always manifest with similar characteristics, indepenent of scene. Shape perception research mainly concentrates on geometrical computations (e.g., curvatures, orientations, symmetries). However identifying types of deformation in liquid shapes will tell more about the behavior and characteristics of the liquid. It is unlikely that spraying droplets occur with very viscous liquids or that runny liquids will heap up in a spiraling manner. We can easily identify these features such as spatter and spiraling and this provides restrictions we use to navigate the perceptual space of liquids. To do so we need to have clear shape representations, but to compute shape from images remains a challenging task (Binford, 1981; Pentland, 1986; Biederman, 1987; Feldman et al., 2013).

Perceiving curvature and geometries of a surface is the first step, but to be able to estimate features as spiraling or elongation, local parts of that surface need to be grouped or to be organized together. Perceptual organization (Palmer, 1999; Wagemans et al., 2012) allows us to link different locations on an object to form higher conceptual features (e.g., blobby, sharp, twisted). Multiple concepts can be assigned to different parts of the object on different scales. The leg of a table can be twisted while the entire frame is very angular.

As mentioned before the perception of shape relies on the optical conditions such as illumination and surface reflectance, but shape itself provides much more information about the mechanics; how this shape came to be or likely will be (Nusseck et al., 2007; Battaglia et al., 2013; Bates et al., 2015). For us, to be able to identify local structures as blobby or irregular, we gain much more detailed information about the objects' origin. Blobby shapes look blobby from many different angles and lighting conditions allowing shape features to be a much more reliable source of information across changing contexts. It even helps us to disentangle external forces such as gravity or gusts of wind. This is a very powerful quality and enables us to generalize. Only a select few have experienced zero gravity environments, yet we can still imagine how a spiraling liquid would look differently in comparison with a more familiar scene in your kitchen.

Motion cues

Interesting is the contribution of motion cues because we are very well able to estimate liquid properties with static stimuli (Van Assen and Fleming, 2016). However, as with other material categories different types of cues contribute to achieve a more precise

percept. We don't need motion to make estimates of viscosity, but it will add precision. In some scenes motion alone is already very predictive of perceived viscosity (Kawabe et al., 2015), but there are contexts in which motion contributes less. Motion adds another rich dimension to liquids that is especially descriptive of both intrinsic and extrinsic forces that work on the liquid, e.g. viscous liquids move slowly. These forces manifest themselves as shape changes over time.

Motion improves the accuracy of shape perception (Norman and Todd, 1994; Caudek and Domini, 1998; Jain and Zaidi, 2011). Doerschner et al. 2011a showed that motion patterns that are produced by specular reflections, such as highlights, are highly depended on surface curvature. The specular reflections move quickly across areas with low curvature and stick in areas with high curvature. Measuring the optic flow (measurement of motion energy in sequential frames) demonstrates that shiny surfaces locally tend to have motion going in multiple directions at different velocities. In contrast, matte surfaces show much more homogeneous directions and velocities. This means that matte surfaces could convey less visual motion than specular images when illumination and shape are constant (Doerschner et al., 2011b). Most liquids have highly specular surfaces but translucent materials tend to obscure specular reflections because of the high amounts of scattering that takes place (e.g., the difference between perceived glossiness of water and milk). It might be that in certain cases the optical material properties might actually influence the magnitude with which motion is perceived.

1.3 Theoretical frameworks

1.3.1 Inverse optics

In material perception there are multiple theoretical approaches that provide a framework on how we extract information from images and assing meaning, how we perceive. One of these frameworks is an inverse optics approach (Barrow and Tenenbaum, 1978; Pizlo, 2001). Inverse optics suggests that our visual system models the physics of our environment. It simulates where the light source is and how light rays reflect of geometries. This would suggest that everything we perceive is measurable in the real world and that the brain should be able to invert physical process to achieve a successful model of our environment. A framework like this would explain our capabilities in perceiving gloss or color. The problem with this theory is that the visual information provided to our brain is insufficient; it is not able to provide adequate information of our surroundings. Chadwick and Kentridge 2015 discusses this in more detail in the context of gloss perception. Inverse optics provides

an in theory working framework, but it is unknown how the actual computations that process the retinal image would look like.

Another question is how far we should take the theory of inverse optics. In the context of liquids it is safe to assume that the complex physics is not reverse engineered by our visual system to extract properties such as viscosity. Does this theory only concern the physics of optics? What kind of systems are in play to estimate properties of deformable materials? This raises another question: do we represent a property as viscosity as an exact measure or is it a more associative concept, e.g. that liquid is as runny as water (with which I have much experience).

1.3.2 Natural statistics

Natural statistics provide a different perspective on how we perceive materials. It utilizes statistical regularities, heuristics, of natural environments (e.g., light is coming from above, vegetation grows towards light). These statistical regularities are all around us and this reflects nicely in real-world illumination (Dror et al., 2004). Fleming et al. 2003 showed that observers are better in estimating reflectance properties under natural illumination than with artificial illumination. When a specific image property occurs regularly with glossy surfaces and these regularities tend to appear across contexts as well, it becomes diagnostic of glossy surfaces. One of these regularities is that images with positively skewed luminance histograms are looking glossier (Motoyoshi et al., 2007). This tends to happen because of highlights which are often associated with glossy surfaces and produce skewed luminance histograms. However, it is possible to create similar skewed histograms by combinations of different shapes, illumination and surface reflectance properties. Therefore the image space in which we successfully can apply statistical regularities becomes more restricted (Anderson and Kim, 2009; Olkkonen and Brainard, 2010).

There are many regularities in the natural world we could utilize. The problem with natural statistics is that in most cases it is not descriptive enough and too many exceptions or false positives occur.

1.3.3 Naïve/intuitive physics

The physics behind liquids are complicated and therefore it is unlikely we use full physical representations of liquids. However, this doesn't mean that there are no physical computations used at all. One approach is that we use simplified physics models; so called naïve or intuitive physics models. These models have a trade-off between accuracy and

simplification of physical laws. Using simplified physics engines would enable us to use mental models that are able to simulate mechanical scenarios. Such mental models would explain our abilities in physical and mechanical reasoning (Hegarty, 2004; Gentner and Stevens, 2014). For example a scene where a wooden block is balancing on the edge of a table, we might mentally assign mass, a centre of mass, include effects of gravity and friction to predict if it will fall or not. Mental simulations like these would certainly explain our unnatural Jenga and dodgeball dodging skills. There is a growing body of research that plays with the idea of simplified intuitive physics models where we are able to perform fast mental simulations and make judgements based on probabilistic outcomes (Spelke, 1994; Nusseck et al., 2007; Hespos et al., 2009; Hespos and van Marle, 2012; Battaglia et al., 2013; Bates et al., 2015; Rips and Hespos, 2015; Hamrick et al., 2016). The comparison of running 1000 fast but more inaccurate simulations and make decisions based on probabilities or run one very accurate but computationally intensive simulation. Again the trade-off between accuracy and computational cost.

It is interesting to find out how much we actually rely on physics to be able to interact with liquids in everyday tasks. We are able to visualize future states of liquids, how a glass with water pours over, how a clump of honey will slowly ooze into a puddle; these examples already speak to our imagination. To be able to visualize this strongly suggests we are able to mentally simulate these scenarios. It is another source of information next to optical and mechanical cues that we can utilize but it seems that predicting future states is a less common higher-level process than estimating viscosity or sliminess.

1.3.4 Visual processing hierarchy

The literature often refers to terms as low-, mid-, and high-level vision. Different levels suggest a hierarchy where in this case each level represents different types of information and processes. Low-level features are computed early in the visual processing hierarchy and apply local computations such as spatial filtering and normalization. These filters represent image structures such as luminance edges and colour gradients and are descriptive of the image. For example a filter being activated by edges under an angle of 135 degrees.

Mid-level features concentrate on pooling and grouping —perceptual organization— it zooms out making concepts more complex addressing non-local regions. Mid-level features are placed between image and object representations, representing concepts as elongated and blobby or opacity and reflectance. These representations are build upon local low-level activations which are grouped together in a systematic way. Mid-level features are important for the representation of surfaces and materials (Adelson, 2000;

Anderson, 2011; Marlow et al., 2012; Paulun et al., 2015; Fleming, 2017). In the case of liquids the mid-level feature space is descriptive of both optical and mechanical cues. This space needs to be navigated to come to accurate materials estimates. Liquids as water or fruit juice won't clump together like maple syrup; by perceiving clumping we can discard a large portion of possible material estimates. By a process of elimination the feature space is reduced until an accurate estimation can be made.

High-level features are in the last stages of visual processing. Here we move from non-local regions to scene content. These concepts feature constancy across size, lighting, viewpoint, and occlusions. This layer provides an interface between cognitive process, language and memory. These interactions are important for material categorization and object recognition. Other examples are intuitive physics simulations which concern scene interactions and are therefore most likely driven by high-level features.

A hierarchical system that starts with low-level filters insinuates information being processed in a bottom-up manner. Information flowing from low-level filters to high-level scene descriptors. However, material estimations are driven by associations and priors as well, which would require top-down input. These top-down feeds could help to steer more efficiently through feature spaces based on prior knowledge and associations. Of this recurrence that propagates through this visual hierarchy little is understood. As the domains in which we study material perception become increasingly more complex we will have to take both bottom-up and top-down information into account. With bottom-up feeds we would need to separate effects of individual cues and top-down feeds would require to limit the influence of prior knowledge. This is practically impossible with human observers. For that reason neural networks have become very popular in vision research since these networks represent very similar hierarchical stuctures. Then we have control over the prior knowledge, the data it is trained on. This provides new insights on how these processes are interacting and contributing to specific percepts.

1.4 Overview

With this background it should have become clear what we stand to gain by studying the perception of liquids. Liquids provide, because of their physical complexity, a challenging case where we must navigate through a large high dimensional perceptual space. We are able to do this and the following chapters will demonstrate to certain extend how. Chapter 2 is specifically looking at the balance between optical and mechanical cues used to perceive liquid properties such as sliminess and runniness. Chapter 3 is a shorter chapter where we try to get a better sense how well and constant we actually are in estimating

viscosity across contexts. Chapter 4 is the largest study where we deterimine if a reduced feature space can explain viscosity constancy across contexts. Chapter 5 presents results of an on ongoing study where a neural network is trained to perceive viscosity. Finally chapter 6 will present the conclusions.

Study 1: Influence of optical material appearance

In this study we specifically look at the contributions of optical cues while estimating a range of liquid properties. Using the same liquid shapes but with different optical appearances we studied which properties (e.g., sliminess, runniness) are to what extend influenced by optical or mechanical cues. The semantic labels we assign to liquids are studied as well to see how we identify and name liquids.

Highlights:

- We are very good at perceiving viscosity
- Optical properties have very limited influence on shape- and motion-based viscosity judgements
- Both optical and mechanical cues separately influence the perception of liquid properties
- We mostly use optical cues to name and identify liquids

Study 2: Viscosity constancy across contexts

We can encounter liquids in many different states and contexts. Here we specifically look at the constancy of viscosity perception despite radical changes in shape. How consistently do we actually perceive shape?

Highlights:

- We are very good at matching viscosity
- Within scene viscosity constancy across noise pertubations is 99%
- Viscosity constancy across scene variations is 95%
- Global shape and motion changes can’t explain viscosity constancy

Study 3: Visual features of liquids

We used the visual perception of flowing liquids to uncover the computations underlying visual inferences about materials. By comparing observers' viscosity ratings with perceived shape features, we show how the brain exploits 3D shape and motion cues to infer viscosity across contexts despite dramatic image changes.

Highlights:

- Observers are remarkably good at visually inferring the viscosity of flowing fluids
- They use multiple midlevel shape and motion features to do so
- Four factors predict perceived viscosity constancy surprisingly well
- The features take wildly divergent stimuli and organize them by viscosity

Study 4: Estimating viscosity with neural networks

Machine learning is an interesting field of research that had major breakthroughs in recent years because of convolutional neural networks. Here we trained a neural network to perceive viscosity. The network is designed to mimic human performance while using phyical viscosity labels for training. Performance is very good and we demonstrate that this network exploits similar cues as the human observers.

Highlights:

- With this dataset particular observers find it hard to rate viscosity accurately
- In both static and moving stimuli conditions the DNN explains human performance very well
- The DNNs make smaller error estimations to the human mean than individual obsevers
- Optical flow can explain the increase in performance for the moving stimuli

Chapter 2

Influence of optical material properties

A similar version of this chapter has been published as:

van Assen, J. J. R., & Fleming, R. W. (2016). Influence of optical material properties on the perception of liquids. *Journal of vision, 16*(15), 12-12.

In everyday life we encounter a wide range of liquids (e.g., water, custard, toothpaste) with distinctive optical appearances and viscosities. Optical properties (e.g., colour, translucency) are physically independent of viscosity, but, based on experience with real liquids, we may associate specific appearances (e.g. water, caramel) with certain viscosities. Conversely, the visual system may discount optical properties, enabling 'viscosity constancy' based primarily on the liquid's shape and motion. We investigated whether optical characteristics affect the perception of viscosity and other properties of liquids. We simulated pouring liquids with viscosities ranging from water to molten glass and rendered them with nine different optical characteristics. In Experiment 1, observers (1) adjusted a match stimulus until it had the same perceived viscosity as a test stimulus with different optical properties, and (2) rated six physical properties of the test stimuli (runniness, shininess, sliminess, stickiness, warmth and wetness). We tested both moving and static stimuli. In Experiment 2, observers had to associate names with every liquid in the stimulus set. We find that observers' viscosity matches correlated strongly with the true viscosities and that optical properties had almost no effect. However, some ratings of liquid properties did show substantial interactions between viscosity and optical properties. Observers associate liquid names primarily with optical cues, although some materials are associated with a specific viscosity or combination of viscosity and optics. These results suggest viscosity is inferred primarily from shape and motion cues but that optical characteristics influence recognition of specific liquids and inference of other physical properties.

2.1 Introduction

In everyday life we continuously interact with our environment and the objects and materials it contains. To be able to do this effectively we need to be able to recognize familiar objects and materials, and infer their physical properties by sight. This is essential to our survival: it allows us to avoid eating rotting food; breaking our ankle on a slippery curb; or burning our hand on a hot pan. One highly challenging class of materials are liquids and gels. It is quite impressive that under typical conditions we can visually infer the properties of liquids and interact with them effectively, despite their erratic nature and the large influence that external forces hold over their shape and flow. We are very well able to distinguish between water, toothpaste, caramel, shampoo, mercury, and numerous other liquids, and can even infer properties such as runniness, sliminess and stickiness without physically touching them. This is important as it allows us to determine their affordances (i.e., whether it can be used for drinking, cleaning, gluing, etc.) and predict their likely behaviour before interacting with them.

Here, we sought to investigate the role of specific visual cues in the perception of liquids and their properties. In principle, there are several distinct sources of information that observers could draw on to recognize liquids and infer their physical characteristics by sight. Broadly, we can divide these into two classes: optical and mechanical. The main purpose of this study was to determine the relative contributions—and interactions between—these two broad classes of information. Some studies approach material perception by asking how the visual system estimates a single physical property of materials (e.g., glossiness, elasticity), and seeking specific visual cues to that property. In this study, by contrast, we look at a wide range of liquid properties to identify whether there are any stimulus or task conditions in which optical and mechanical cues interact to affect the perception of liquids.

A liquid's optical material appearance can tells us many things about the liquid. For example, water is colourless and transparent, while milk is translucent; caramel and chocolate-sauce have distinctive colours, whereas molten solder is lustrous. Because specific optical characteristics are associated with particular liquids, we could use the optical appearance—or low-level image correlates—to narrow down the range of expected behaviors of the liquid. In addition to the large literature on the perception of surface colour (see Foster, 2011, for a review), a growing body of research has investigated the estimation of optical properties such as gloss (Beck and Prazdny, 1981; Nishida and Shinya, 1998; Fleming et al., 2003; Motoyoshi et al., 2007; Ho et al., 2008; Kim et al., 2012, see Chadwick and Kentridge, 2015 for a recent review), translucency (Fleming et al., 2004;

Fleming and Bülthoff, 2005; Xiao et al., 2014), transparency (Fleming et al., 2011; Faul and Ekroll, 2012; Schlüter and Faul, 2014) and surface texture (Landy and Graham, 2004; Dong and Chantler, 2005; Emrith et al., 2010; Liu et al., 2015). These findings suggest that human observers are generally very good at inferring optical material properties under a wide range of conditions, and thus it is plausible that observers could base judgments about liquids on such cues.

In contrast, it is the mechanical properties of liquids that determine the way they move and adopt particular shapes in response to external forces. Probably the most important mechanical parameter distinguishing different liquids and gels is viscosity. For example, water is very runny and therefore prone to splash and spread out in puddles, whereas, toothpaste is thick and therefore tends to pile up into clumps when poured. Thus, the visual system could use the distinctive shape and motion caused by different viscosities to recognize liquids and predict their behaviours. Previous research has shown that we can infer viscosity both from shape (Paulun et al., 2015) and motion cues (Kawabe et al., 2015). Thus, again, it is plausible that human judgments about liquids could rely on their mechanical properties.

In this study we sought to determine the relative contributions of optical and mechanical cues to the perception of liquids and their properties. We ask the following questions: Do observers recognize specific liquids based primarily on optical properties—like colour, gloss or translucency—or is viscosity also important for determining a liquid's identity? Are judgments of viscosity biased by a liquid's optical properties? What about the perception of other properties—like temperature, or stickiness—which cannot be so easily inferred from the motion or shape of the liquid? Such properties are potentially extremely important for determining the affordances of materials, but little is known about whether participants can infer them through visual information.

A given material can change both its optical and mechanical properties depending on the prevailing conditions: for example the sugar concentration or temperature of syrup affects its viscosity, while small concentrations of dirt can make water cloudy without affecting the way it flows or splashes. Thus, both sources of information are imperfect cues to material identity. While it is commonly argued that shape dominates other cues in object recognition (Biederman, 1987; Landau et al., 1988), liquids are highly mutable, so it is plausible that colour and other optical characteristics might be more diagnostic than shape. At the same time, if shape and motion can be computed accurately across a wide range of different optical conditions (Todd et al., 1997; Todd, 2004; Nefs et al., 2006; Khang et al., 2007; Vangorp et al., 2007; Doerschner et al., 2013; Dövencioğlu et al., 2015), then viscosity could be estimated in a way that is unaffected by the surface material

appearance, enabling 'viscosity constancy'. Thus, there are grounds for believing that optical and mechanical properties may contribute to different extents depending on the specific judgments that observers are asked to make: whether it is estimating viscosity; rating other properties of liquids; or identifying (e.g., naming) specific materials, like paint, toothpaste or molasses. To test the contributions of optical and mechanical properties in the perception of liquids and their properties, we therefore asked participants to perform three tasks: (1) viscosity matching; (2) subjective rating of liquid properties and (3) identifying which liquids correspond to verbal labels.

For these experiments we used physically-based computer simulations of a wide range of liquids. The viscosities ranged from water to molten glass in six approximately perceptually uniform steps (established in an unpublished pilot experiment with the same stimuli using maximum likelihood difference scaling). Each liquid was rendered with nine different optical characteristics. Although the computer simulations are not absolutely perfect (careful observation reveals a few visible artifacts) they are accurate enough to elicit vivid and compelling impressions of distinct liquids, and were computed at higher resolutions than used in previous studies on the perception of liquids (Paulun et al., 2015; Kawabe et al., 2015). Moreover, only by using computer simulations is it possible to vary mechanical and optical properties independently in a parametric and perfectly controlled way. Only computer graphics allows us to render identical 3D shapes with different optical properties, enabling us to perfectly isolate the relative contributions of the two classes of cue.

In the experiments observers were asked to adjust the viscosity of a match stimulus along a high-resolution viscosity scale (64 steps) until it appeared to have the same physical properties as a test stimulus that had different optical properties, in an asymmetric matching task. Observers also rated six different properties of the test stimuli, (runniness, shininess, sliminess, stickiness, warmth and wetness). These two tasks were performed with both static and animated stimuli. Finally observers participated in a liquid naming experiment to see how optical or mechanical properties interact to determine the identity of familiar liquids such as chocolate sauce, mouthwash or milk.

2.2 Methods

In the first experiment observers were asked to perform two tasks on each trial: an asymmetric viscosity-matching task, followed by a liquid property-rating task. The matching task showed a test stimulus with a specific viscosity and optical appearance, and observers could scroll through a standard set of liquids with fixed optical appearance, but finely vary-

ing viscosities, to select another stimulus that had the same apparent viscosity as the test. The test and match stimuli were sampled from different points in time in the animation sequence to encourage observers to base their responses on an internal representation of the physical properties of the liquid, rather than simply by identifying the stimulus with identical shape. Following the matching task, participants were asked to perform a series of ratings in which the same test stimulus was presented together with rating sliders for six different liquid properties: runniness, shininess, sliminess, stickiness, warmth and wetness.

Across participants, we varied (1) whether the stimuli were single static frames or 1-sec animation sequences and (2) whether the test or match stimuli were taken from the earlier time point.

In a second set of experiments we measured how participants assigned names to the stimuli based on their mechanical and optical properties. First, one group of observers were presented with all 54 stimuli (6 viscosities $\times$ 9 optical appearances) and were asked to provide names for each material. Then, a second group of subjects filtered the word list to select the most descriptive and plausible liquid names corresponding to the stimuli. Finally, a third group of participants were provided with each name in the list and were asked to identify all of the stimuli from the 6 $\times$ 9 array that fitted the description. The observers were allowed to select multiple liquids, allowing us to measure the extent to which each verbal term designated a mechanical or optical appearance (or both).

2.2.1 Stimuli

All stimuli used in this study can be downloaded here:
http://doi.org/10.5281/zenodo.154570

Simulation

The stimuli were generated using RealFlow 2014 (V. 8.1.2.0192; NextLimit Technologies, Madrid, Spain). This software enabled us to simulate and render liquids up to the standards used by the visual FX industry. We used the “Hybrido” particle solver, which makes it possible to specify the dynamic viscosity of the liquids in real physical units (Pa·s). Hybrido is a FLIP (Fluid-Implicit Particle) solver using a hybrid grid and particle technique to compute a numerical solution to the Navier-Stokes equations describing viscous fluid flow. All information for the fluid simulation is carried by discrete particles, but the solution to the equations is carried out on a grid. Once the grid solve is complete, the particles gather the information required from the grid to move forward in time to the next frame. The fluid

boundary is then derived from the position of the particles by a meshing algorithm (when visible artifacts occur, it is primarily due this step of the algorithm, not the underlying physics solver). For the match stimuli, a set of 64 different viscosities was simulated with logarithmically evenly placed steps from 0.001 Pa·s to 100 Pa·s (roughly corresponding to a range from water to molten glass in approximately perceptually uniform steps). The following equation can calculate the step number back to viscosity:

$$\mu = b \cdot (10^{\frac{d}{s-1}})^{n-1} \tag{2.1}$$

Where μ is the viscosity in Pa·s, *b* the starting value of the scale, in our case 0.001, *d* the range of the scale in decades, in our case 5 (10^{-3} to 10^{2}), *s* the amount of steps of the scale (64), and *n* the step number of which we want to calculate the viscosity. Liquid density was held constant at one kilogram per liter. The number of particles used varied between 2 and 4.5 million particles depending on the viscosity, the only changing parameter in the simulation.

The simulated scene (see Figure 2.1) consisted of a m^2 plane with a shallow wall around its perimeter and an irregularly shaped solid object (height = 17.5 cm, diameter = 19 cm) that was rigidly attached to the centre of the plane. The liquid emerged from an 'emitter', located approximately 30 cm above the object (outside the frame of view). Gravity was the only external force acting on the liquid, which had no initial velocity on emerging from the emitter. The orifice of the emitter had a rounded cross shape, yielding distinctive ridges in the shape of the liquid, whose durability and distinctness varied with viscosity.

The simulated animations had a total duration of ten seconds (300 frames at 30fps). For the experiments using static stimuli, the test and match images consisted of frames 90 and 150 from the animation (i.e. a 2-second time difference) in the first condition, and

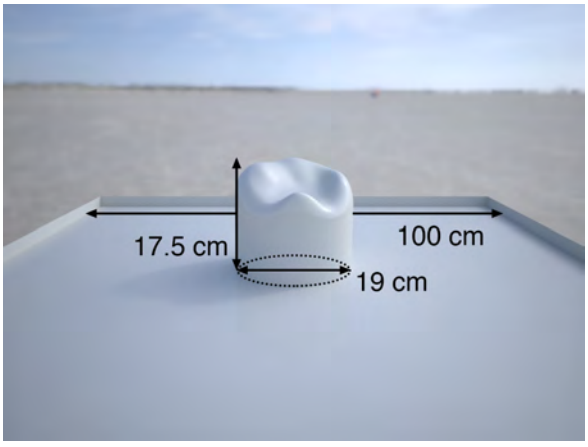


Figure 2.1: The dimensions of the simulated scene.

150 and 90 in the second condition. The duration of the moving stimuli was one second: frames 80-110 or frames 140-170.

For the target stimuli, six different viscosities were selected which were evenly spaced on the existing 64-step scale. In this case steps 10, 19, 28, 37, 46, 55 corresponding to dynamic viscosity values of 0.005, 0.027, 0.139, 0.72, 3.73, 19.3 Pa·s. Figure 2.2 shows an overview of the static test stimuli. Video A.1 shows the full ten second animations of the six different viscosities with the same optical material.

Rendering

The render engine used to generate the final image frames was Maxwell (V. 3.0.1.3; NextLimit Technologies, Madrid, Spain). Nine different optical materials were developed with diverse appearances, varying in their opaque, transparent and translucent properties. The match stimulus set (consisting of 64 viscosities) was rendered with a translucent 'green goo' appearance. The test stimuli consisted of approximations of the following materials: caramel, metallic car paint, chocolate, copper, a matte blue material, milk, water and wine. These materials were selected to represent a wide range of different appearances that we could encounter in liquid form, including both common (e.g. colourless transparent) and unusual (e.g. matte blue) appearances. Video A.2 shows a loop of the one-second animations used during the experiment. It shows the nine different optical materials with the same viscosity.

The images were rendered at an 800×600 resolution and the scene was lighted using an HDR light probe depicting a beach scene (from the Maxwell Resource Library by Dosch Design).

2.2.2 Observers

Matching and rating tasks

48 observers took part in the first experiment with static and animated stimuli and the two temporal orderings of test and match (i.e., four groups, with twelve observers per condition). The average observer age was 25.3 (SD = 4.45). 33 observers were female and 15 male.

Naming experiments

42 German speakers participated in the three experiments to match names with liquids. Ten observers took part in the free-naming ('brainstorming') session with static stimuli,

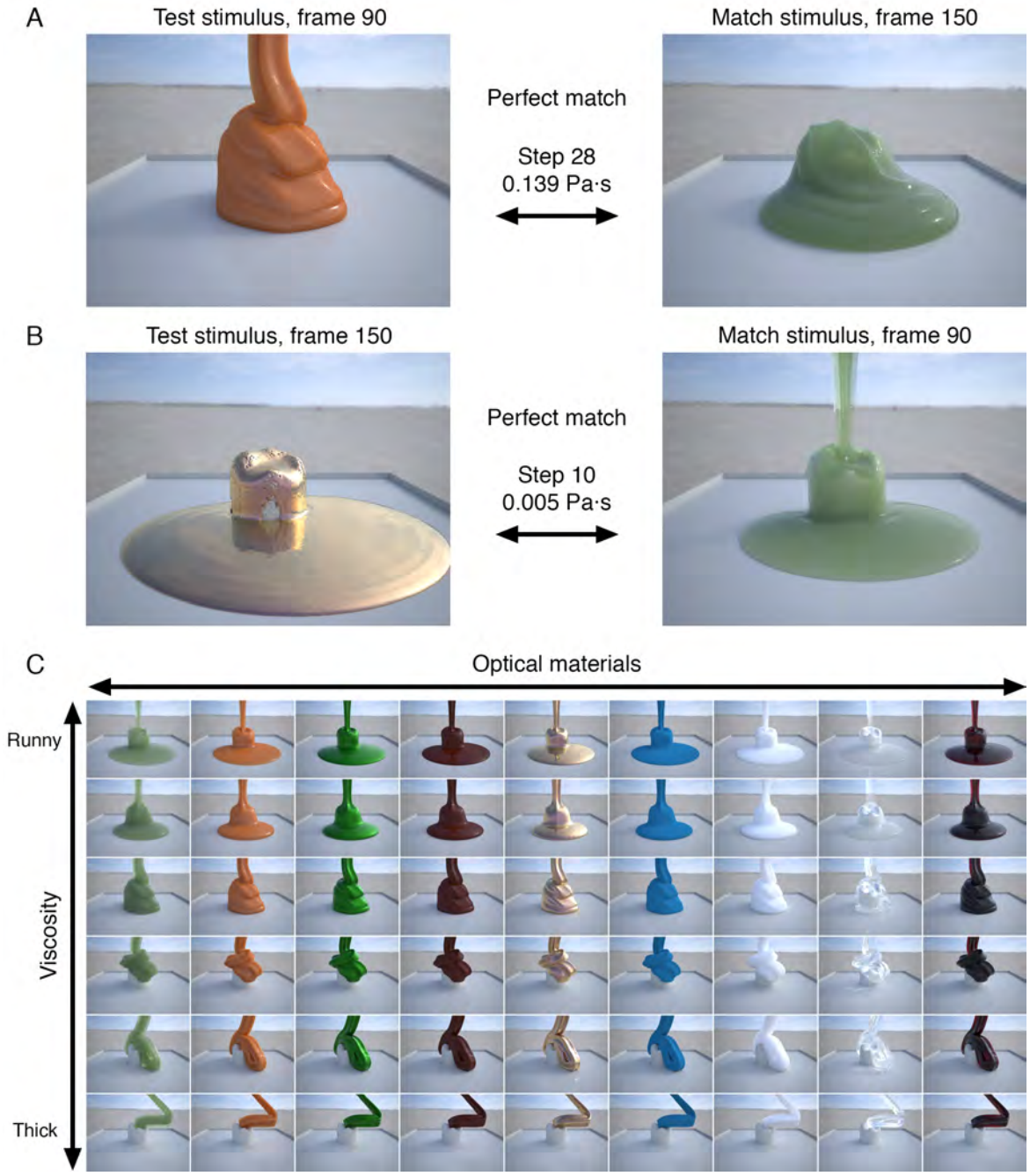


Figure 2.2: (A) An example trial with the physically correct match stimulus. (B) Another trial with inverted time points. (C) An overview of static stimuli with the nine different optical materials in the x-axis and the six different viscosities on the y-axis. The optical materials are approximations of the following materials: green goo, caramel, metallic car paint, chocolate, copper, a matte blue material, milk, water, wine.

and a further ten observers with animated stimuli. Six other observers took part in the 'filtering' session to select a sub-set of terms from the brainstorming sessions. Finally, 16 observers participated in the main experiment, in which participants identified which stimuli corresponded to each verbal item. The average age was 24.9 (SD = 4.08), 27 were female and 15 male.

All observers gave written consent prior to the experiment and were paid for participating. All observers reported having normal or corrected-to-normal vision.

2.2.3 Procedure

Matching and rating tasks

All experiments were performed in accordance with the declaration of Helsinki, and prior approval was obtained from the local ethics committee of the University of Giessen. The experiments were performed on an Apple Mac Mini with a Dell U2412M 24-inch monitor using factory default settings, gamma of 2.2 and a resolution of 1920 x 1200 pixels. Matlab 2015a (v. 8.5.0.197613) and the Psychtoolbox library (v. 3.0.11; Brainard, 1997; Pelli, 1997) were used to run the experiment, although Psychtoolbox was upgraded to (v. 3.0.12) over the period when different observers participated.

Observers completed a short training session before starting the experiment. This consisted of a single trial to familiarize the participant with the interface for the matching and rating tasks and to ensure that the concepts on the six rating scales (liquidness, shininess, sliminess, stickiness, temperature and wetness) were clearly understood. Each trial consisted of the matching task followed by all six ratings for a given test stimulus. For the viscosity-matching task, the test stimulus was presented on the left hand side of the screen, and the match stimulus was presented simultaneously on the right hand side of the screen. Observers had to scroll through the viscosities of the match stimulus, with the left and right arrow key on the keyboard. A 'page turning' animation occurred with every button press, revealing the new match stimulus, to avoid apparent motion between the different stimuli. Once the match stimulus on the right appeared to have the same physical properties as the target stimulus on the left, the observer could confirm by pressing the 'space' bar to proceed to the rating task for the same target stimulus. Here, the observer had to indicate their subjective rating for each of the six properties by using the mouse to move the randomly placed dots along the continuous rating bars (with seven tick marks). When the observer interacted with the dot on the rating bar, the dot would turn green. When all six dots were green the observer could continue with the next trial by pressing 'space'. The observer had to complete a total of 108 trials (2 blocks, each consisting of 9

materials $\times$ 6 viscosities in random order). There were no time limits, and the experiment took observers 45 to 90 minutes to finish.

Naming experiments

For the naming experiments the same Apple Mac Mini was used with the same Dell U2412M monitor as in the other experiments. The 'brainstorming' experiment also used Matlab and the Psychtoolbox library. On each trial, one of the 54 test stimuli (9 optical materials $\times$ 6 viscosities) was presented and observers were instructed to "name the liquid you see in the image". There were four empty lines where observers could enter names for the liquids. Only one response per stimulus was required, although subjects were encouraged to provide multiple verbal terms if they applied. The brainstorming session resulted in a combined word list of 2156 entries, 1262 for the static stimuli and 894 for the moving stimuli. From this list, ten names for each stimulus were selected, removing many duplicate and less descriptive entries. The resulting list of 540 words was used for the 'filtering' experiment.

Different software was used for the 'filtering' and 'name matching' experiments because of better interfacing possibilities. In this case a Flask (v 0.10.1) based framework was used compiled with Python 2.7.1. The front end was written using HTML5 technology displayed in Safari (v. 7.1.7). These browser-based experiments were displayed in 'presentation mode' and therefore showed no interface of the browser itself. On each trial in the filtering experiment, an animated liquid stimulus was presented along with a randomized list of ten names generated for that stimulus in the previous brainstorming session, 54 lists in total. Observers were asked to order the three most appropriate and descriptive names to the top of the list. This top three was weighted accordingly (3 points 1st choice, 2 points 2nd choice and 1 point 3rd choice) during the selection process. All scores above 90% of the highest score were selected from the list. This means that if there was a close second both words were selected, which happened eleven out of 54 times. Duplicate answers were filtered out, resulting in 49 words for the main name matching experiment.

The name matching experiment used the same Flask and browser based presentation system as the filtering experiment. A new set of observers performed the task. On each trial, they were presented with a liquid name and a 6 $\times$ 9 grid containing static thumbnails of all stimuli. The viscosities were ordered vertically and the optical materials horizontally. When the observer dragged the mouse over a stimulus in the stimuli grid, a full-size animation for the corresponding stimulus would appear. If the observer thought that a given stimulus corresponded to the verbal item for the current trial, they could select it with a simple checkbox (subsequent unchecking was also possible but was rarely used in

practice). Multiple stimuli could be selected for each name (i.e., each trial) but only one answer was required. Finally the observers were asked to give a confidence rating for their response before continuing to the next trial. This experiment had 49 trials in which the names from the list were linked to the 54 different stimuli. There was no requirement for all of the stimuli to receive a name.

All experiments were performed in German and have been translated to English for presentation here.

2.3 Results

Raw data from all experiments can be downloaded here:
http://doi.org/10.5281/zenodo.154570
For each of the matching and rating tasks, we tested four different versions: the static and animated stimuli with the test stimulus from an earlier or later time point in the animation sequence than the match. (see Methods for details.)

2.3.1 Viscosity-matching task

Figure 2.3 shows the results from the viscosity-matching task for the four different conditions. The first notable observation is that observers are generally very good at matching viscosity: for all optical materials, the matching function is approximately linear with slope close to one. A linear regression can explain the data extremely well with a slope close to one for static stimuli: with the match from later than the test $\beta = 0.91$, $R^2 = 0.98$, $p < 0.001$ and for the reversed time points $\beta = 1.02$, $R^2 = 0.97$, $p < 0.001$. Especially for the moving stimuli, observers matched the liquids close to perfectly for the entire tested viscosity range $\beta = 1.002$, $R^2 = 0.99$, $p < 0.001$ and for the reversed time points $\beta = 0.92$, $R^2 = 0.99$, $p < 0.001$.

There is however, a systematic additive bias in the responses, which is most pronounced for the static stimuli. For the non-reversed condition (i.e. match stimulus from a later time point in the animation than the test stimulus) stimuli there is a slight overestimation of viscosity. In other words, the liquids were perceived as having the same viscosity when the match stimulus was thicker than the test stimulus. This presumably reflects an imperfect compensation for the time offset between test and match, rather than a systematic overestimation of viscosity. This interpretation is supported by the observation that when the time points for test and match are swapped (i.e., test stimulus from a later time point than the match stimulus) the bias inverts. Evidently, in the absence of strong

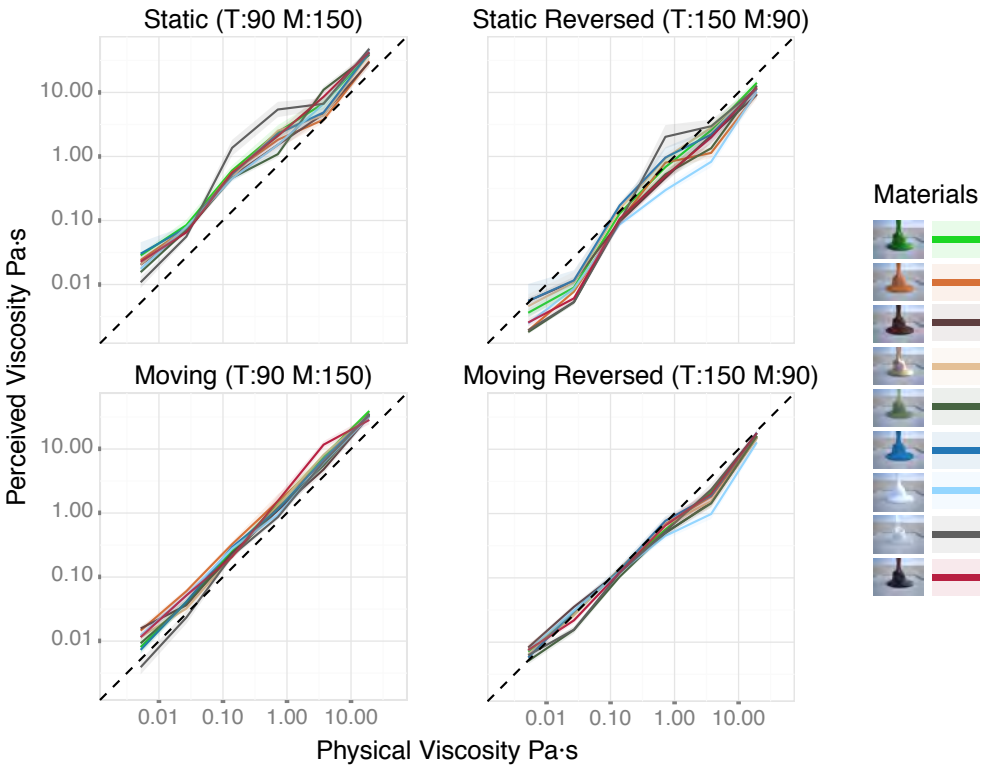


Figure 2.3: Mean results of the matching task for the four different conditions, with static and moving stimuli, and reversed time points between match (M) and test (T) stimulus. Error envelopes represent standard error of the mean. Time points for the Moving conditions refer to a range of 30 frames, in this case frames 80-110 and 140-170.

visual cues to indicate the precise point in time, it is difficult for observers to compensate for the difference in time point between test and match. Put differently, when asked to match viscosity in this task, there is a bias towards selecting similar shapes. This tends to lead to errors because the shape of runnier liquids evolves more rapidly than for thicker liquids. Thus, the shape adopted by a given material at a particular point in time is often somewhat better approximated by a runnier fluid at an earlier point in time or a thicker fluid at a later point in time.

To test more rigorously the hypothesis that participants simply selected the most similar shape, we (i) ran a control experiment and (ii) developed a simple image similarity metric based on the Euclidean distance. The control experiment was exactly the same as the asymmetric matching task in the main experiment, except that instead of matching viscosity twelve new observers were instructed to match shape. The match was only performed with static stimuli and all stimuli were of the green goo material. The Euclidean similarity matric used grayscale versions of the match and test stimuli with the same optical material, which were subtracted from each other. The mean pixel value of the

resulting image is compared with other match/test stimuli combinations where the lowest mean value is the best Euclidean match. This allows us to derive a predicted match for each test stimulus, by identifying which of all the match images has the smallest Euclidean error (difference) to each test image. Figure 2.4 plots these predictions in comparison with observers' data for the static stimuli. Both results further support the interpretation that the additive bias is due to observers tending to match shape, while only partially compensating for differences in time point. The Euclidean predictions for runny liquids diverge more because of faster evolving shapes resulting in bigger differences between the two time points. Observers seem to partially compensate for this by not picking the most similar shape (in Euclidean terms). The hypothesis is further supported by the shape matches made in the control experiment. Performance was practically identical when observers were asked to match based on shape rather than viscosity, suggesting that viscosity judgements are very similar to shape similarity judgements. Our interpretation of this finding is that it is not very helpful to think of 'viscosity perception' as a fixed process of creating a single, unified internal estimate of the physical parameter of the liquid, which can then be accessed psychophysically. Instead, depending on the specific task (e.g. matching viscosity, rating runniness) and stimulus context (i.e. other stimuli in the experiment), participants latch onto different cues in a highly flexible way (here focussing mainly on shape similarity between test and match stimuli).

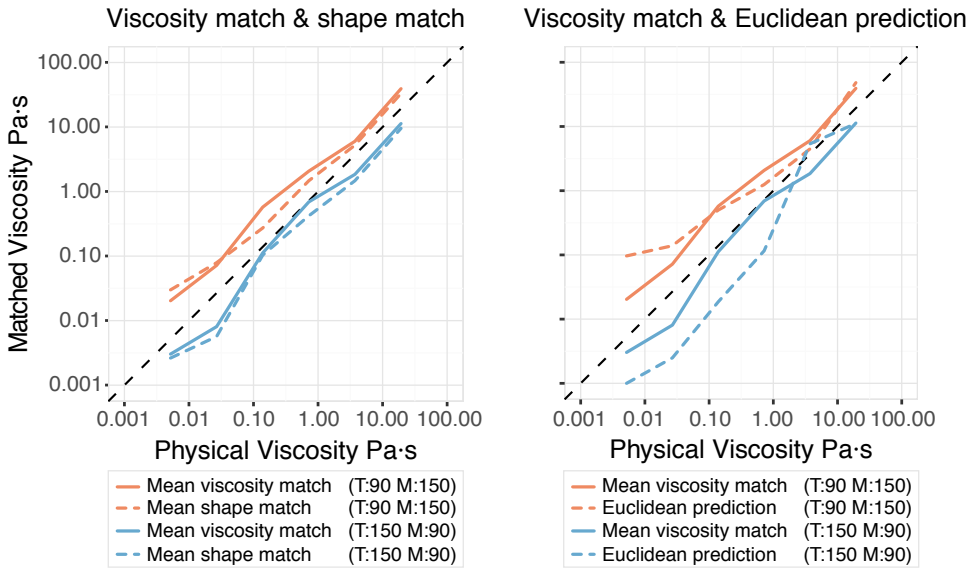


Figure 2.4: Mean viscosity match over all nine materials for both time conditions with static stimuli in comparison with the shape match task (left) and predictions based on match stimulus with the most similar shape to each given test (using a simple Euclidean shape metric; right). Note that the solid data series depict identical data in both panels; only the predictions (dashed lines) differ.

Returning to the results of the main experiment, another notable aspect is the negligible differences between the nine optical materials, especially for the moving stimuli. This is confirmed by the linear regression analysis performed earlier where linear models based on viscosity explain 97-98% of the variance for the static stimuli. This means that at most 2-3% of the variance can be accounted for by the optical material differences and noise. With moving stimuli this is even down to 1%. Thus, optical material appearance barely influences viscosity judgments, i.e., observers have very good viscosity invariance across changes in optical appearance, at least when reliable motion and/or shape cues are present.

2.3.2 Rating liquid properties

Observers were asked to rate runniness, shininess, sliminess, stickiness, wetness and warmth. Figure 2.5 shows the scores observers gave for each material at the six different test viscosities. To save space only graphs from the moving stimuli variation are shown. The graphs for the other variations, which are broadly similar, can be found in the appendix.

There is a clear difference between properties that are driven mainly by mechanical cues (i.e., cues based on shape and motion), and optical cues, based on optical material appearance. As expected, runniness is clearly scored primarily on the viscosity of the stimulus and optical material appearance has almost no effect. A linear model based on the viscosity explains 98% of the variance leaving 2% unexplained by the optical material appearance and noise.

Conversely, shininess is driven primarily by optical cues. As expected, the matte blue material is seen as the least shiny, and the lustrous copper-metal as the most shiny. There is almost no effect of viscosity on perceived shininess: most materials have a certain shininess independent of their viscosity, as indicated by the flat curves. The only exception seems to be the milk-like material. We believe this effect is caused by the high degree of subsurface scattering for this material. When the material's shape is thin, there is little scattering, so the body colour appears darker, and the specular reflections have higher contrast. By contrast, when the material has more volume, scattering makes the body colour whiter, reducing the contrast of highlights (Pellacini et al., 2000). From the third viscosity step on we see a notable decline in perceived shininess, shown at point 'A' of Figure 2.5. From this viscosity on, the material gathers into thicker, more voluminous clumps, creating a more diffuse, matte appearance. Thus the interaction is probably not due to the perceived viscosity per se, but rather simply due to the shape.

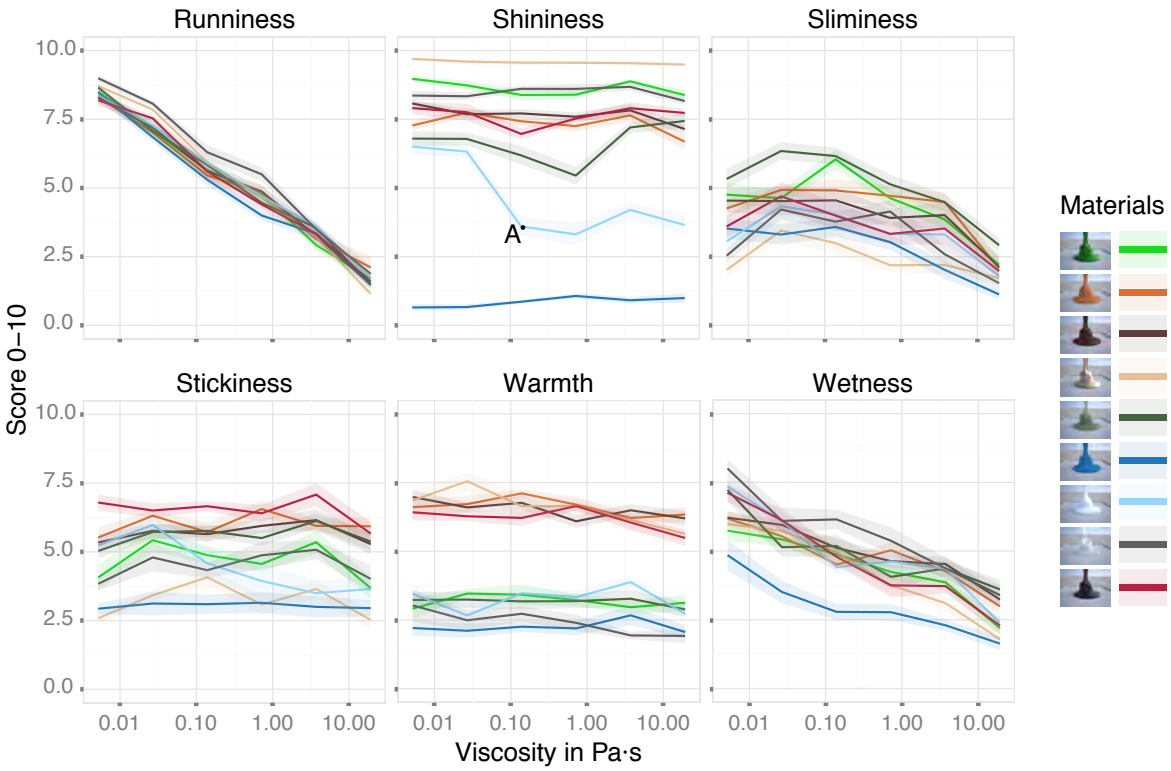


Figure 2.5: Mean rating scores for the six different liquid properties with moving stimuli. Error envelopes represent standard error of the mean. Point A (Shininess plot) indicates the point at which the liquid gathers into voluminous clumps, affecting the perceived shininess for the milk-like material.

Sliminess is a property that depends on both mechanical and optical cues. There are certain optical materials like green goo that appear slimier than others. At the same time, there is also a certain (intermediate) viscosity range that observers associate with sliminess. It is interesting to see that there appears to be little interaction between optical and mechanical cues. The different materials are shifted from each other vertically, but follow roughly the same curve.

Stickiness is mainly driven by optical cues. For example the matte blue material does not look sticky at all while a wine-like materials appears to be stickiest.

Wetness decreases with increasing viscosity. The matte blue material appears substantially less wet than all other materials. This is consistent with previous findings that specularity is associated with wetness (Sawayama and Nishida, 2015).

Somewhat surprisingly, the warmth ratings do not show a substantial effect of viscosity. One might expect a runny metal or chocolate coloured material to appear warmer than a more viscous variant. The instructions clearly stated that participants should rate the expected temperature, as it would feel were the participant to put their finger in the liquid. However, participants did not seem to consider runniness as a cue to increasing

temperature. It is possible that a forced choice paradigm might reveal a tendency to associate runnier liquids with higher temperatures, but if present, the association is not strong enough to show up in this experiment. Another notable result is that there is a clear bimodal distribution of warm and cold materials. This appears to be influenced by the 'warmth' of the colour of the liquid, where red, brown and orange materials are warm and green, blue and transparent materials are cold. It is unclear whether this was simply a tacit association, or whether participants deliberately chose to base their warmth judgments on colour, despite the explicit instructions to attend to the expected temperature.

2.3.3 Model

As noted, most differences amongst the nine optical materials appear to be shifts in scores on the y-axis. This suggests that although both optical and mechanical cues contribute to the perceived properties of liquids, the interactions between the two classes of information are generally relatively weak. We quantified this observation by fitting models to each of the nine materials for the six liquid properties shown in Figure 2.5. For each of the rated properties, we took the mean of all optical materials and fitted a linear and quadratic model to this. The best AIC score of the mean-based model defines the type of model for the individual materials. AIC or Akaike information criterion is a statistical model fit measure based on the likelihood function and number of predictors. To test the hypothesis that most of the data can be explained by only shifting a fitted model on the y-axis, we took the slope of the mean-based model and fit only the intercept ('Fixed slope model'). This we compared with a fit where each material had an independently fitted slope ('Free slope model'). The results of the average AIC values from the nine different materials are shown in Figure 2.6. A lower value means a better fit of the model to the original data. Since AIC weighs in the complexity of the model and our fixed slope models are less complex we can see if a decrease in complexity compensates for the decrease in goodness of fit. In the cases where the orange bar is shorter in Figure 2.6 our fixed slope model outperformed the free slope model. This means that in these cases, the model without interaction between optical and mechanical cues explains the data better. Another measure, AIC_c, or the second-order corrected Aikaike information criterion assigns greater penalty for extra model parameters and is mostly applied in cases when the sample size (n) is small compared to the number of parameters (k) where $n/k < 40$ (Burnham, 2002), which holds in this case. In all six cases AIC_c prefers our fixed slope model. Overall, based on these results, it is safe to say that interactions between optical and mechanical cues are relatively limited. Shininess, with the outlier at point A of Figure 2.5 seems to be the

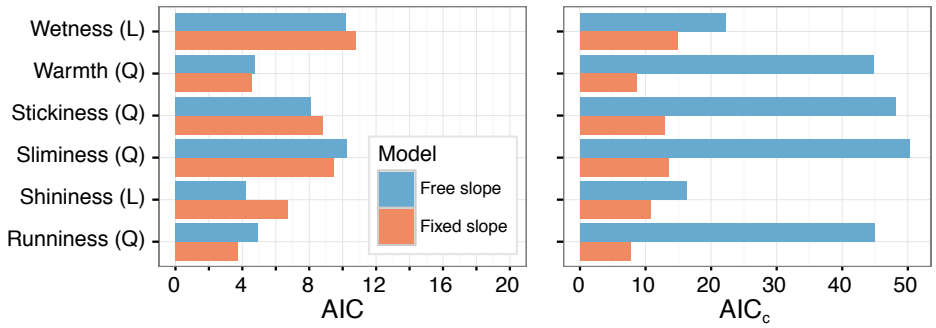


Figure 2.6: Overview of average AIC and AIC$_c$ values from the nine different materials for each liquid property. A lower value means that the model is a better fit. The free slope model fits the intercept and the slope. The fixed slope model only fits the intercept with a predetermined slope. The Q and L show if it is a linear or quadratic based model. The AIC criterion takes the amount of parameters used into account for an optimal trade-off between goodness of fit and complexity of the model. AIC$_c$ is similar to AIC but assigns greater penalty for extra model parameters.

reason why our fixed slope model doesn't perform well where in the other negative two cases the differences between the two models are much smaller.

2.3.4 PCA analysis

Another way of representing the rating data, to gain insights into the relative contributions of mechanical and optical cues, is using a principal component analysis. Each stimulus can be represented as a point in a 6D feature space, where each feature represents one of the six subjective rating scales. PCA allows us to summarize the relationships between the different stimuli as well as the relationships between the different liquid properties. Figure 2.7 plots the data from the experiment with moving stimuli with standard time ordering, in the space spanned by the first two principle components.

Caution is required in interpreting these plots as the different ratings are not necessarily measured on a consistent scale. Although participants were asked to rate each property on a 0–10 range, they may have used very different internal scales for mapping the perceived differences between different liquids onto each scale. Thus, for example, a step of 0.1 on the Runniness scale is not commensurable with a step of 0.1 on the Warmth scale. This means that we cannot draw strong conclusions about the metric distances between different samples in the PCA space. Nevertheless, it is interesting to observe the orderly arrangement of the samples in the feature-space, which are systematically organized by both optical and mechanical properties.

The different dimensions plotted in Figure 2.7A reveal that runniness and shininess are approximately perpendicular to each other. As noted above, runniness is mainly driven by

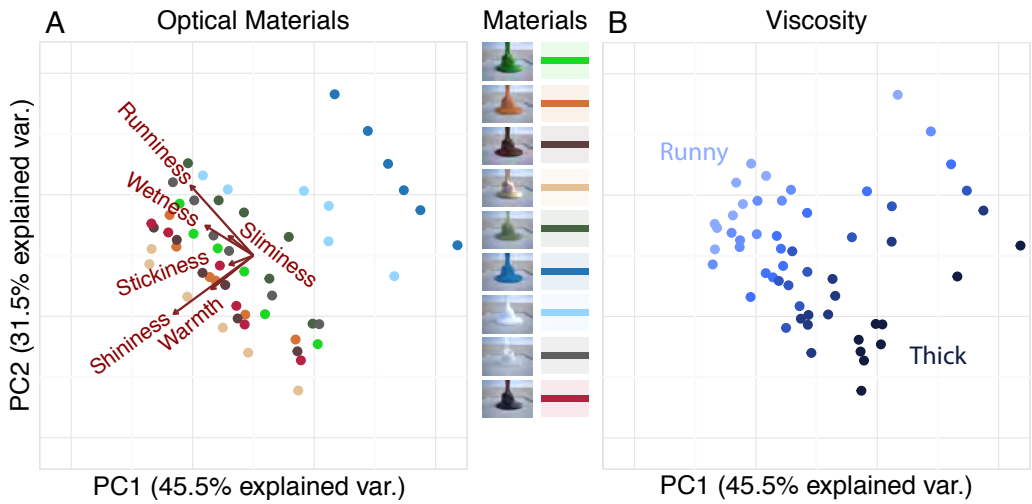


Figure 2.7: (A) Samples in the PCA space (first two components), color-coded by optical material. Vectors represent projections of the different liquid property dimensions. (B) The same data points color-coded by viscosity instead of optical properties.

mechanical cues (viscosity), and shininess mainly by optical cues (material appearance). That runniness and shininess are perpendicular to one another in the PCA space, confirms that we tend to separate optical and mechanical cues when judging liquid properties. In Figure 2.7A, the different optical materials are systematically organized along the shininess axis where Figure 2.7B shows that the different viscosities clearly follow the runniness axis. It is also notable that for the range of viscosities and optical appearances we used here, and for the particular set of liquid properties we asked participants to rate, optical and mechanical cues play approximately equal roles. The spread of samples in terms of their optical properties is roughly the same as the spread in terms of the viscosities (although we cannot directly compare magnitudes across features, it is nevertheless interesting that across all features there is a roughly even spread of influence of optical and mechanical properties).

2.3.5 Naming experiment

Figure 2.8 shows the results of the name-matching task in which observers were asked to select one or more stimuli for each of the 49 different liquid names that were generated in the 'brainstorming' and 'filtering' experiments.

For every participant and for each verbal item, we have a complete 6×9 binary array indicating whether the corresponding image was deemed to match the verbal item, along with a scalar confidence rating. Pooling across subjects gives us an integer array per verbal item, containing the number of votes each stimulus received across observers

(Figure 2.8B). For display purposes, we can reorder the array into a 54-vector for each verbal item. Example response vectors for a several liquid names are shown in Figure 2.8A (a complete list is presented in the appendix).

For most stimuli, the participants' responses were sparse: in other words, each name corresponded to only a small subset of the 54 candidate images (mean = 2.8 items, SD = 3.4). Moreover, there was a high degree of consistency between participants in the set of stimuli that were selected for each name. This can be measured by the kurtosis of the distribution of responses over all possible words and stimuli, where sixteen votes is the maximum score (i.e. one vote per participant). The kurtosis is 17.74 making the distribution highly leptokurtic meaning that in many cases multiple participants matched a stimulus with a word or none did (Figure 2.8C). If it is sixteen it means that for one word all sixteen observers chose a specific stimulus, which happened two times. Participants were very confident matching stimuli to words with an average confidence interval of 7.3 on a 0-10 scale. Together these findings suggest that observers associate liquid names with specific appearances, and thus that visual appearance is quite diagnostic of liquid identity for a wide range of common liquids.

To gain a more thorough insight into the extent to which liquid identities are associated with specific ranges of optical and mechanical properties, we computed two indices to measure how selective participants were in terms of the optical and mechanical properties of the stimuli they chose. We define the 'optical focus' of the responses as the extent to which the responses to a given verbal item were restricted to a particular optical material, specifically, the kurtosis of the sum votes for each optical material. Analogously, we define the 'mechanical focus' as the extent to which the responses to a given item were restricted to a particular range of viscosity values, specifically the kurtosis of the sum votes for each viscosity (Figure 2.8B). Note that these two quantities are independent and not mutually exclusive, so that an item could have a low degree of focus for both properties (indicating that the verbal term is not very specific, e.g. 'liquified dough'); a high focus for one property but not the other (indicating that it specifies a particular optical appearance, but not a specific viscosity, or vice versa, e.g. 'chocolate pudding' or 'gum'); or a high focus for both properties (indicating that the name specifies a particular combination of optical appearance and viscosity, e.g., 'grape juice'). In Figure 2.8A, example items with low focus are coloured grey, items with high optical focus only are indicated in blue, items with high mechanical focus only are indicated in red, and items with high focus for both optical and mechanical properties are indicated in green. Note, that due to the reordering of the array into a vector, periodic responses every nine steps indicate that observers selected based

on the optical material (i.e., high optical focus), and an adjacent sequence of nine high values indicates that observers selected based on viscosity (i.e., high mechanical focus).

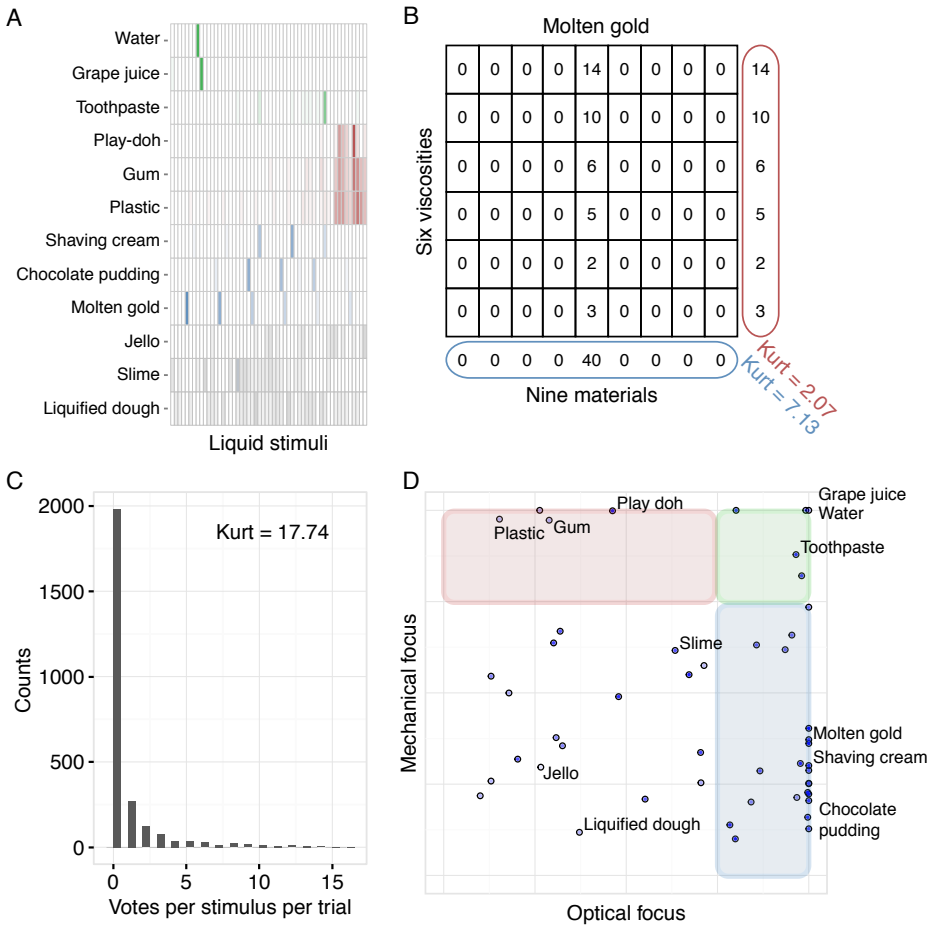


Figure 2.8: (A) Raw data of a sample of the 49 words. The 54 columns represent the 54 stimuli where the first nine optical materials of the runniest liquid are on the left. Periodic behaviour (blue) suggests optical focus, sequential behaviour (red) suggests mechanical focus, green is a combination of both and grey are more noisy names. (B) Raw data of the 'Molten gold' word. Six rows for six viscosities and nine columns for nine materials. The kurtosis of the sum of each row and column is used to calculate the optical focus (blue) and mechanical focus (red). (C) The distribution of votes per stimulus per trial with a maximum of sixteen votes for the sixteen participants. (D) The optical and mechanical focus for each word plotted in a single 2D space. The red area has high mechanical focus, the green area both high mechanical and optical focus, the blue area high optical focus. The intensity of the dots represents the confidence interval given by observers.

These 'focus' indices allow us to summarize the relative importance of optical and mechanical properties for all 49 liquid names in a single 2D space, as shown in Figure 2.8D (a complete overview is presented in the appendix). The x-axis shows optical focus and the y-axis mechanical focus. The intensity of the name dots represents the mean confidence ratings observers gave for each item. Note that items with lower focus values also tended to receive lower confidence ratings. Thus, it could be that items with low focus values

could simply be liquids for which none of the images corresponded well to the name. Thus, we should be cautious about concluding that some liquid names do not specify very precise appearances: it could simply be that the stimulus set did not contain appropriate images.

Most names are associated primarily with the liquid's optical appearance, as indicated by the blue region. Only four of the 49 names were associated with one specific viscosity but no specific optical appearance (red region). There are a few liquid names that specify a particular combination of optical and mechanical properties e.g. water that needs to be both runny and transparent. These results suggest that although we are very well able to perceive different viscosities, optical material appearance seems to be more a more distinctive feature than viscosity, and is therefore assigned more linguistic value by observers. Alternatively, it could be that a given class of liquid is generally prone to vary more in viscosity than in optical appearance, relative to the range of values that we used (e.g., "chocolate sauce" comes in lots of different thicknesses, but they are all brown).

2.4 Discussion

There are at least two routes by which optical properties could affect the perception of liquids: (1) via learned associations, or (2) by aiding (or hindering) the perception of shape and motion cues that are the basis for estimates of liquid properties. The former is specific to liquid perception, while the latter reflects general processes of mid-level vision. Our findings suggest that the extent to which observers rely on optical or mechanical information about liquids and their properties depends on the context and task. When asked to make visual matches of viscosity, shape and motion cues dominate, and optical material appearance barely influences perceived viscosity. This suggests that both learned associations and effects of shading on shape and motion estimates only weakly affect viscosity matches, at least when motion and shape cues to viscosity are strong. Although it is surely possible to find combinations of lighting and reflectance that do adversely affect shape and motion cues to viscosity (as occurred to some extent with the 'milk' material in Experiment 1), under typical viewing conditions, shape and motion processing is robust enough to derive viscosity-diagnostic information from the richly structured patterns that pouring liquids generate on the retina. In contrast to the viscosity-matching task, the rating task showed that subjective ratings of different liquid properties are based on mechanical cues, optical cues or a combination of both, depending on the specific property. Moreover, the pattern of responses suggests that processing of mechanical and optical cues is independent because of very limited interactions between the two: most

rating patterns could be well explained by a simple linear combination of the two kinds of information. The liquid naming experiment suggests that in most cases, we tend to assign names to liquids based mainly on their optical material appearance. This could mean that the optical material appearance is more diagnostic of the liquid (or more invariant) than its mechanical properties, at least for the range of appearances that we considered.

The finding that optical properties have only a weak effect on viscosity judgments makes intuitive sense because the physical processes determining viscosity are independent of those that affect the way the fluid scatters, reflects and absorbs light. In principle, any given optical appearance could co-occur with any possible viscosity and therefore optical characteristics do not provide a direct visual cue to viscosity. However, we reasoned that if a specific liquid with familiar viscosity properties is identified (via optical cues), this could bias or interact with viscosity estimates. Our findings suggest, however, that if this occurs, it is to a very small extent, at least when strong motion and/or shape cues to viscosity are present. Especially with moving stimuli, observers show close to perfect performance at matching viscosity across variations in optical materials. This suggests that when observers are judging a mechanical intrinsic property of the liquid like viscosity they rely primarily on shape and motion cues. As mentioned before with other scenes where mechanical cues are less dominant the influence of optical cues might increase. We do think that with our stimuli, designed to study viscosity, mechanical cues will keep their dominant role and therefore we will continue our studies investigating the perception of viscosity without taking potential influences of optical characteristics into account.

However, this is not to say that there is no role of optical properties in the perception of liquids and their properties more generally. In both the ratings and the naming task, some properties and liquids were associated with specific optical cues. However, our results provide an initial indication that optical and mechanical cues do not interact much with each other. This impression is amplified by the results of the second task in which observers had to rate six liquid properties: runniness, shininess, sliminess, stickiness, wetness and warmth. In most cases the various properties were determined primarily by either optical or mechanical cues on their own, e.g. ‘runniness’ decreases with increasing viscosity, but is unaffected by the optical properties of the liquids, while ‘shininess’ varies as a function of the specular reflectance of the material, and is barely influenced by viscosity (apart form the translucent milk-like material discussed above). There are however, some properties that are affected by both optical and mechanical characteristics. For example, both mechanical and optical cues play a role in the perception of ‘sliminess:’ green goo looks significantly slimier than copper-like liquids, even when the shape and motion are identical, but there is also a certain viscosity range that is considered to be slimiest

(neither too thick nor too runny—like Goldilocks' porridge). Nevertheless, even though both types of cue influence perceived sliminess, the interaction between the two is limited. All scores followed approximately the same curve, merely shifting additively up and down as a function of the optical characteristics (see Figure 2.5). This tends to suggest that the visual system treats the two kinds of information as distinct cues, which are then combined according to a simple 'weak fusion' process (Landy et al., 1995) to arrive at a subjective rating of 'sliminess'. Alternatively, it is possible that the influence of the optical and mechanical cues on the ratings proceeds via top-down associations. Specifically, it could be that the image cues serve to identify a specific liquid (e.g. green goo), whose cross-modal properties (e.g. sliminess) are recalled from memory. It is difficult to design experiments that tease apart the relative role of bottom-up and top-down contributions to ratings of high level properties of materials (Fleming et al., 2013).

Some caution is required in generalizing the conclusions of the matching and rating experiments. Here, we used a somewhat restricted range of stimuli consisting of one single scene of pouring liquids. It is almost certainly the case that other stimuli—such as those shown in Figure 2.9—can yield more extreme percepts of many of the features we tested here. For example, none of the stimuli in our experiment appeared as 'sticky' as the example shown in Figure 2.9A. Additional cues to stickiness presumably include the distinctive strands that span surfaces that have been stuck together and pulled apart, or in terms of motion, prolonged adhesion to other surfaces in the scene. Likewise, the molten metal in Figure 2.9B clearly conveys a stronger sense of high temperature than any of the stimuli in our experiments, presumably due to the visible glow, and other cues such as smoke or steam. If motion and shape cues to materials are extremely weak (e.g., in the limit a stationary liquid in a container), then optical cues will presumably carry a relatively stronger weight in determining perceived viscosity or other liquid properties. Nevertheless, we believe that the broader conclusion that different properties of liquids combine shape, motion and optical cues with different weights will withstand further scrutiny. This is for the simple reason that, while optical properties are almost always an ambiguous (i.e. unreliable) predictor of viscosity, shape and motion cues tend to be highly diagnostic of viscosity as soon as the liquid flows.

In our study, liquid names are mainly dominated by optical material appearance. The names considered descriptive of liquids in most cases span a range of possible viscosities. For example: the name 'chocolate' is assigned to all viscosities as long as the optical material is chocolate. This presumably reflects the fact that different concentrations and temperatures of chocolate yield a wide range of viscosities, but changes to the surface colour and optical appearance are less common. However, there are exceptions to the

Figure 2.9: (A) Example of a sticky material. (B) Example of a hot liquid. Images used under CC0 Public Domain license.

dominance of optical qualities. The term 'water' specifies both a specific colourless transparent appearance and a specific (runny) viscosity. 'Plastic' needs to look viscous but can have a wide range of different optical materials. Thus, specific recognizable liquids can be associated with both optical and mechanical properties. It seems that under many conditions, the optical material appearance (primarily colour, gloss and translucency parameters) are sufficiently distinct to specify many common liquids. Where the optical material appearance is not sufficiently specific for communicating a particular physical state, speakers may use additional terms that are specific to a liquid's mechanical aspects, such as sauce, paste, mouse, syrup or cream. Our observers did not report any problem with using multiple terms to specify appearances—including materials in viscosity states that they have not personally experienced before. We suggest that this approach to linguistic labelling of fluids—with basic level terms for optical appearance and qualifiers for viscosity—may reflect how we prioritize the visual cues that are used to identify liquids in general (i.e., optical appearance may dominate mechanical under many circumstances).

Chapter 3

Viscosity constancy across contexts

A similar version of this chapter is being prepared for publication:

van Assen, J. J. R. & Fleming, R. W. (2018). Viscosity constancy across contexts.

Despite radical retinal image changes our visual system allows us to perceive physical properties of materials in a consistant way. Liquids are a particularly interesting class of materials where intrinsic properties (e.g. viscosity, density) and external forces (e.g. gravity) enable a wide spectrum of possible shapes. Previous work has shown that despite the physical complexity of liquids we are surprisingly good at estimating properties such as viscosity. Here we investigated how constant we really are in perceiving viscosity. We used three stimuli sets: (1) A standard match set with 64 viscosity steps simulated in one scene. (2) The same scene simulated with 7 viscosities and 8 different noise perturbations. (3) 8 different scenes with different interactions (e.g. stirring, rain, and smearing) simulated with 7 viscosities. In the first experiment we defined the perceptual responses of the standard match set using maximum likelihood difference scaling. In experiment 2 and 3 observers matched viscosity using the standard match set with the noise perturbation set and scene variation set. We find that our stimuli evoke very constant impressions of viscosity with an impressive constancy of 95% across scenes. Small but systematic under- and overestimations are being made for certain scenes revealing the limits of our viscosity estimation ability. Global shape and motion information is varying across contexts and cannot account for the observers' constancy. This suggests that more localized tailor-made features are exploited to achieve viscosity constancy.

3.1 Introduction

All objects we interact with in everyday tasks have shape and material properties. To be able to interact successfully with these objects we need to be able to recognize and estimate properties of materials by sight. The perceived material properties are essential in inferring object affordances. Is something edible or poisonous, is a surface cold or glowing hot, is an object soft or hard; all basic estimations that are essential to our survival. Sharan et al. 2014 have shown that we are able to make these material estimations quickly and accurately. Liquids are especially interesting due to their mutable shapes and erratic nature. This means that there is an extensive spectrum of possible liquid appearances influenced by optical appearance, shape and motion. We are able to identify different liquids (e.g., water, honey, molasses) and estimate their properties such as runniness or stickiness. Maybe more important, we are able to identify the liquid in a glass of water to be the same as the water in a river. It is quite a remarkable feature that despite large image differences we are able to identify the materials as being the same, with the same properties. Here, we sought to investigate this ability further, to quantify how constant we actually are in perceiving liquids.

Identifying surfaces and objects across large image changes is most important for interactions with our surroundings. This perceptual constancy is one of the major challenges of object and material perception (Maloney and Wandell, 1986; Bülthoff et al., 1995; Tarr et al., 1998; Kraft and Brainard, 1999; Anderson, 2011; Foster, 2011; Motoyoshi and Matoba, 2012). Liquid perception is a relatively new field of study with specific research looking at intuitive physics (Bates et al., 2015), motion (Kawabe et al., 2015), 2D features (Paulun et al., 2015) and relative contributions of optical and mechanical cues (Van Assen and Fleming, 2016). One thing these studies have in common is that we are good at perceiving liquids and estimating their properties, such as viscosity. In this study we want to quantify how constant we are in perceiving viscosity across different contexts.

Studying constancy is especially interesting when it fails. When this happens it could tell us more about critical visual information that is necessary to remain constant, information that might be missing in the image. Identifying which type of information is causing a in- or decrease in constancy will tell us much about the cues we find most informative in an image.

Optical cues are one source of information we use to perceive liquids. A liquid's optical material appearance can tells us many things about the liquid. Orange juice (without pulp) basically has the same viscosity as water; what allows us to discern the two is the optical material appearance. Liquids occur as opaque, transparent, and translucent materials in

various colours allowing a larger range of optical appearances than most other material categories (e.g., metals, paper, wood). Van Assen and Fleming 2016 showed that optical cues definitely influence certain perceived properties of liquids. For example sliminess is perceived by a combination of optical and mechanical cues. Slimy objects need to be within a certain viscosity range to be perceived slimy (mechanical cue) but the optical cues add to the perception of sliminess as well. If you have exactly the same liquid shape and the optical appearance of copper or green goo then the green goo is perceived as slimier. Much research has been performed studying different aspects of material appearance such as colour (see Foster 2011, for a review), gloss (see Chadwick and Kentridge 2015, for a review), surface textures (Landy and Graham, 2004; Dong and Chantler, 2005; Emrith et al., 2010; Liu et al., 2015), transparency (Fleming et al., 2011; Faul and Ekroll, 2012; Schlüter and Faul, 2014) and translucency (Fleming et al., 2004; Fleming and Bülthoff, 2005; Xiao et al., 2014). We are very good at perceiving very fine differences in optical material appearances and we do this with great constancy across a large range of illumination and shape conditions.

In this particular study we are interested in viscosity constancy. The physics behind liquids and its internal properties are very complex. Despite this, we have a good sense how runny or viscous a liquid is. Van Assen and Fleming 2016 demonstrated that viscosity perception is mostly driven by mechanical cues, not optical cues. Therefore we keep the optical material appearance constant for all our test stimuli in this study. Glue and water are both transparent and yet we can see large difference in viscosity largely due to shape and motion information. Runny liquids tend to move faster, spread out and take the shape of its container; viscous liquids tend to pile up, but as time continues it will spread out into its container shape as well. Static images prove that motion is not required to perceive viscosity, but it does add additional information we can exploit to improve our estimations.

In this study we ask the following questions: How constant do observers perceive viscosity? Is this constancy driven by shape cues? Is this constancy driven by motion cues? We used physically-based computer simulations to simulate a wide range of scenes. Seven viscosities are simulated (roughly from water to molten glass) and observers had to match the viscosity with a fine-scaled 64-step standard match set. For this match set we performed a maximum likelihood difference scaling (MLDS) experiment (Maloney and Yang, 2003) to see if the perceptual distances within our match set are equal. One test set depicts pouring liquids, which is simulated eight times with different noise perturbations. The other test set consists of eight completely different scenes with different liquid interactions (e.g., rain, stirring, water wheel). Observers matched both these test sets with the standard match set selecting the most similar viscosity. Finally we compare the

constancy results with a 3D shape similarity metric and optical flow analysis to see if there are correspondences between constancy and shape or motion cues.

3.2 Methods

Three experiments were performed. Experiment 1 was a maximum likelihood difference scaling (MLDS) experiment (Maloney and Yang, 2003). MLDS provides a robust estimate of suprathreshold perceptual differences. It computes a perceptual scale with perceptual differences between stimuli. The MLDS experiment was performed with our 64-step viscosity match set (see Figure 3.1A).

In Experiment 2, observers were asked to perform a matching task where they had to match the viscosity of the test stimulus on the left with a match stimulus on the right. Observers could scroll through a standard set of match stimuli with 64 very fine viscosity steps. Both test and match stimuli were displaying a similar scene of animated pouring liquids. The test stimuli had an additional noise force field creating shape perturbations that are similar to gusts of wind blowing against the liquid (Figure 3.1B). Eight different versions of the noise force field were simulated creating eight variations of pouring liquids with varying shapes. These noise perturbations were simulated with seven different viscosities resulting in 56 different test stimuli (7 viscosities $\times$ 8 noise variations).

Experiment 3 was a similar matching task where only the test stimuli were different. Instead of a pouring liquid scene with noise perturbations, eight completely different scenes were simulated with varying liquid interactions (e.g. stirring, rain, see Figure 3.1C). These eight scenes were simulated with the same seven viscosities as in Experiment 2.

3.2.1 Stimuli

Simulation

The liquid stimuli were simulated with RealFlow 2014 and 2015 (V. 8.1.2.0192/V. 9.1.2.0193; NextLimit Technologies, Madrid, Spain). The viscosity of the standard set of match stimuli was simulated with 64 logarithmically spaced steps from 0.001 Pa·s to 100 Pa·s (equivalent from water to molten glass). The seven viscosities of the test stimuli (Experiment 2 and 3) were evenly spaced values of this same 64-step scale (steps 8, 15, 22, 29, 36, 43, 50) ranging from 0.004 to 7.74 Pa·s. Multiple particle solvers are available in the RealFlow software, here we used 'Hybrido' which allowed us to specify the dynamic viscosity of the liquids in real physical units, Pascal-second or Pa·s. Hybrido is a FLIP (Fluid-Implicit Particle) solver using a hybrid grid and particle technique to compute a numerical solution

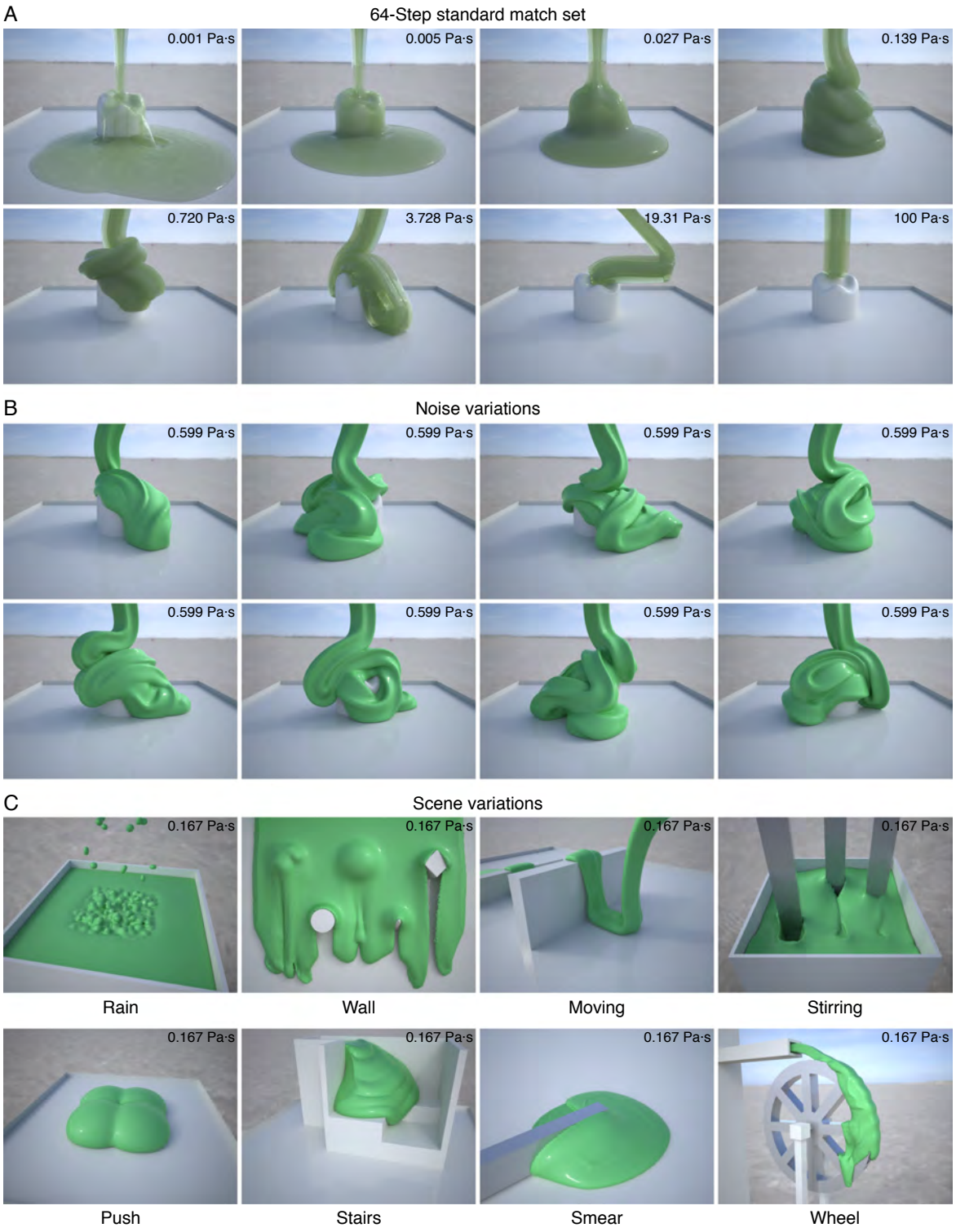


Figure 3.1: (A) The match set, a pouring liquid scene with 64 viscosity steps. Here eight logarithmic equally spaced viscosity steps are displayed from the runniest 0.001 Pa·s to the thickest 100 Pa·s. (B) Test set of the pouring liquids scene with noise perturbations. Here eight noise variations are shown of the same viscosity. (C) Test set of eight different scenes, here displayed with a simulated viscosity of 0.167 Pa·s. See Video B.1 for the animated stimuli.

to the Navier-Stokes equations describing viscous fluid flow. The discrete particles carry all information for the fluid simulation, but the solution to the equations is carried out on a grid. Once the grid solve is complete, the particles gather the information required from the grid to move forward in time to the next frame. To give the particles a continuous shape a meshing algorithm is applied. In our simulations viscosity was the only changing liquid property, e.g. density was held constant at one kilogram per litre. All scenes were simulated in a space of roughly one cubic meter. The amount of particles in the simulations varief from scene to scene and between viscosities but roughly 2 to 5 million particles were being used. 300 frames were simulated at 30 frames/s resulting in 10-second animations.

Rendering

The images were rendered using the Maxwell render engine that is built in Realflow 2014/2015. The images were rendered at an 800×600 resolution and we used a HDR light probe depicting a beach scene (from the Maxwell Resource Library by Dosch Design). The liquids of both noise and scene variations were rendered with a glossy green opaque material looking similar to liquid paint. The liquids in the standard match set were rendered with a green translucent material. The influence of different optical materials on viscosity judgements is minimal (Van Assen and Fleming, 2016).

3.2.2 Observers

Sixteen observers participated in the MDLS experiment. Groups of twelve observers participated in the two other experiments. The average observer age was 25.3 (SD = 5.41) of which 23 were female and 17 were male. All observers gave written consent prior to the experiment in accordance with the Declaration of Helsinki and prior approval was obtained from the local ethics committee of Giessen University. All observers were paid for participation and reported having normal or corrected-to-normal vision.

3.2.3 Procedure

The MLDS experiment was performed on a Dell T3500 workstation with Matlab 2015a (v. 8.5.0.197613) and the Psychtoolbox library (v. 3.0.12) (Brainard, 1997; Pelli, 1997). The monitor used was a Dell U2412M 24-inch, using factory default settings, a gamma of 2.2 and a resolution of 1920×1200 pixels. No training was performed because the instructions were very simple. A triads setup was used where the top stimulus was the test stimulus and the observer had to pick one of the two bottom match stimuli that differed most from

the test stimulus. This could simply be indicated by pressing the left or right arrow key on the keyboard, choosing the left or right stimulus. In total there were 224 trials. There was no time limit during the trials.

Both matching experiments were performed on the same system, a Dell T7610 with Matlab 2015a (v. 8.5.0.197613) and the Psychtoolbox library (v. 3.0.12). The display received an update between experiment two and three. Where the scene variations experiment was performed on a Dell U2412M 24-inch monitor using factory default settings, a gamma of 2.2 and a resolution of 1920 $\times$ 1200 pixels. For the noise variations experiment we used an Eizo ColorEdge CG277 27-inch monitor with a resolution of 2560 $\times$ 1440 and a gamma of 2.2. A training session took place before the main experiment, where observers could get acquainted with the interface and task. During this training session all the viscosities and scenes/noise variations passed by once in the form of eight normal trials. During each trial the test stimulus was shown on the right and the match stimulus on the left. The full animation of ten seconds was shown which looped throughout the trial. By using the left and right arrow key on the keyboard, observers could scroll through the 64 different stimuli of the standard match set. By pressing space the active answer was confirmed and after which the next trial was loaded. There was no time limit during the trials. Each trail was repeated three times in random order. Each experiment resulted in a total of 168 trials (7 viscosities $\times$ 8 noise/scene variations $\times$ 3 repetitions).

3.2.4 Shape similarity

To provide context on how similar the liquid shapes of our stimuli are, we developed a shape similarity metric. This metric uses the 3D liquid shape information based on the meshes from the liquid simulations and is represented in a voxel space. Voxels are practically 3D pixels, cubes. First the coordinate system of the mesh is aligned with the centre of mass. This means that coordinate (0,0,0) is on the centre of mass. The mesh is then transformed into a voxel representation. From this the AND and the OR products are calculated for every stimulus of the test set to every stimulus of the match set. The time frames were kept the same. The AND results store the pixels that overlap for both liquids. The OR results store the pixels where at least one liquid is positive. These pixels are counted and then the AND result is divided by the OR result which provide a ratio of overlapping liquids in total liquid space. How much the two liquid shapes have in common. Finally the liquid combination with highest common liquid ratio is selected as the correct match answer.

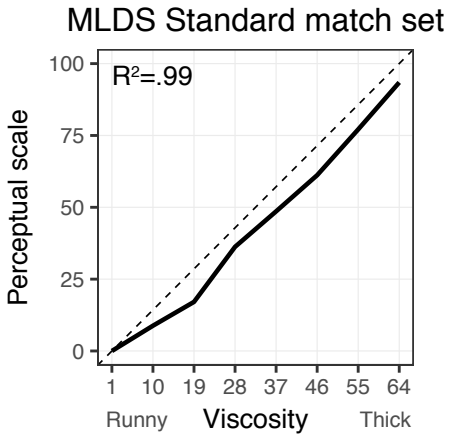


Figure 3.2: The perceptual scale of the match stimuli set. The x-axis shows the steps of our 64-step scale. The y-axis shows the perceptual scale. The R^2 value shows the results of a linear fit between the physical and perceptual viscosity scale

3.2.5 Optical flow

Motion is another rich source of information that next to shape is informative about the mechanical properties of liquids. Kawabe et al. 2015 showed that the mean speed of optical flow is highly predictive of perceived viscosity. To evaluate whether motion can predict viscosity matches with our stimuli, we used the same iterated pyramidal Lucas-Kanade method (Bouguet, 1999) to calculate optical flow. The match was selected based on the most similar flow vector.

3.3 Results

3.3.1 MLDS results

Figure 3.2 shows the results of the MLDS experiment. There are two noticeable results. (1) The perceptual scale is practically linear and correlates very highly with the physical viscosities ($R^2 = 0.99$, $F(1,6) = 704.7$, $p < 0.001$). This means that physical viscosities with logarithmic step sizes have the same perceptual distances. Viscosity step 1 has in both physical and perceptual space the same distance to step 10. (2) There seems to be a slight offset. This means that perceptual viscosity is slightly underestimating the physical viscosity.

Here the most important conclusion is that our 64-step standard match set will be able to provide a consistent matching standard without any peculiar anomalies.

3.3.2 Noise variations

Figure 3.3 shows the results of Experiment 2 where different noise perturbations were matched with the standard set. The first thing to notice is how good the observers are in matching viscosity ($R^2 = 0.99$, $F(1,54) = 7609$, $p < 0.001$). We consistently get very linear results for every noise variation. There is a systematic underestimation of runny liquids, they are perceived as more viscous, and the same happens for the thick liquids, which are perceived runnier. When we look at our shape similarity metric we see that it is a very good predictor ($R^2 = 0.96$, $F(1,54) = 1173$, $p < 0.001$). Interestingly the shape similarity model seems to make the same under and overestimation errors. This is further supported by the fact that the model seems to explain why the perceived viscosities (RMSE = 4.25) are slightly better than the physical viscosities (RMSE = 5.59), although the differences are small.

The constancy across variations is striking. Figure 3.5A shows a different representation of the consistency. It shows the RMSE of each variation in relation to the mean across variations. Taking the mean of these error distances is a representation of constancy, which in this case is RMSE = 0.72. In other words there is 1 % error between variations or 99% variation constancy.

This is further supported by a low error across observers, Figure 3.5C. The y-axis shows the RMSE in relation to the physical truth and the x-axis how precise each observer is within their own repetitions. Overall we can say that observers are extremely consistent in matching viscosities across noise perturbations.

3.3.3 Scene variations

Figure 3.4 shows the results of Experiment 3, where different scene variations were matched with the standard match set. Compared to the noise variations there is more variation across scenes, but we are still very good in matching scenes by their viscosity ($R^2 = 0.89$, $F(1,54) = 447.1$, $p < 0.001$). We still get very linear results although the relation with the physical viscosity varies more. The stairs scene (Figure 3.1C,Video B.1) shows practically perfect matches without any over- or underestimation. When we look at our shape similarity metric we see much more variation and some very bad predictions ($R^2 = 0.04$, $F(1,54) = 2.391$, $p = 0.13$). This is not completely unexpected, the liquid shapes vary much more.

The constancy across scenes has decreased. Figure 3.5B shows larger errors, the mean error across scenes is RMSE = 3.19. This translates back to 5% error or a scene constancy of 95%. A very impressive result for much more radical shape changes in different contexts.

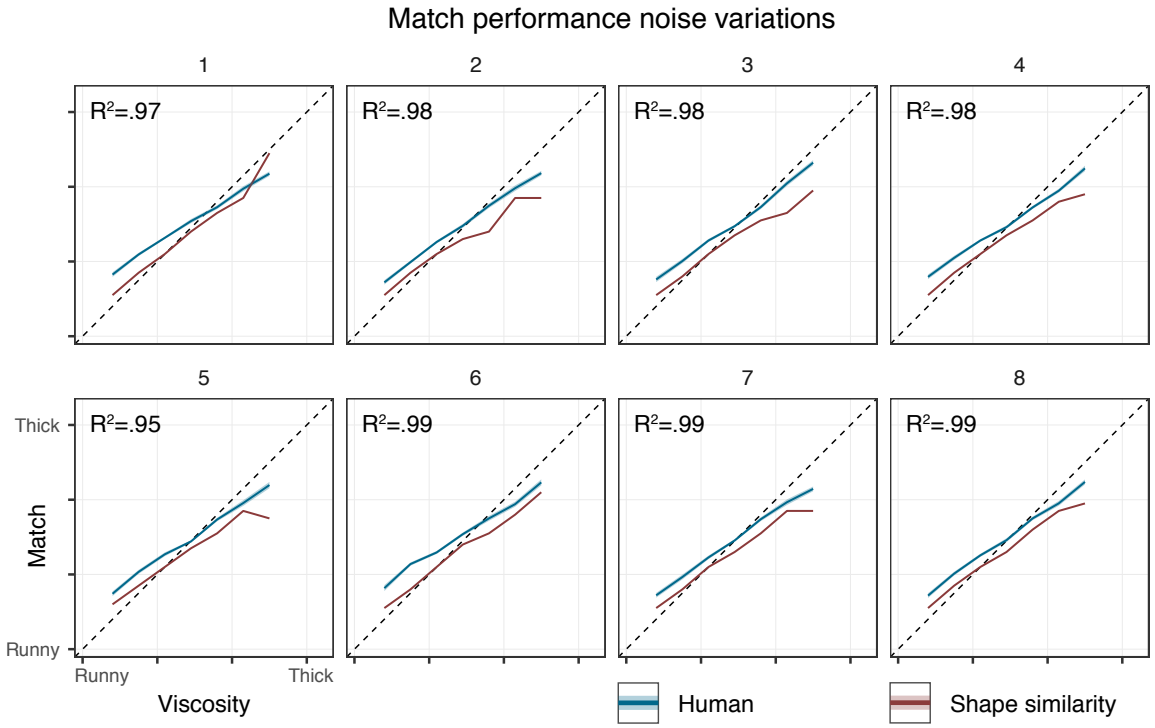


Figure 3.3: The matching performance of the noise variations. The x-axis shows the viscosity of the test stimulus and the y-axis shows the viscosity of the matched stimulus. The blue line shows the perceived viscosity with very small error ribbons that show the standard error of the mean (SEM). The red line is the predicted match of our shape similarity metric. The R^2 value shows the results of a linear fit between the perceived and predicted matches.

In line with the findings we find lesser agreement across observers as well (Figure 3.5D). There seems to be one observer who is very inconsistent within repetitions and there is an overall increase of 37% in individual errors.

3.3.4 Shape similarity

One hypothesis is that the variation across scenes depends on the shape variation across viscosities for the same scene. If a runny liquid looks very different from a thick liquid in the same scene there is much shape variance within the scene. If there is little shape difference across all viscosities in the scene it is plausible that it is harder to estimate viscosity differences. Especially if this is tested with a matching task where the match set does show large shape differences across viscosities. One example could be the Push scene (Figure 3.1C,Video B.1). The liquid shape across viscosities does not vary as much as with other scenes and therefore might be less descriptive for viscosity judgements. This is supported by the more horizontal slope in Figure 3.4.

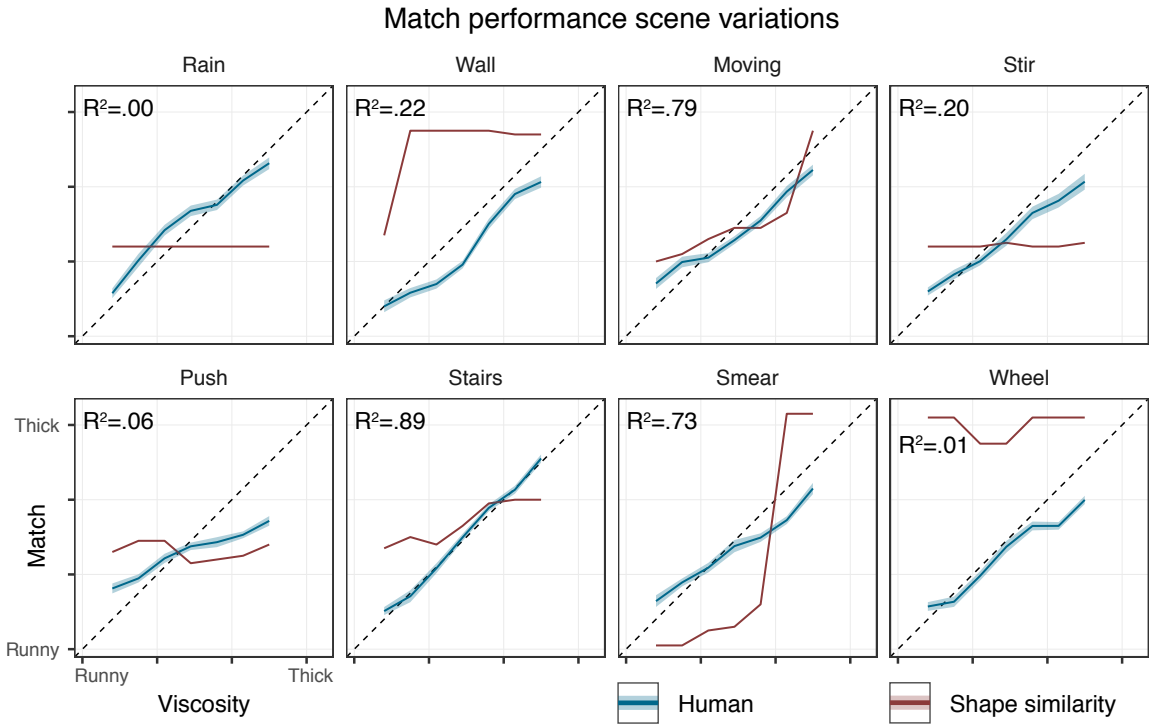


Figure 3.4: The matching performance of the scene variations. The x-axis shows the viscosity of the test stimulus and the y-axis shows the viscosity of the matched stimulus. The blue line shows the perceived viscosity with error ribbons that show the standard error of the mean (SEM). The red line is the predicted match of our shape similarity metric. The R^2 value shows the results of a linear fit between the perceived and predicted matches.

To test this hypothesis we ran the same AND/OR voxel operations but instead of using the match set it was performed with the other six viscosities of that scene. This provides a measurement of how much the ratio between shape overlap and the sum of the shapes, changes across viscosities. We find that there is no significant correlation. This means that variance in global shape differences do not explain differences in perceived viscosity. It could be that more local shape information causes these differences. A set of distinct features that vary less or more within a scene.

3.3.5 Motion information

One other possibility is that we make matches based on similar motion flow between the test and match stimulus. This is predicted by the match stimulus that is most similar to the test stimulus in terms of optical flow. Across scenes the optical flow matches are not very predictive of the perceived viscosity (R^2 = 0.14, F(1,54) = 8.706, p < 0.01). We find that the optical flow matches are only significantly predictive for the stairs scene (R^2 = 0.81, F(1,5) =

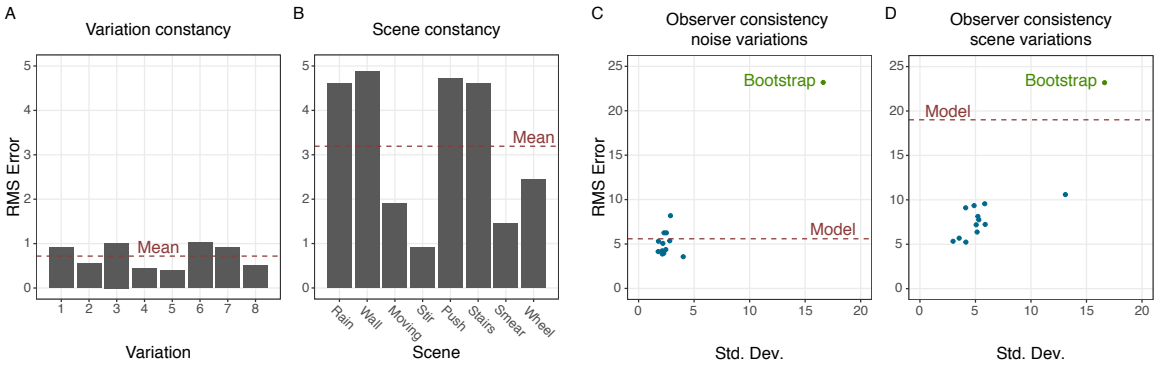


Figure 3.5: (A) Constancy across variations, x-axis shows the eight different variations and the y-axis the RMSE between each matched variation and the mean of the matched variations. The red dotted line shows the mean. (B) Exactly the same graph only for experiment 3 with eight scene variations instead of noise variations. (C) Individual observer consistency for Experiment 2. The x-axis shows the mean standard deviation within repetitions, the y-axis shows the error between individual observers and the physical truth. The green bootstrap dot shows the performance of 1000 random draws. The red dotted line shows the accuracy of the shape similarity model. (D) The same graph as C only for experiment 3.

21.22, $p < 0.01$). This might explain why the observers perform so well in this scene ($R^2 = 0.99$, $F(1,5) = 847.7$ $p < 0.001$). Clearly this is a scene where motion patterns between match and test stimulus in- or decrease with similar magnitudes across viscosities. However, as other scenes don't have this commonality with the match set, observers need to rely on additional, maybe less informative, cues.

3.4 Discussion

How we achieve perceptual constancy is an important challenge in material perception. How is it possible that despite radical image changes we are able to perceive materials consistently? Here our purpose was to quantify how constant we are in perceiving viscosity of liquids. We did this by using matching experiments where we have one scene simulated with 64 fine viscosity steps (Match set). We had two different test sets, one depicting pouring liquids with eight different noise perturbations and eight completely different scenes with various liquid interactions (e.g., smear, push, stir). Using a matching task instead of ratings enables us to keep the influence of individual mental scales to a minimum, context is provided. Instead of thinking how viscous this particular liquid is in some internal viscosity reference space, we asked the observers to match this to a predefined and perceptually quantified reference space. Because of this we can make clear comparisons between contexts as they are all matched with the same reference set.

We see that observers are very good at matching viscosity ($R^2 = 0.99$ for the noise variations and $R^2 = 0.89$ for the scene variations). It is quite incredible that we can match

two very different liquid shapes with this precision. Our viscosity constancy is very high (99% for the noise variations and 95% for the scene variations). This demonstrates that we are very well able to generalize estimations of material properties across various contexts. This will set the benchmark for future models designed to estimate viscosity.

It is important to ask what scene invariant information we use to make these constant estimations. This question concerning perceptual constancy translates to many different fields. Using our shape similarity metric we demonstrated that shape similarity of the whole liquid shape is not the very predictive. Across scenes these shapes are very different. Even by aligning them and choosing the most similar shape match it is very hard to predict matches across scenes ($R^2 = 0.04$). The entire liquid shape varies very much across contexts and other information is required that stays invariant across scenes.

These experiments were performed with animated videos and providing very rich motion information. Previously it has been shown that motion can be very predictive of viscosity estimates (Kawabe et al., 2015). We applied a similar analysis on the scene variations stimuli set. In many cases the mean optical flow for each stimulus was not very predictive of viscosity ($R^2 = 0.14$). There was one exception, the Stairs scene (Figure 3.1C,Video B.1) showed a significant effect ($R^2 = 0.81$). This means that for each viscosity of the Stairs scene there was an equivalent match set stimulus with similar optical flow. Observer matched in this particular scene the actual viscosity very accurately as well ($R^2 = 0.99$) suggesting the motion contributed positively to other available cues. This supports the idea that we combine cues using 'weak fusion' processes (Landy et al., 1995; Ernst and Bülthoff, 2004). In this particular case one scene shares similar motion cues with the match set and therefore observers are more accurately matching viscosity. When the domain of motion contains informative cues it will contribute to accuracy.

There remain two open questions. (1) If global shape and motion features can't explain viscosity constancy across scenes what can? (2) How does our visual system knows which cues are informative for that particular context? To answer the first question, we think that mid-level shape and motion features provide rich cues of more localized groupings of information on the object. More often mid-level features have been suggested to deliver key contributions in material perception (Adelson, 2000; Anderson, 2011; Marlow et al., 2012; Paulun et al., 2015). With mid-level shape and motion features we think of more local details such as blobbiness, pulsing and spread. The shapes of the liquids in different contexts (Figure 3.1,Video B.1) have very distinct features and they are not mutually exclusive, a liquid can be blobby, pulse and spread out at the same time. Each liquid shape can therefore be represented in mid-level feature space, which is used to estimate viscosity. It is important to emphasize the difference between our binary shape-to-shape similarity

measurement and mid-level shape features. A mid-level shape feature is a higher-level concept than a binary shape comparison. A spherical blob on the left side of a liquid shape can be a very distinct feature, but this blob can occur as easlily on the right side of the shape as well providing the same information. In this case our shape similarity metric would say the shapes differ, while in mid-level feature space the information stays the same, invariant across shapes.

To answer the second question is maybe even harder. How do we know which visual cues are most informative in any particular context? What makes a feature distinguishable? We demonstrated that a larger shape variance within a scene is not explaining performance differences across scenes. One could argue that, similar to the previous argument, variance in mid-level shape features is key. If a scene contains only variance in blobbines and spread across viscosities it might be harder to detect viscosity changes compared to scenes that vary across ten shape features. The missing puzzle piece is how to recognize distinct features as informative in the first place. What makes spiralling recognizable as spiralling? There must be some heuristics, rules, descriptors of the features that allow us to identify them as such. To find these rules would require a similar approach as suggested by Paulun et al. 2015, only performed in the 3D domain, which is a difficult problem to solve. How do you specify rules that measure spiralling in 3D and are insensitive to scale, orientation or other factors? There is one certainty; our visual system is able to exploit a rich, dynamic set of cues that stay invariant across contexts, enabling great constancy in viscosity perception.

Chapter 4

Visual features of liquids

A similar version of this chapter has been published as:

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

Perceptual constancy —identifying surfaces and objects across large image changes— remains an important challenge for visual neuroscience (Maloney and Wandell, 1986; Bülthoff et al., 1995; Tarr et al., 1998; Kraft and Brainard, 1999; Anderson, 2011; Foster, 2011; Motoyoshi and Matoba, 2012). Liquids are particularly challenging because they respond to external forces in complex, highly variable ways, presenting an enormous range of images to the visual system. To achieve constancy, the brain must perform a causal inference (Biederman and Gerhardstein, 1993; Gilchrist et al., 1999; Riesenhuber and Poggio, 2000) that disentangles the liquid's viscosity from external factors—like gravity and object interactions—that also affect the liquid's behaviour. Here, we tested whether the visual system estimates viscosity using 'mid-level' features (Adelson, 2000; Anderson, 2011; Marlow et al., 2012; Paulun et al., 2015) that respond more to viscosity than other factors. Observers reported the perceived viscosity of simulated liquids ranging from water to molten glass exhibiting diverse behaviours (e.g. pouring, stirring). A separate group of observers rated the same animations for 20 mid-level 3D shape and motion features. Applying factor analysis to the feature ratings reveals that a weighted combination of four underlying factors (Distribution, Irregularity, Rectilinearity and Dynamics) predicted perceived viscosity very well across this wide range of contexts ($R^2 = 0.93$). Interestingly, observers unknowingly ordered their mid-level judgments according to the one common factor across contexts: variation in viscosity. Principal Component Analysis reveals that across the features, the first component lines up almost perfectly with the viscosity ($R^2 = 0.96$). Our findings demonstrate that the visual system achieves constancy by representing

stimuli in a multidimensional feature space—based on complementary, mid-level features—which successfully cluster very different stimuli together and tease similar stimuli apart, so that viscosity can be read out easily.

4.1 Results and discussion

If the estimation of viscosity proceeds hierarchically—through a weighted combination of mid-level features describing dynamic 3D shape properties—it should be possible to identify such features and use them to predict perceived viscosity across variations in other scene variables. To test this hypothesis, we simulated liquids with a wide range of viscosities interacting with a variety of different scenes (see Video C.1 and C.2). In Experiment 1 we made detailed measurements of viscosity perception in a simple scene in which each liquid poured vertically onto an object on a plane (Figure 4.1A). The 10 sec animations, depicting liquids with 32 different viscosities, were divided into six (1,67 sec) time periods. On each trial, observers viewed eight videos of liquids with different viscosities from the same time period, and rated the perceived viscosity by adjusting sliders for each video. Results are shown in Figure 4.1B. Consistent with previous work (Paulun et al., 2015; Kawabe et al., 2015; Van Assen and Fleming, 2016), we find that observers are excellent at judging viscosity: the regression between their ratings and physical truth was $R^2 = 0.96$, $F(1,190) = 4941$, $p < 0.001$. There was also a mild tendency to see later time periods as runnier. The range of responses across observers is shown in Figure 4.1E.

Comparing viscosities across liquids is relatively straightforward if all other scene factors are held constant. The deeper challenge is to achieve constancy—i.e., generalization across contexts. To investigate constancy, in Experiment 2 we created a series of scenes in which liquids underwent qualitatively different behaviours, such as oozing through holes, being stirred in a container, or interacting with a waterwheel (Figure 4.2A, Video C.1). Seven viscosities were simulated, and observers again rated viscosity, this time for the entire ten seconds of each animation (see Supplementary Information for details). We found a significant decline in viscosity constancy across scenes, as indicated by the different rates at which the columns in Figure 4.2B change from light to dark. Nevertheless observers were still very well able to differentiate and order the seven simulated viscosities across qualitatively different behaviours, yielding a regression between the ratings and physical truth of $R^2 = 0.92$, $F(1,54) = 656.7$, $p < .001$. The range of responses across different individuals is shown in Figure 4.2E.

Next, we sought to identify a set of mid-level shape and motion cues that predict viscosity perception. Rather than identifying potential cues through physical analysis, we

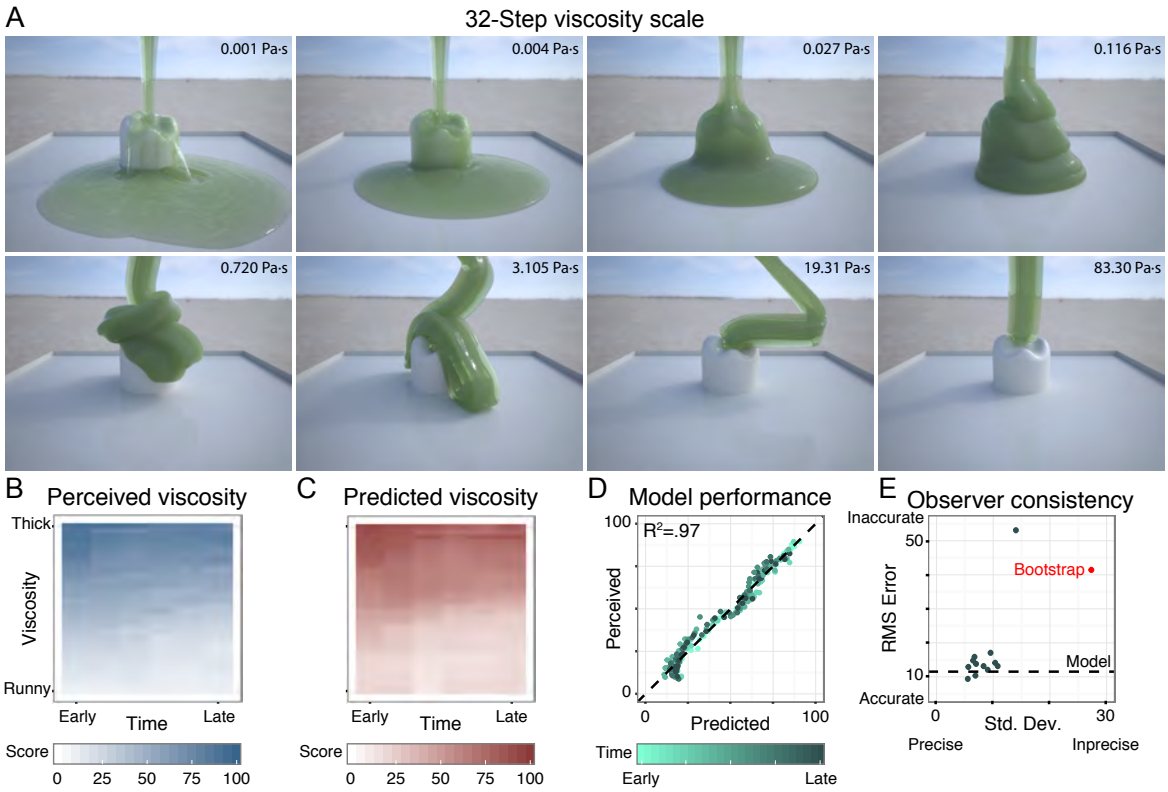


Figure 4.1: (A) Eight equally spaced viscosity stimuli spanning the full range of viscosities (frame 90 of 300). (B) Mean viscosity ratings for all videos in Experiment 1. (C) Predicted viscosity for same stimuli, based on the four-factor model. (D) Scatterplot comparing model predictions to mean responses across observers and repetitions. Darker greens indicate later time periods. (E) Root mean square errors relative to ground truth viscosities, and standard deviation of responses across repetitions for each observer (dots); red dot indicates bootstrapped estimate of random performance based on 1000 random draws.See also Figure C.1, Video C.1, Video C.2 and Table C.1.

took a data-driven approach, in which we selected a broad set of hypotheses through phenomenology, which could then be tested, rejected and refined through experimentation. To do this, we viewed the 'pouring liquids' stimulus set and brainstormed features that (1) described aspects of the stimuli's 3D shape and motion; (2) varied across stimuli and (3) could be described to participants verbally. We also asked four naïve observers to brainstorm a list of features describing the liquids. Although the terms they identified were not identical to ours, they were judged by another group of observers to overlap substantially with our list, suggesting we had identified a reasonable set of features to test. Importantly, we view the initial feature list as a superset of potential cues—i.e., hypotheses—which we sought to cull through subsequent analyses.

To do this, in Experiment 3 and 4, two new groups of observers viewed the same videos as in Experiments 1 and 2, but instead of rating viscosity, they rated the twenty features

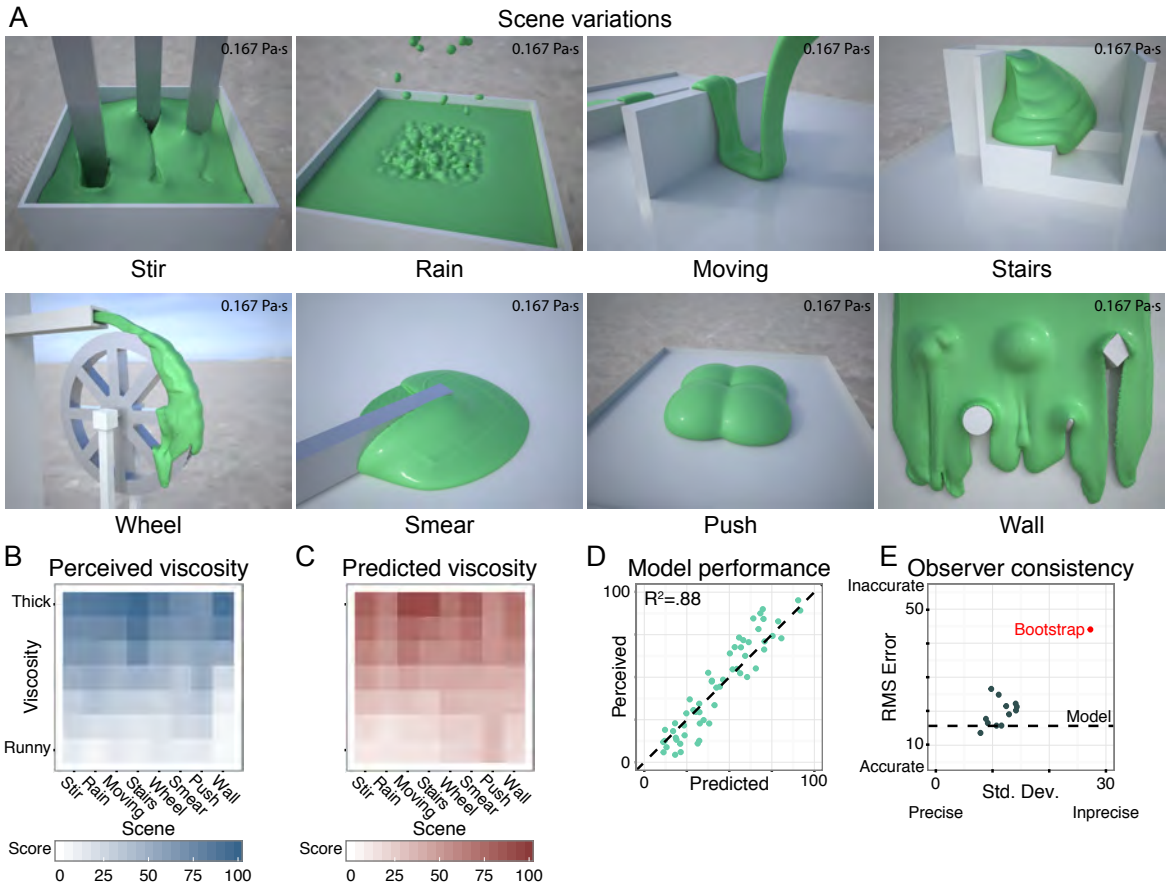


Figure 4.2: (A) Eight different scenes simulated with the same viscosity of 0.167 Pa·s. (B) Mean viscosity ratings for all scenes in Experiment 2. (C) Predicted viscosity for same stimuli, based on the four-feature model. (D) Scatterplot comparing model predictions to mean responses across participants and repetitions (from B and C). (E) Root mean square errors relative to ground truth viscosities, and standard deviation of responses across repetitions for each observer (dots); red dot indicates bootstrapped estimate of random performance based on 1000 random draws. See also Figure C.2, Video C.1, Video C.2 and Table C.1.

(e.g. 'compactness', 'elongation', 'pulsing', 'clumping'; see Table C.1 for a complete list with specific instructions). None of the features referred to the liquids' material properties. Instead they targeted the stimulus's 3D shape and motion characteristics to test the hypothesis that viscosity is inferred from specific weighted combinations of such cues.

Results for three of these features with pouring liquids (Experiment 3) are shown in Figure 4.3A (see Figure C.1 for all 20 features). Unlike viscosity ratings, the feature judgments often varied in complex, non-monotonic ways as a function of viscosity and time period. This means the different features provide potentially complementary cues about the liquid. Although some individual features predict viscosity perception in some scenes, few features predict all the data well on their own. Instead, the brain likely combines multiple cues to achieve more robust estimates of viscosity. There were strong correlations

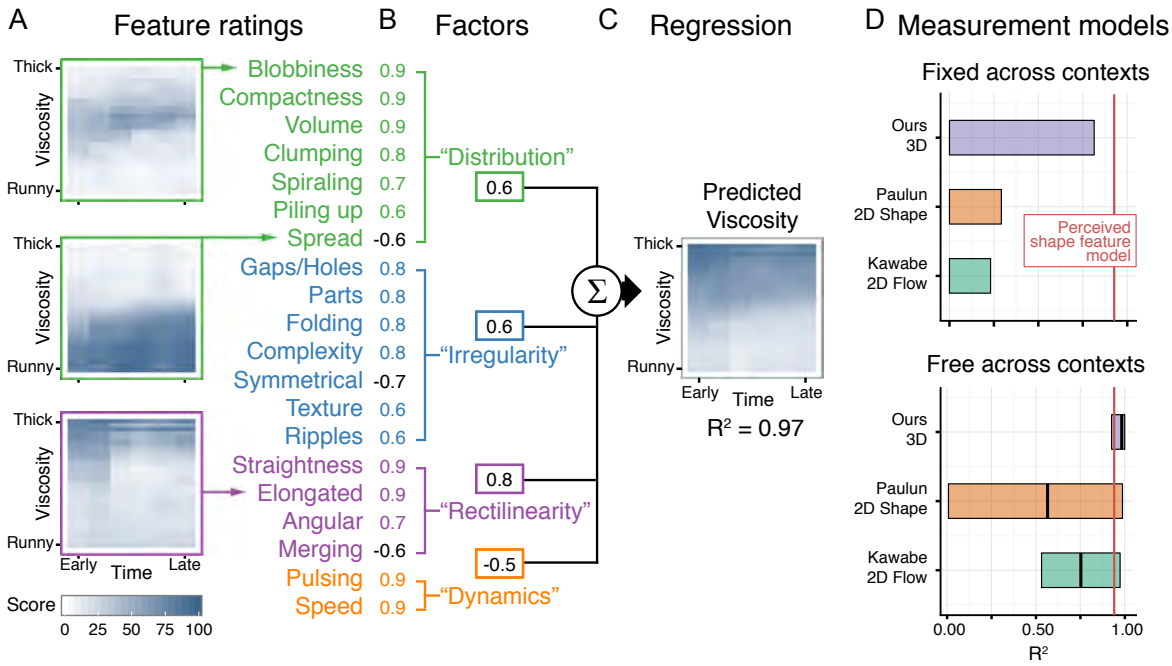


Figure 4.3: (A) Example ratings for three features from Experiment 3. (B) Factor analysis weights for the twenty perceptual features into the corresponding four factors, for which they have the largest weights. (C) Multiple linear regression combines the four factors into a viscosity prediction for the data from Experiment 1. (D) Performance of the measurements based models compared with the perceived shape feature model (red line). These regressions were performed on the second stimulus set with eight different scenes. Free and fixed across contexts refers to one set of weights for all scenes or a separate set of weights for each scene. See also Figure C.4.

between features (Figure C.3A), suggesting a smaller number of true underlying factors describing the liquids' shape and motion.

To test this, we performed a factor analysis (Figure 4.3). A Horn test (Horn, 1965) revealed four basic factors (each a weighted combination of the feature ratings; Figure C.3): (1) 'Distribution'—describing the extent the liquid clumped together vs. spread out; (2) 'Irregularity'—describing how complex and detailed its shape was; (3) 'Rectilinearity'—capturing how straight and angular the liquid appeared; and (4) 'Dynamics' describing its motion properties. When a new group of nine observers were asked to judge these factors directly, the responses correlated significantly (mean $R^2 = 0.52$) with the factors derived from the feature ratings on the same stimuli. The process of applying factor analysis allowed us to narrow down the broader list of twenty features to four more refined and specific hypotheses.

To test whether these factors could predict viscosity perception, we performed a multiple linear regression, using the factors derived from Experiment 3 to predict the viscosity ratings data from Experiment 1. The model predicts the viscosity data extremely well, $R^2 = 0.97$, $F(4,187) = 1386$, $p < .001$, far better than random predictors (a bootstrapping analysis

with 1000 repetitions revealed 185 predictors would be required to achieve equivalent non-significant performance). This indicates that a simple weighted linear combination of dynamic 3D shape features is sufficient to explain perceived viscosity. Note, again, that the combination of factor analysis and regression allows us to reduce our initial hypotheses, and to quantify the relative roles of individual cues. Of course, on its own, our finding does not strictly imply that the estimation of mid-level features is prior to the inference of viscosity. It is logically possible that observers derived their feature judgments from the perceived viscosity. However, we suggest that the detailed—often non-monotonic—feature ratings makes this unlikely. On grounds of parsimony, it seems more likely that viscosity is inferred from the mid-level features than vice versa.

The key challenge of viscosity perception is to achieve constancy across dramatic changes in the liquid's behaviour. To test how well the model predicts viscosity constancy, we applied the factor loadings and regression weights derived solely from the 'pouring' scene (Experiments 1 and 3) to the feature ratings from Experiment 4, to measure how well the model predicted viscosity perception in the other eight scenes (Experiment 2). Results are shown in Figure 4.2D. Despite having no new training data or additional free parameters, the model generalizes to the eight new scenes remarkably well, ($R^2 = 0.88$, $F(1,54) = 391.4$, $p < .001$). These results confirm that a relatively small number of mid-level stimulus characteristics—related to how fast they move, how much they spread out or clump together, how irregular they are and how rectilinear—determine the perception of viscosity across a very wide range of contexts.

To test the robustness of these conclusions, we also ran the factor analysis and regression 'in reverse', using the data from the 8 scenes (Experiments 2 and 4) to build a model for predicting perceived viscosity. As before, this model predicts its training data very well ($R^2 = 0.96$, $F(4,51) = 294$, $p < .001$). When used to predict the viscosity ratings from the pouring scene (Experiment 1), this model also generalizes well ($R^2 = 0.77$, $F(1,190) = 637$, $p < .001$; again with no free parameters), although not as well as the original model, which is unsurprising given that only about a third as much training data was available (56, rather than 192 data points). To quantify the similarities between the two models, we computed representational dissimilarity matrices (RDM, Kriegeskorte et al., 2008) describing the differences between stimuli in their respective factor spaces (Figure C.3C). The RDMs correlated highly for both Experiment 3 (pouring scenes: $R^2 = 0.65$, $F(1,18334) = 33470$, $p < .001$) and Experiment 4 (8 scenes: $R^2 = 0.58$, $F(1,1538) = 2090$, $p < .001$), suggesting that the models learned similar representations of the stimuli from the feature ratings. Together these findings further suggest that representing stimuli using multiple complementary factors enables viscosity constancy.

Of course, some caution is required in interpreting these results. Although the range of liquid behaviour we tested was broad, there may be some conditions where other, untested, features could predict viscosity perception even better. Indeed, while these factors account for viscosity perception once a given stimulus is identified as a liquid, it is highly unlikely that they are sufficient to determine whether a given stimulus is a liquid in the first place. Many non-liquid forms could appear as 'distributed', 'irregular', 'rectilinear' and 'dynamic' as one of our stimuli, without appearing to be a liquid of a specific viscosity. Thus, although these factors are important for viscosity estimation, they do not explain all aspects of liquid perception across all possible conditions. Nevertheless, the broader conclusion is that the visual system can achieve a high degree of constancy by representing liquids in a feature space incorporating multiple, complementary measurements. A similar approach has been proposed to account for errors of gloss perception (Kim et al., 2012; Fleming, 2012); our results suggest that such an approach predicts both successes and failures of constancy in material perception more generally.

Why do these features work? The key challenge of constancy is that movies of the same liquid in different scenes are very different from one another in the image domain, while movies of different liquids in the same scene are much more similar (Figure 4.4A). Somehow the visual system must remap the representational space to organize the stimuli by their viscosity. We find that this is exactly what the mid-level features achieve. To investigate this, we performed Principal Component Analysis (PCA) on the data from the second stimulus set (eight scenes). Figure 4.4A depicts each stimulus in the pixel similarity space by performing PCA on the rescaled grayscale pixel data of the entire video sequence. This represents the raw input to the visual system. The ellipses show standard errors around the mean for the seven viscosities. The substantial overlap of the ellipses indicates that raw retinal image similarities provide a poor basis for viscosity perception, demonstrating the extent of the challenge confronting the visual system. In contrast, Figure 4.4B shows the PCA space of the features ratings, which reveals a clear and systematic ordering of the stimuli by viscosity. It is important to emphasize that observers were simply instructed to rate different shape and motion features—viscosity was never mentioned. Despite this, the first principal component of the ratings is highly correlated with the actual viscosity ($R^2 = 0.96$, $F(1,54) = 1212$, $p < .001$). This demonstrates that despite massive physical variations across scenes, observers unknowingly arranged the stimuli according to the one common factor across these scenes: the viscosity. This impressive ability strongly suggests that the visual system achieves constancy by identifying features that transform the perceptual space to extract invariant material properties and negate the effects of other scene variables.

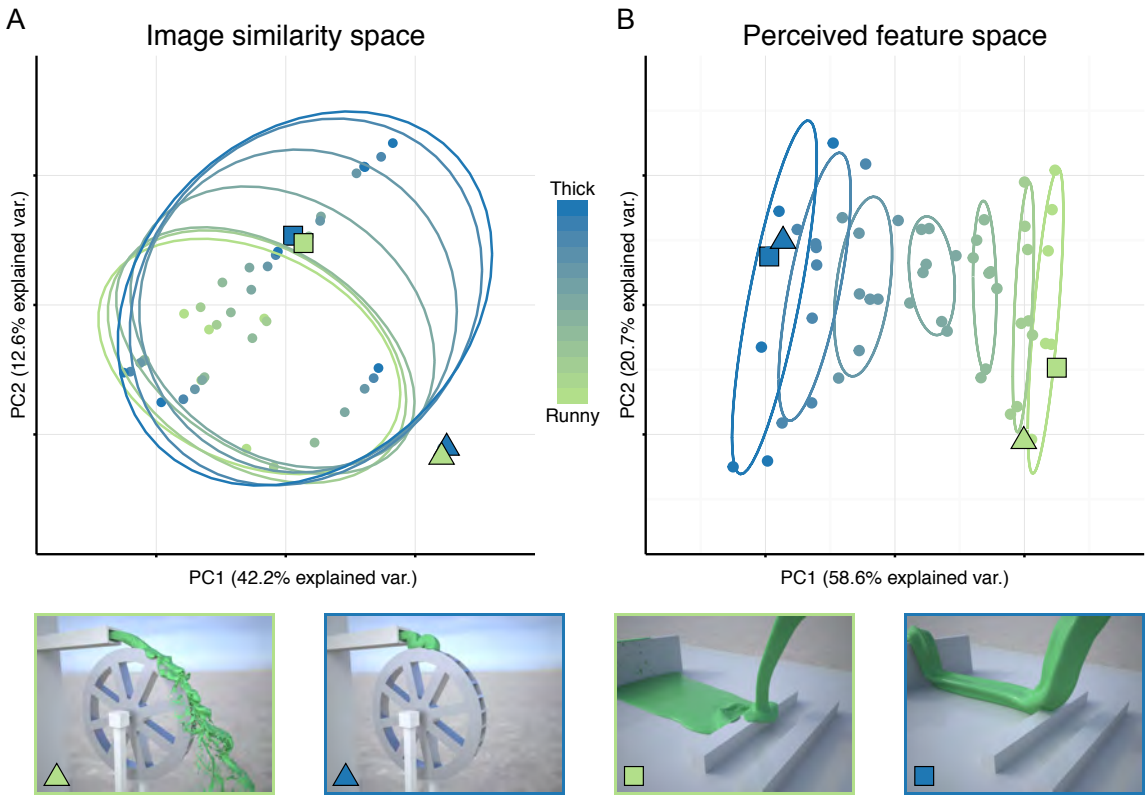


Figure 4.4: (A) PCA space performed over all stimulus pixel values (36 million dimensions), a measurement of image similarity, representing the raw input to the visual system. (B) PCA space of the perceived shape features (twenty dimensions). Note that changes of viscosity almost perfectly are aligned with the first component ($R^2 = 0.96$). The triangle and cube data points are the most runny (green) and viscous (blue) stimuli of two scenes. In image similarity space the two viscosities are very similar and grouped together, in perceived feature space the stimuli are separated by viscosity. This demonstrates that while the retinal input is dominated by extrinsic scene factors rather than viscosity, the perceptual feature space unravels the common quantity across these scenes: the liquids' viscosity.

An important current debate in material perception research is the extent to which 2D image quantities are sufficient for material judgments (Motoyoshi et al., 2007; Sharan et al., 2013), or whether 3D surface structure plays a crucial role (Marlow et al., 2015; Marlow and Anderson, 2015; Marlow et al., 2017). To provide some perspective on this, we developed a model using four shape metrics computed directly from the liquids' 3D meshes (see Figure C.2 and Supplemental Results), which we compared against two previous models, based on 2D optical flow (Kawabe et al., 2015) and 2D shape (Paulun et al., 2015) (Figure 4.3C). The 2D models generalize poorly across our stimuli (Kawabe et al.: $R^2 = 0.23$; Paulun et al.: $R^2 = 0.29$). Only the novel 3D model predicts viscosity perception moderately well with fixed weights across scenes ($R^2 = 0.81$). Unsurprisingly, when we perform separate regressions for each scene independently (free weights), all three measurement models perform somewhat better. However, the results suggest that 3D shape does contribute

to the robustness of viscosity perception beyond simple 2D image measurements. How might 3D information be used? Simply representing local 3D structure at every surface point would be insufficient to infer viscosity. Some degree of perceptual organization is required to group and summarize raw 3D measurements into quantities that relate to viscosity. We suggest that the mid-level perceptual factors pool and organize local 3D estimates to create robust viscosity cues.

Together, these results indicate that despite the extremely complex physics underlying fluid flow, we can predict viscosity perception using a small number of quite simple mid-level cues. A fascinating open question is how the visual system identifies which stimuli are liquids in the first place, and then extracts information from features that robustly generalize across an even wider range of contexts than tested here (e.g., when spatial scale or the liquid's density also vary). In the long run, models of viscosity perception should be combined with a 'front-end' that allows predicting viscosity perception directly from image sequences, presumably via 3D estimates. Such a model should also seek to predict the effects of lighting and the liquid's optical properties on perceived viscosity, although these effects are generally small (Van Assen and Fleming, 2016). One approach to liquid detection and feature selection would be through sophisticated—potentially innate (Spelke, 1994; Hespos et al., 2009; Hespos and van Marle, 2012; Battaglia et al., 2013; Rips and Hespos, 2015)—physics-like internal models that capture the typical behaviour of fluids. For example, Battaglia and colleagues suggest (Bates et al., 2015) that the visual system predicts liquids' future states through simulations based on internal models. They show that humans far outperform simple heuristics at predicting where and how liquids flow. Such 'intuitive physics' approaches could potentially account for the successes of viscosity constancy we observe in our experiments: an internal model could be fit to liquids in a wide variety of different poses and contexts. Nevertheless, a challenge for any type of model is to predict the partial failures of viscosity constancy that we observe (e.g., the differences between the columns in Figure 4.2B).

An alternative approach would be to learn features from observation using large quantities of training data (e.g., learning of optimal speed or disparity encoding from natural scene data, Burge and Geisler, 2014, 2015). Given the recent success of convolutional neural networks (CNNs) in predicting visual object recognition and its neural correlates (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014; Güçlü and van Gerven, 2015; Cichy et al., 2016) the visual system could likely learn to recognize liquids—and features diagnostic of viscosity—from sufficient training data. An interesting topic for future investigation is whether similar features emerge for both liquid detection and viscosity estimation. Indeed, it would be fascinating to test whether CNNs arrive at similar features to the ones we

identify here. Nevertheless, as such algorithms acquire their features through supervised training, a major challenge for their use as models of human perception is to explain where the training labels come from during human learning. We might associate certain ranges of viscosity with different liquids but are not explicitly taught a fine viscosity scale that relates across materials, and yet, as we find here, we eventually become surprisingly good and precise at identifying them across a wide range of conditions.

4.2 Methods

4.2.1 Stimuli

Two stimuli sets were used in the four experiments. Set 1 contains a pouring liquid scene that was simulated with 32 viscosity steps. Each 10 sec long animation was divided into six time periods of 1.67 sec each for Experiments 1 and 3. This results in a total of 192 stimuli (32 viscosities $\times$ 6 time periods). Set 2 consisted of eight different scenes each simulated with seven viscosity steps. The duration of each stimulus in Set 2 was the full 10 sec because (1) In Experiment 1 we found that time had very little effect on perceived viscosity, and (2) due to the very wide range of speeds across scenes, there were long time periods for some scenes with viscous liquids, where the liquid had not yet entered the scene, which obviously would have made viscosity estimation impossible. Thus, Set 2 contained 56 stimuli (7 viscosities $\times$ 8 different scenes).

Simulation

Stimuli were generated using RealFlow 2014/2015 (V. 8.1.2.0192/V. 9.1.2.0193; NextLimit Technologies, Madrid, Spain). The pouring liquid scene (Experiments 1 and 3) consisted of 32 different viscosities ranging from 0.001 to 80.30 Pa·s. In Experiments 2 and 4, seven different viscosities were tested ranging from 0.004 Pa·s to 7.74 Pa·s in eight different scenes. Viscosity values were selected from a logarithmically spaced scale of 64-steps between 0.001 Pa·s and 100 Pa·s (equivalent from water to molten glass). RealFlow provides multiple particle solvers; in this case the “Hybrido” particle solver was used, making it possible to specify the dynamic viscosity of the liquids in real physical units (Pa·s). Hybrido is a FLIP (Fluid-Implicit Particle) solver using a hybrid grid and particle technique to compute a numerical solution to the Navier-Stokes equations describing viscous fluid flow (Bridson, 2015). Discrete particles carry all information for the fluid simulation, but the solution to the equations is carried out on a grid. Once the grid solve is complete, the particles gather the information required from the grid to move forward in time to the next frame. Finally, a

meshing algorithm uses the particles to calculate the fluid boundary. When visible artifacts occur it is mostly due to this stage where the mesh is being calculated, not the underlying physics solver. The density of the liquids was held constant at one kilogram per litre. The number of particles varied across scenes, with a maximum of roughly 5 million particles. All scenes were simulated in a space of roughly one cubic meter. Gravity was the main external force acting on the liquid, however in some cases an additional noise force field was used to achieve better scene-liquid interaction. The simulated animations had a total duration of ten seconds (300 frames at 30fps).

Rendering

The render engine used to generate the final image frames was Maxwell (V. 3.0.1.3; NextLimit Technologies, Madrid, Spain). The images were rendered at an 800×600 resolution and the scene was lighted using an HDR light probe depicting a beach scene (from the Maxwell Resource Library by Dosch Design). The liquid of the pouring scene (Experiments 1 and 3) was rendered with a translucent material. The liquid in all other scenes was of a green opaque material. Previous research has shown that optical material appearance of liquids barely influences viscosity judgements (Van Assen and Fleming, 2016).

4.2.2 Observers

Groups of twelve observers rated perceived viscosity in the first two experiments. In Experiments 3 and 4, where shape features were rated, two separate groups were formed. Each group rated only ten of the twenty shape features. In experiment 3 twelve observers participated in each group and in experiment 4 ten observers participated in each group. In total, over all four experiments 68 observers participated. The average observer age was 25.0 (SD = 4.82). 38 observers were female and 30 male. In the control experiments a total of 21 observers participated (4 in the brainstorming experiment, 8 in the semantic (word-list) matching experiment, and 9 in the factor rating experiment). The average age was 24.3 (SD = 3.89), 14 observers were female and 7 male. All observers gave written consent prior to the experiment and were paid for participating. All observers reported having normal or corrected-to-normal vision. Experiments were conducted in accordance with the Declaration of Helsinki and prior approval was obtained from the local ethics committee of Giessen University.

4.2.3 Procedure

Experiment 1 and 2: Rating viscosity

The experiments were performed on a Dell T3500 with a Dell U2412M 24-inch monitor using factory default settings, gamma of 2.2 and a resolution of 1920 $\times$ 1200 pixels. Matlab 2015a (v. 8.5.0.197613) and the Psychtoolbox library (v. 3.0.12) (Brainard, 1997; Pelli, 1997) were used to run the experiments. Observers completed a short training session before starting the experiment. The training was a single trial where the maxima and minima of the stimuli were presented and the observer could get acquainted with the interface, in case of experiment 2 and 4 all eight scenes were shown as well. During each trial, eight stimuli were shown with a rating bar transparently projected over each stimulus (Video C.2). There was no time limit and once all rating bars were set the observer could continue to the next trial. Corrections during the trial were possible and the observer was free to choose in which order the stimuli were rated. Each stimulus was repeated four times during the experiment but the position and combinations with other stimuli were chosen randomly for each trial.

Experiment 3 and 4: Rating shape features

The same setup as in Experiment 1 and 2 was used in Experiment 3 and 4. Experiment 3 and 4 where split up in two groups of observers, each rating ten of the twenty shape features. The stimuli were organized by viscosity on the screen. This was done to make it easier to rate the shape features. In the case of Experiment 3, 32 stimuli of the same time period were shown simultaneously, with Experiment 4, seven stimuli of one scene were shown. There were no repetitions in Experiment 3 and in Experiment 4 every trial was shown twice, in random order. Each shape feature was presented in the top left of the screen and an additional description of the shape feature was provided for clarity. All experiments were performed in German and have been translated to English for presentation here, see Table C.1 for a full list of shape features and descriptions.

Control experiment 1: Brainstorming new word list

We asked four observers to brainstorm 'shape features' while viewing videos of the pouring liquids (full 10s duration). There was a short training stage in which we explained the concept of shape features with examples using cars and plants. We carefully used examples that would not overlap with features in liquids. Individually, each observer wrote down as many shape features as possible, after which the four observers were instructed to work

together to pick the most descriptive twenty features. This closely resembles the way we selected the features ourselves.

Control experiment 2: Semantic matching of word lists

In this experiment, eight observers were asked to rate the similarity between our original word list (A) and new words generated in control experiment 1 (B). The videos of the pouring liquids were shown to provide some context. For each word in one list, the observer had to select similar words from the other word list. Observers were not required to choose similar words if there were none, and a maximum of three similar words for each item was allowed. The similarity of each of the matching words was then rated as 'high similarity', 'intermediate similarity', and 'little similarity'. This experiment was performed in both directions, so wordlist A was matched with wordlist B and vice versa. This enabled us to judge the similarity between the two word lists.

Control experiment 3: Factor ratings

In this control experiment we asked nine observers to rate the four factors (Distribution, Irregularity, Rectilinearity and Dynamics) directly, instead of the 20 features. Apart from this, the experimental procedure was the identical to the main experiments 3 and 4 in which the 20 features were rated.

4.2.4 Measurement models

2D Motion flow model

Kawabe et al. 2015 showed that the mean speed of optical flow is highly predictive of perceived viscosity. To evaluate whether motion cues are able to predict viscosity in our stimuli, we used the same iterated pyramidal Lucas-Kanade method (**?**) to calculate optical flow. We found that flow speed correlated poorly with perceived viscosity in the pouring liquids scene ($R^2 = 0.01$, $F(1,190) = 2.297$, $p = 0.13$). There are at least two possible reasons for this: first, the liquid was translucent, which could hinder optical flow computations; second, stimuli of intermediate viscosity tend to fold, which yields strong contours amplifying optical flow in certain ranges of viscosity. Optical flow for the other eight different scenes and opaque liquids also did not perform very well ($R^2 = 0.23$, $F(1,54) = 16.24$, $p < .01$). Only when we perform regression analysis for each specific scene do we see that for some specific scenes optical flow is a good predictor, with an average of $R^2 = 0.75$, and minima an maxima between $R^2 = 0.53$ and $R^2 = 0.97$. The large variations in performance of the

motion predictor suggest that the visual system likely uses other cues in addition to speed to infer viscosity.

2D Image statistics model

Paulun et al. 2015 found that twenty simple 2D shape statistics derived from the liquids' silhouette predict perceived viscosity surprisingly well. The statistics include measurements of shape, area, curvature, spatial distribution and perimeter, among others. We applied the same measurements to our stimuli, having excluded frames where there was not enough liquid (fewer than 300 pixels, i.e, <0.06% of image) and areas with only one-pixel width (to avoid errors in the contour measurements). Paulun et al. did not apply a regression but simply took the mean of the normalized measurements. Without fitting they found the model predicted perception in their stimuli extremely well (r = .99, p < .001). We applied the model to the second stimulus set (eight scenes) and found a much poorer fit (R^2 = 0.29, F(1,54) = 22.27, p < .001). Like Paulun et al. we used only a single predictor, the mean of all normalized measurements across our eight scenes. Performing a regression for each scene independently yield highly variable performance, ranging from R2 = 0.01 to R2 = 0.99. This shows that in some cases simple 2D shape measurements are sufficient to predict viscosity very well. However such cues are not flexible or invariant enough to achieve similar performance across contexts. Generalizing the model to use all 20 features as separate predictors in a regression (rather than the mean across measurements) yields R^2 = 0.80, F(20,35) = 7.19, p < .001, compared to R^2 = 0.81, F(4,51) = 55.88, p < .001 for our 3D model with only four predictors. The difference in performance is likely due to the fact that the four 3D measurements generalize better across scenes and contain less covariance than the twenty 2D measurements.

3D Shape measurements model

One advantage of computer-simulated liquids is the generation of detailed 3D meshes of the liquids. From these, we derived four 3D measurements (Figure C.4) that were loosely inspired by some of the perceptual features in the regression model. Specifically, (1) mean absolute curvature weighted by the shape index (Koenderink and Van Doorn, 1992), which emphasizes angular features, (2) the sum of absolute vertical normal coordinates, which captures the tendency of liquids to form horizontal planes as they spread out, (3) the vertical position of the centre of mass, which tends to be higher when the liquid piles up, and (4) total absolute curvature of the liquid, which tends to be large when the surface has many local convolutions. As the pouring liquids sequence is divided into six periods,

we compensated for large differences in mesh size over time by normalizing the median value of each feature over the different time periods. This was not necessary for the stimuli used in Experiment 2 and 4 where the entire 10 second time sequence was shown and we could simply take the average measurement value over 300 frames. We did apply normalisation of each measurement across the scenes. To compare performance with the other models we applied a multiple linear regression on the second stimulus set with eight scenes. We find the 3D model performs much better than the other two ($R^2 = 0.81$, $F(4,54) = 55.88$, $p < .001$). When we apply the regression separately for each scene the mean is $R^2 = 0.98$ across scenes. It is important to note however, that this performance is achieved even though the mesh measurements do not correlate with the perceived feature ratings across contexts (mean $R^2 = 0.04$). This means that although a linear combination of the mesh measurements can explain perceived viscosity relatively well, there is no direct correspondence between these measurements and the features that our observers judged.

4.2.5 Quantification and statiscal analysis

All experiments were performed in Matlab using Psychtoolbox (v. 3.0.12) (Brainard, 1997; Pelli, 1997). All analyses were performed in R. The code is publicly available and can be downloaded here: http://doi.org/10.5281/zenodo.1136202. All dependencies of external packages used in R are clearly documented in the code. No observers were excluded from the analysis.

Factor analysis

We performed a maximum likelihood factor analysis using the R 'psych' package. To determine how many factors there are in the dataset, we applied Horn's parallel analysis (Horn, 1965). We applied the Harman method to calculate the scores, applying the loadings to the actual data.

Representational Similarity Analysis (RSA/RDMs)

For a comprehensive description of Representational Similarity Analysis, we refer to [19]. The representational dissimilarity matrices (RDMs) in Figure C.3C were calculated using the Euclidean distances between observations in the 4D factor space, with each dimension representing one of the factors. The linear regression performed to quantify similarity is

performed on the lower triangles of the two matrices (i.e., diagonal and upper triangle excluded from analysis).

Principal Component Analysis

To perform PCA on the raw image similarity space (Figure 4.4A), we halved the images to 400 x 300 pixels, and converted the images to grayscale using the following conversion values (0.2989 * R + 0.5870 * G + 0.1140 * B). The resulting dataset contains 36 million dimensions for each of the 56 stimuli (i.e., over 2 billion observations in total). We include this PCA data as a separate, comma separated file.

4.2.6 Data and software availability

All data, analysis code, and stimuli are available on Zenodo at http://doi.org/10.5281/zenodo.1136202. Any questions should be directed to the Lead Contact (mail@janjaap.info).

Chapter 5

Estimating viscosity with neural networks

This chapter is based on findings in an ongoing project.

In computer vision neural networks are being applied on a wide variety of visual tasks. Most of the work using DNNs (Deep Neural Networks) is concentrating on predicting a predefined label with most optimal precision. In visual perception more human aspects are added to the challenge, introducing human error. Labelling large datasets with human defined labels is in many cases not an option. Here we investigated if relatively simple architectures can predict human errors in estimating viscosity of liquids while being trained on the physical truth. Perceiving intrinsic properties of liquids is a challenging visual task because of the complex behaviour and mutable nature of liquids. We simulated a training set of 2 million images or 100.000 animated sequences depicting in 10 different scenes with 16 different viscosities. The different scenes varied in liquid interactions, e.g. pouring or stirring, making some liquids easier or harder to interpret. We defined two networks, one for static stimuli and one for animated sequences of 20 frames. In the case of animated stimuli we applied a slow-fusion technique where parallel pathways with different temporal inputs slowly fuse into one. We asked observers to rate the viscosity for a subset of the stimuli. We find that across scenes there are big differences in observer performance. Using Bayesian optimization in combination with the viscosity ratings we specified the hyper parameters for the network, e.g. kernel sizes, learning rates. This means that the network is trained using physical viscosity labels but the hyper parameters optimize towards are perceived viscosity predictions. Previous work has shown that next to motion cues, mid-level shape features are very predictive of human viscosity estimations. Here we find that both network and human observer show a high increase in performance

when motion cues are available. This demonstrates that both human observers and network utilize the same cues to make viscosity estimations. The network presented here is the best image based viscosity predictor we have encountered so far.

5.1 Introduction

In previous chapters we established that the perception of liquids is a very challenging and intriguing problem. Liquids can have many different appearances because of their highly mutable shapes, which are visually influenced by internal and external forces. For our perception we use a wide range of cues, which mostly can be categorized as optical cues, shape cues and motion cues. Our visual system reweights the influence of each type of cue based on the available information in the image. If there is no motion information our judgments tend to use shape cues, if the motion and shape cues are not very pronounced, we might assign more meaning to the optical cues. When estimating viscosity, shape and motion cues tend to be dominant.

Within the cue groups there are specific features that can be informative of liquid properties. Optical cue driven features (e.g., color, sub surface scattering, glossiness), shape cue driven features (e.g., angular, spread, spiraling), and motion cue driven features (e.g., optical flow, pulsing, localized motion differences) provide a very large feature space. In order to navigate this feature space we need to be able to detect the most informative features in an image. For our visual system this seems a simple task; we can easily rate how angular or blobby a shape is. More challenging is to implement this quality in a model . What correspondences in an image should a model utilize to be able to estimate blobbiness? Blobbiness is a mid-level concept that can occur localized in various scales and orientations. Previous 3D shape metrics have attempted to capture the essence of certain shape features, but they did not generalize across contexts. Here we hope to obtain more insight using an image based neural network model.

The recent success of convolutional neural networks (CNNs) in predicting visual object recognition and its neural correlates are very impressive (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014; Güçlü and van Gerven, 2015; Cichy et al., 2016). This has inspired many scientists to apply these techniques on difficult visual tasks. The problem with most neural networks is to obtain conclusive results on what the network is actually doing, what it has learned to do, and how to represent this in an understandable manner. The results presented here are part of an ongoing research project. For now we only are able to report the performance of the model and not yet a thorough analysis of the inner workings of the model.

The neural networks were trained on a dataset of 100.000 computer-generated liquids. The training labels corresponded with the different viscosity steps that were simulated. We developed two networks: one using static images as input and the other an animation of twenty frames using a slow-fusion technique (Karpathy et al., 2014). Human observers rated a subset of 800 stimuli and assigned them with perceived viscosity labels. The network, which trains on physical viscosity labels, was optimized using Bayesian optimization. The Bayesian optimization was specifically used for optimizing the hyper parameters (e.g., learning rate, kernel sizes) for the 800 human labels. This means that the network trained on physical labels but the architecture is optimized for perceived labels. Training was relatively short with only 30 epochs. When training will be extended the networks tends to converge on the physical labels and away from the human perceived labels. Observers showed great differences in perceived viscosity across scenes. This is mostly due to the lack of clear shape cues in the small 64×64 pixel images. The task here is to cause the network to make similar errors in viscosity estimations.

As the project progresses we hope to identify specific combinations of filter patterns which correlate with human feature concepts we previously have identified as being very predictive of perceived viscosity (Van Assen et al., 2018). The networks trained here were provided with labels of the physical viscosity (supervised learning). Human observers have never been taught these physical viscosity labels (unsupervised learning) and despite this we are surprisingly good at identifying liquids and their properties in a wide range of contexts. We would like to emphasize that we identify neural networks as a tool to gain new insights on high dimensional problems. This doesn't mean that an accurately performing neural network solves viscosity perception. We want to learn how neural networks solve this problem to obtain knowledge on possible workings of our visual system. The visual system might still have very different inner workings or representations of concepts compared to DNNs (Deep Neural Networks). There may not be one, but many solutions to viscosity perception, of which the human visual system is the most efficient one we have encountered so far.

5.2 Methods

5.2.1 Stimuli

Large amounts of training data are necessary to properly train neural networks. With more variation in these training sets networks tend to generalize better across newly introduced liquid scenes. Here we generated a stimulus set of ten different scenes. Each

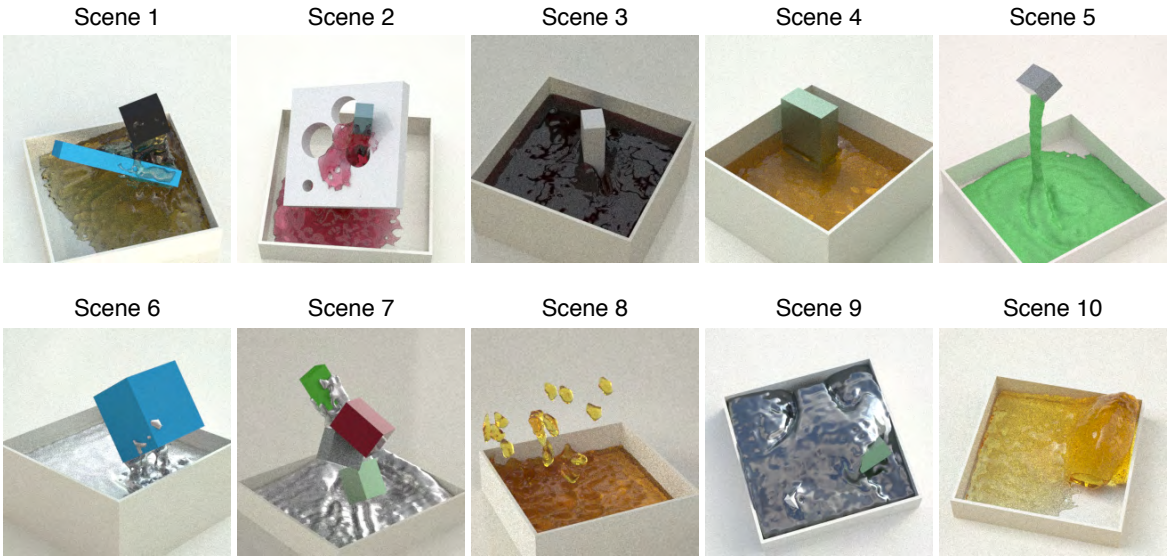


Figure 5.1: Stimuli overview with in this case the ten different scenes simulated at viscosity step 8 or 0.074 Pa·s. Different liquid interactions were simulated, as pouring, rain, stirring and dipping. Optical material properties and illumination maps were randomly assigned with the white plane and square reservoir staying constant. See Video D.1 for the animated version.

scene had it's own specific liquid interactions (e.g. dipping, rain, stirring, spraying over various geometries). Each scene was simulated with sixteen different viscosities steps from 0.001 Pa·s to 10 Pa·s (roughly similar to a range from water to molasses). Each scene and viscosity were simulated several times to create the large amount of necessary images. Parameters such as liquid emitter velocity, emitter direction, initial liquid volumes and scene geometries that interact with the liquid were randomized. This process was repeated 125 times and of these 125 variations five different render variations were made, changing illumination maps, optical material properties of both liquid and scene geometries, and camera position. Twenty sequential frames were rendered providing moving stimuli of a 0.67 second duration (30 frames per second). This resulted in a training set of 20.000 unique simulations and 2 million images (10 scenes $\times$ 16 viscosities $\times$ 125 scene variations $\times$ 5 optical variations $\times$ 20 frames). Figure 5.1 shows an impression of the different scenes. A subset of this set was used for experiments with human observers, 800 in total (10 scenes $\times$ 16 viscosities $\times$ 5 scene variations).

Simulation

The stimuli were generated using RealFlow 2015 (V. 9.1.2.0193; NextLimit Technologies, Madrid, Spain). Viscosity values were selected from a logarithmically spaced scale of 16-steps between 0.001 Pa·s and 10 Pa·s. The "Hybrido" particle solver was used which

simulates the dynamic viscosity of the liquids in real physical units (Pa·s). Hybrido is a FLIP (Fluid-Implicit Particle) solver using a hybrid grid and particle technique to compute a numerical solution to the Navier-Stokes equations describing viscous fluid flow (Bridson, 2015). A meshing algorithm uses the particles to calculate the fluid boundary and creates a mesh. The density of the liquids was held constant at one kilogram per litre and gravity was the only simulated external force. The simulated animations had a total duration of four seconds (120 frames at 30fps). Only the last twenty frames were used for the final stimuli. Each scene had specific parameters that were randomly assigned for each simulation. The random values were drawn from predefined ranges to limit the occurrence of artefacts. For example, in some scenes the liquid emitter was changing position during simulation, where the initial position, size, rotation, and trajectory of the emitter were randomly assigned. The simulation space for each scene was one cubic meter. The white container in the scenes was placed on the simulation border making this container 1m^2 large. The height of the container changed depending on the scene.

Rendering

The render engine used to generate the final image frames was Maxwell (V. 3.0.1.3; NextLimit Technologies, Madrid, Spain). This render engine is build into Realflow 2015. The images were rendered at a 256 $\times$ 256 resolution where the sampling rate was kept lower than normal to save time generating the 2 million images. Because of the lower sampling rate some noise was detectable. The illumination maps were randomly assigned from a set of 234 light probes, which were normalized and white balanced. The illumination maps came from different sources, some from scientific databases (Debevec, 2008; Adams et al., 2016). There were two categories of materials, solids (12) and liquids (13), which were randomly assigned to the different objects in a scene.

5.2.2 Observers

Eight observers participated in each of the four experiments. The average observer age was 25.3 (SD = 2.95). Twenty observers were female and twelve observers were male. All observers reported having normal or corrected-to-normal vision. All observers gave written consent prior to the experiment and were paid for participating. Experiments were conducted in accordance with the Declaration of Helsinki and prior approval was obtained from the local ethics committee of Giessen University. With enough time observers could participate in two experiments, both the static and the moving condition with stimuli of the same size. In this case the static condition always was performed first since these

stimuli contained more restricted information about the liquid. For the 64px condition this happened six times and for the 256px condition five times, resulting in 21 unique observers.

5.2.3 Procedure

Four experiments were performed, all with the same experimental setup. Only the stimulus resolution and moving/static conditions changed: Experiment 1, 256 $\times$ 256 pixels, static stimuli; Experiment 2, 64 $\times$ 64 pixels, static stimuli; Experiment 3, 256 $\times$ 256 pixels, moving stimuli; Experiment 4, 64 $\times$ 64 pixels, moving stimuli. The experimental setup was a Dell T3500 system with Matlab 2015a (v. 8.5.0.197613) and the Psychtoolbox library (v. 3.0.12) (Brainard, 1997; Pelli, 1997). The stimuli were displayed on an Eizo ColorEdge CG277 27-inch monitor with a resolution of 2560 $\times$ 1440 and a gamma of 2.2. A training session was performed to get the observers acquainted with the task and interface. The training session consisted of four trials in which the maximum and minimum viscosity were included. The task was to simply rate the viscosity, which was done with a horizontal rating bar below the stimulus. The rating bar marker reacted to the x-position of the mouse. Once the marker on the rating bar was at the desired position, the observer could confirm the answer by pressing 'space' on the keyboard after which a new trial was loaded. In total 800 trials were tested, 10 scenes $\times$ 16 viscosities $\times$ 5 variations. There was no time limit for the trials.

5.2.4 DNN Architecture

Convolutional Neural Networks (CNNs) are established as a powerful class of models for visual recognition problems. The last few years many new techniques have been developed enabling researchers to solve a larger range of visual problems with more precision. The novelty in the network presented here is that it is designed to predict human-like behaviour while being trained on physical viscosity labels. With human behaviour we mostly refer to making similar prediction errors. Figure 5.2 shows the architectures we applied in this study. The two networks presented here, for static and moving stimuli, are not very deep. More advanced, deeper, networks are able to predict the physical labels of viscosity better, but are not mimicking human estimations and errors at all. The networks here were trained using the Linux build of Matlab 2017b (v. 9.3.0.713579).

Static stimuli

The single-frame architecture is relatively simple. A network dealing with the temporal domain is more challenging. Therefore the most optimal hyper parameters of the multi-frame network was used as a base for this single-frame architecture. The only difference was that the parallel pathways were removed. Using a shorthand notation the architecture is C (64, 3, 1)-R -P -C (32, 5, 1)-N -P -C (100, 3, 1)-R -P -F (4096)-R -D -F (1). C(f, k, s) is a convolutional layer with *f* filters of $k \times k$ kernel size and *s* stride. F (n) is a fully connected layer with *n* nodes. P are max pooling layers with 2×2 regions and a stride of 2. R are ReLU layers as described in Nair and Hinton, 2010. D is a dropout layer with a dropout probability of 50%. The final layer is a regression output layer. The learning rate was set to $1.1105e^{-5}$, momentum to 0.43325, and L2 regularization to $4e^{-9}$. These were the most optimal settings according to the Bayesian optimization.

Moving stimuli

For the moving stimuli we applied a slow fusion model (Karpathy et al., 2014). There are parallel pathways that slowly fused over time providing the higher layers with more global information in both spatial and temporal domains. Each pathway had a specific part of the image sequence as input. Between the pathways there was an overlap of input images; in this case for the first convolutional layer the temporal extent T = 8 with stride 4 and for the second convolutional layer T = 2 and stride 2. The third convolutional layer had access to the full input range of 20 frames. The network was trained on continuous labels, not categorical, therefore a regression output layer was used.

The 800 stimuli of the experiments and scene 10 were excluded from the training set and were used for network validation. Bayesian optimization was used to determine the optimal settings for the hyper parameters; in this case the learning rate, L2 regularization, momentum, kernel sizes and filters for the three convolutional layers and the dropout probability of the dropout layer. The Bayesian optimization was fitted to the human labels of the 800 stimuli. This means that the hyper parameters were set to achieve lowest error predicting human labels, not the physical viscosity labels. Training was still performed with the physical viscosity labels. The Bayesian optimization ran for 52 iterations of 25 epochs after which the most optimal parameter settings were provided.

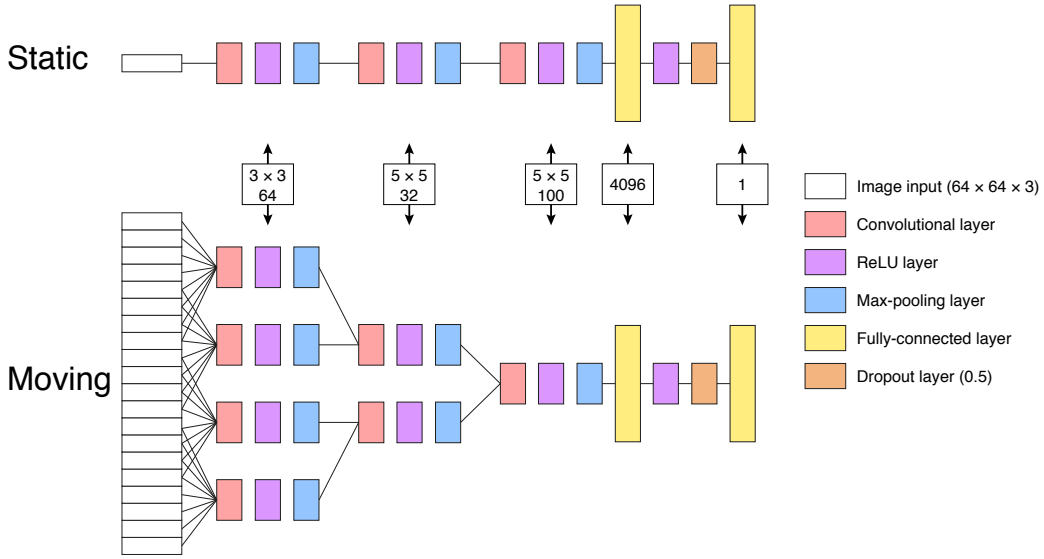


Figure 5.2: The two network architectures, one for static stimuli and one for moving stimuli. The hyper parameters were held constant for both static and moving stimuli networks. The only difference is that with moving stimuli there are parallel convolutional layers that slowly fuse over time. The dropout layer randomly sets input elements to zero with a 50% probability, against over fitting.

5.3 Results

5.3.1 64px vs. 256px

We will only report the results of the 64px (pixel) conditions because between the 64px and 256px conditions the differences are surprisingly small. Figure D.1 (static) and Figure D.2 (moving) show the differences and errors between the 64px and 256px conditions. The errors between observers are 70% (static) and 69% (moving); larger than the errors between the 64px and 256px conditions, which means that the differences are far below observer noise. Another reason not to further report on the 256px conditions is that the neural networks were only trained on 64px images; resulting in positive performance benefits.

5.3.2 Static stimuli

Figure 5.3 shows the results for Experiment 1 where static stimuli of a 64px size were rated. The first observation is that observers do not perceive the viscosity very accurately (R^2 = 0.18, F(1,158) = 34.68), $p < .001$). This is not in line with previous studies where observers were able to estimate viscosity very accurately (Paulun et al., 2015; Kawabe et al., 2015; Van Assen and Fleming, 2016; Van Assen et al., 2018). However, the stimuli used here are visually much more restricted, limiting the amount of clear shape cues. In this case the

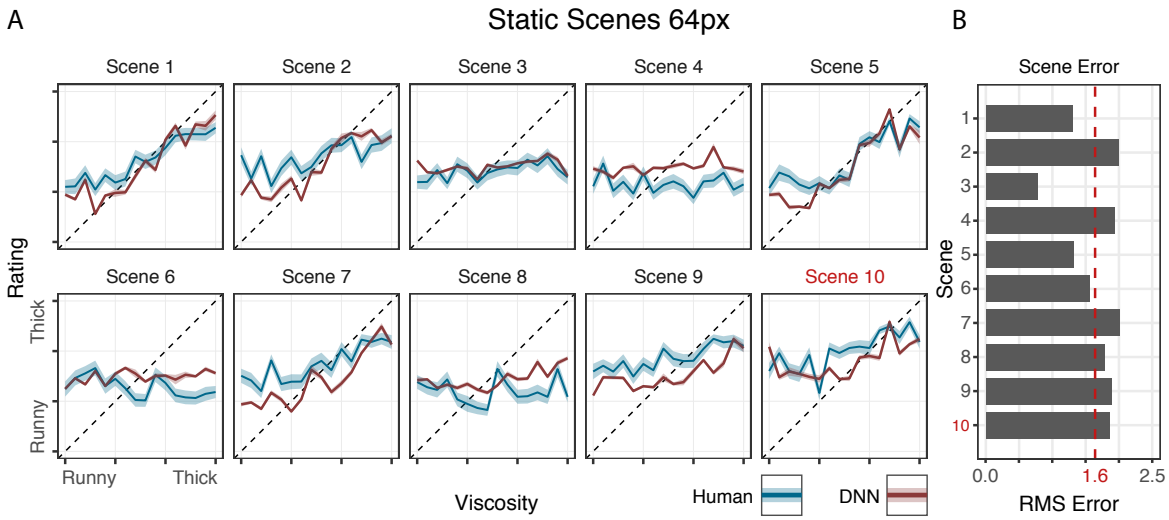


Figure 5.3: (A) Viscosity ratings for the 10 different scenes with static stimuli of a 64px size. The x-axis shows the tested viscosity steps (1-16). The y-axis shows the perceived/predicted viscosity steps. The error ribbons show the standard error of the mean (SEM). Blue lines are human viscosity ratings and red lines are DNN viscosity predictions. The dotted line shows the physical truth. The DNN was not trained on any stimuli predicted here, where scene 10 (red) was completely left out of the training set and tests generalization to other scenes. (B) The x-axis shows the Root Mean Square Error for each of the 10 scenes on the y-axis. This is the error between human observations and the DNN predictions. The dotted line shows the mean error across scenes.

fluctuations between small differences in viscosity are very noticeable which partially is caused by the low agreement between observers. The agreement is an average of 3.8 RMSE to the mean observer, errors of 24% of the total viscosity scale. See Figure 5.5A for a clear overview of observer consistency. Only when we average the ratings across all scenes we see a better linear fit ($R^2 = 0.82$, $F(1,14) = 62.12$, $p < .001$). This means two things: (1) observers find it hard to estimate viscosity accurately and consistently for individual stimuli, (2) averaged across scenes we are able to estimate general changes in viscosity.

The DNN model performs very well. The mean prediction error is only 1.6 RMSE or 10% of the viscosity step scale. An error of this size on a continuous scale is a very good result. By just looking at the scene plots we see that the DNN and human observers behave very similarly. The data plotted here is averaged over the five variations. Figure 5.5A shows the errors between observers and the DNN network for all 800 stimuli. The magnitude of the DNN errors is small compared to individual observers and the DNN and the observers are making similar errors as well. Averaged across all scenes and variations the DNN explains the perceived viscosity extremely well ($R^2 = 0.95$, $F(1,14) = 294.4$, $p < .001$). This network is not only making similar mistakes for individual stimuli but it predicts the overall perceived viscosity extremely well.

The generalization in this network is very good. Scene 10, the scene on which the network was never trained, shows that the DNN is only performing 13% above the mean error across scenes (Figure 5.3B). The network doesn't seem to be able to predict the specific fluctuations of the perceived viscosity, but it is not the worst performer.

Figure 5.4: (A) Viscosity ratings for the 10 different scenes with moving stimuli of a 64px size. The x-axis shows the tested viscosity steps (1-16). The y-axis shows the perceived/predicted viscosity steps. The error ribbons show the standard error of the mean (SEM). Blue lines are human viscosity ratings and red lines are DNN viscosity predictions. The dotted line shows the physical truth. The DNN was not trained on any stimuli predicted here, where scene 10 (red) was completely left out of the training set and tests generalization to other scenes. (B) The x-axis shows the Root Mean Square Error for each of the 10 scenes on the y-axis. This is the error between human observations and the DNN predictions. The dotted line shows the mean error across scenes.

5.3.3 Moving stimuli

With moving stimuli we find a big increase in perceived viscosity accuracy (R^2 = 0.68, F(1,158) = 335.3, p < .001). Figure 5.4 shows more individual characteristics for each scene, where scene 4, 6 and 8 seem to be especially difficult for viscosity estimation, even with motion cues. This can be demonstrated by the amount the slope of the linear fit should be adjusted to match the physical truth, which in these three cases is 96% more than with the other scenes. Exluding these three scenes from the regression, the performance increases substantially (R^2 = 0.87, F(1,110) = 743,p < .001). The overall trend of viscosity, averaged across scenes, shows a very good fit (R^2 = 0.96, F(1,14) = 322.7, p < .001), clearly the addition of motion helps to perceive viscosity. Especially when the images are only 64 $\times$ 64 pixels it is hard to perceive reliable liquid shape information. Individual observers confirm the

increase in performance. Figure 5.5B shows that the agreement across observers is higher and the errors are lower, diverging to an optimum.

The DNN shows a similar increase in performance. The network architecture here really seems to latch onto the additional temporal motion information. The mean prediction error is only 1.4 RMSE or 9% of the viscosity step scale. Since the network mimics the human patterns even better, and the observers became more precise, the network became better in predicting the physical viscosity as well (R^2 = 0.98, F(1,14) = 786.5, p < .001). It even outperforms the human observers a bit. Figure 5.5B shows the errors between observers and the DNN network for all 800 stimuli. The magnitude of the DNN errors has decreased even more and the correlation, the similarity of the errors is much better; it outperforms individual observers. Averaged across all scenes and variations the DNN explains the viscosity extremely well (R^2 = 0.97, F(1,14) = 466, p < .001). It is impressive how good this network is showing similar behaviour as human observers. This is a network trained in the physical truth, the dotted line in the plots and yet it converges on patterns much closer to human performance.

The generalization is a bit poorer for the moving condition where scene 10 has 32% more error than the mean. The prediction seems especially off for the first data point, the runniest liquid, for the other datapoints it almost perfectly predicts the non-monotonic pattern until viscosity eight. The network still performs worse for other scenes suggesting that the network still generalizes relatively well.

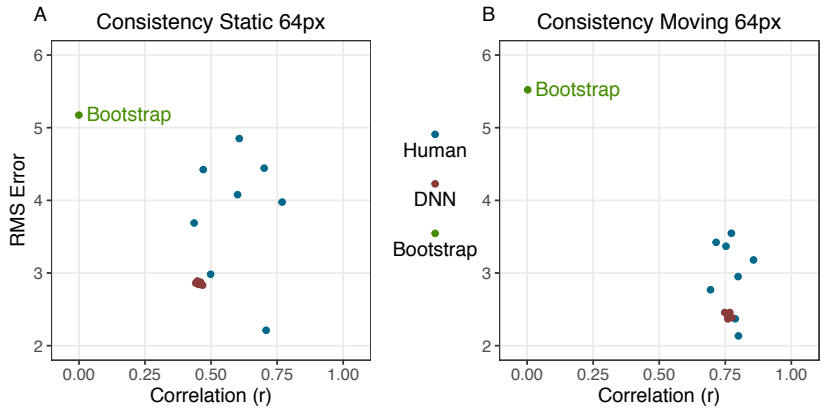


Figure 5.5: (A) Observer and DNN cosistency with static 64px stimuli. With on the x-axis the correlation and on the y-axis the Root Mean Square Error in relation to the observers' mean. This plot demonstrates not only the magnitude of the errors but the similarity of the errors as well. Blue dots are individual observers, red dots are individually trained DNN networks and the green dot shows a bootstrapped estimate of random performance based on 1000 random draws. (B) Exactly the same plot only with the moving 64px stimuli condition.

5.3.4 Optical flow

The stimuli are very restricted in how much visual detail they provide about the liquid. The images are small and are simulated with a relatively low spatial resolution. A large proportion of liquid details are not available in comparison with stimuli used in previous studies (Van Assen and Fleming, 2016; Van Assen et al., 2018). This is probably the reason we see such an increase in performance between the static and moving conditions. This increase is much larger compared to results presented in Van Assen and Fleming, 2016 where both static and moving stimuli were tested as well. We think this is caused by the restrictions in resolution, making the shape features less reliable and informative. This would also explain a much better performance for the moving conditions, where viscosity estimations might rely much more on cues that relate to movement. To test this we applied an optical flow analysis similar to Kawabe et al., 2015 to measure the liquid movement in each stimulus. The optical flow measurements are a bit less reliable with these stimuli since there is quite some pixel noise. To correct this a threshold was used to cut off very low movement created by the noise, however this might also partially exclude slow liquid movement.

We performed linear regressions for each scene where the optical flow is fitted with human performance (mean $R^2 = 0.47$). When we perform a regression of the optical flow with the DNN predictions we see a similar result (mean $R^2 = 0.48$). This is understandable since the DNN and human ratings are very similar. It is more interesting when we calculate the correlation across the R^2 values for each scene. This tells us when optical flow explains performance in- and decreases between the perceived and physical viscosity. This relationship is much better ($r(8) = 0.86$, $p < .01$). It is practically the same for the DNN predictions ($r(8) = 0.87$, $p < .01$). This indicates that optical flow does explain performance in- and decreases between the perceived and physical viscosity. This is a clear indication that observers heavily relied on the motion cues in our motion stimuli. As our network shows very similar patterns we conclude that our model utilizes the same motion cues.

5.4 Discussion

In this study we tried to train a neural network to predict human performance matching perceived viscosity to an image. The network is trained on physical viscosity labels but the hyper parameters are fine-tuned to best match 'human' characteristics. We asked observers to rate 800 stimuli in four different conditions, 64px static stimuli, 256px static

stimuli, 64px moving stimuli and 256px moving stimuli. The network was only trained on the 64px conditions.

We see very small differences between the two stimuli sizes. We expected that certain shape related features would have been more informative with the larger stimuli. The liquid simulations and meshes were generated with a relatively low resolution. Possible the used simulation resolution is not allowing clear localized mid-level shape cues. Because of this small difference we trained the networks with 64px images which resulted in a positively large performance difference during training.

The ten different scenes show nice unique characteristics. The contrast between performance in scene 4 and 5 is quite large, demonstrating a large range of scenes in which we can estimate viscosity well or not at all. This is especially useful as we want our model to focus on both mistakes and successes of the human observers. The network is trained for a relatively short period of 30 epochs. If we would continue training, the network would have converged on the physical training labels and moved away from perceived human labels. The network performs really well; in many cases it predicts scene specific characteristics. The variance in performance for each specific stimulus is quite large. Both DNN and human observers are not very constant, especially in the static stimuli condition. However, in both static and moving conditions the DNN outperforms the average individual observer errors. It is making smaller mistakes to the observer' mean than individual observers. This differs with the moving stimuli. The DNN is performing extremely well, making small or mostly similar errors as human observers (Figure 5.5B). With an overall predictive power of $R^2 = 0.95$ for static stimuli and $R^2 = 0.97$ for moving stimuli. We conclude this relatively shallow network clearly captures some of the human rating behaviour.

We can use a large range of cues to estimate the viscosity of liquids. Van Assen and Fleming 2016 showed that optical material appearance has little influence on viscosity judgements, which is mostly a mechanical property. Therefore it is to be expected that viscosity estimates are mainly driven by mechanical cues. Mechanical cues can be divided in two groups: shape cues and motion cues. In our stimuli tested here, the liquid shape information is scarcely available. This mostly is because of image resolution and liquid simulation resolution. The performance of human observers with static stimuli is therefore not very good ($R^2 = 0.82$), this is averaged across variations and scenes. We noticed a large increase of performance when observers rated the moving stimuli ($R^2 = 0.96$). One could argue that this is mainly due to the addition of motion cues. We measured the optical flow of the stimuli and conlcude that optical flow doesn't explain perceived viscosity directly (mean $R^2 = 0.47$). However, optical flow does explain performance in- and decreases

between the perceived and physical viscosity. This is a clear indication that observers heavily relied on the motion cues in our motion stimuli.

This study is still an on-going project. The neural network is now described as a black box that magically predicts perceived viscosity. The next step is to study the specific filter activations and how they change over time. The kernel sizes of the network are relatively small. The kernels are the patches that scan through the image. For the first layer they are only 3×3 pixels. Therefore at first sight it is very hard to interpret the filter activations and the temporal domain is making it even more complex. Although at this point we can't say much about the internal representations of the network, we can confirm that this is the best image based viscosity estimation model we have encountered so far.

Chapter 6

Conclusions

Materials are all around us. Every interaction with our surroundings is based on our knowledge of materials. Whether we are typing text on a keyboard, opening a coconut with a machete or laying down in a hammock, every interaction is based on our perception of the materials surrounding us. To be able to do this properly we need to be able to estimate material properties: is it soft or hard, sharp or blunt, runny or thick; properties that will influence our actions. How our visual system interprets these properties is key: it somehow decodes all this information from retinal images. We use visual information allowing us to estimate these properties consistently across many contexts. This information, which is invariant across contexts, is something we try to determine and measure in images. Which cues make us perceptually constant?

6.1 Estimating viscosity

In every study presented here we asked observers to rate or match viscosity. Viscosity is a physical property of liquids that make it look runny or thick. Water or syrup, paint or tar; all have large contrasts in viscosity. We are very good in estimating viscosity. When the visual information on the liquids is of high quality we can estimate viscosity with an accuracy of 99%. That leaves 1% of error for that single observer who was not paying attention for a few trials. When the detail of the liquids 'image decreases, we become less accurate, but by actively switching between different sources of image information, we are still able to get relatively good estimations. This suggests that not a single process, but multiple, connected processes allow us to estimate properties and recognize materials. We are able to actively switch between these processes and combine them for greater accuracy as well.

6.2 Viscosity constancy

We established that we can estimate viscosity very accurately. It is the more impressive we achieve this very consistently across a wide range of possible liquid appearances. Not only for liquids, but for many other material classes, or aspects thereof, we exhibit perceptual constancy (Maloney and Wandell, 1986; Bülthoff et al., 1995; Tarr et al., 1998; Kraft and Brainard, 1999; Anderson, 2011; Foster, 2011; Motoyoshi and Matoba, 2012). We can perceive viscosity being the same for a liquid that is being stirred in a bowl or smeared out over a plate. We notice when wine and an unnatural matte blue liquid are of the same viscosity. To be able to do this our brain needs to use causal inference (Biederman and Gerhardstein, 1993; Gilchrist et al., 1999; Riesenhuber and Poggio, 2000) to keep effects of gravity or scene geometries apart. Across a range of eight different scenes we find that observers have a constancy of 95%. This was measured between error differences per scene and the mean error across scenes. The main question is: what information do we use that determines this constancy across scenes. Image statistics are not able to achieve this, motion alone is not able to achieve this, even more advanced 3D shape metrics are not able to achieve this without an absurd amount of statistical reweighing. We seem to use higher-level features of shape and motion that stay invariant across scenes.

6.3 Mid-level shape and motion features

The impression that we heavily rely on mid-level features is gaining momentum in the field of material perception (Adelson, 2000; Anderson, 2011; Marlow et al., 2012; Paulun et al., 2015). Mid-level features are representations of concepts between low and high-level features. Low-level features are for example edge or contrast detectors; high-level features represent objects more in the context of a scene such as denim jeans, toothpaste or if it is windy. Mid-level features bring us from simple edge activations to the classified toothpaste. Mid-level features are more localized, regional cues (e.g., surface complexity, piling up or merging). Different regions of an object can have different mid-level features, only adding more contexts to interpret the object. We find that mid-level features are very predictive of perceived viscosity. There are many features that can play a role. Some features might be more descriptive for one particular scene than other scenes.

We found that a single element, comprised groups of features, can be descriptive of a wider range of feature habits (e.g., gaps, holes, folding, symmetry, texture, describe irregularity). Four of these mid-level feature groups could explain 88% of the perceived viscosity variance across eight completely different scenes. However, this was under the

strongest statistical restrictions; weights of the four feature groups were derived from a single scene. It is very acceptable that, on a scene-by-scene basis, we slightly adjust the weight of the most informative features to make a final estimate. If we allow for this freedom the 88% quickly rises.

6.4 The next step

What do we need to properly measure features in an image? The first challenge was to get better 3D shape estimates from 2D images (Binford, 1981; Pentland, 1986; Biederman, 1987; Feldman et al., 2013). Once a shape representation was available, a set of rules, heuristics, had to be defined by which a feature can be identified. For example spread, which was measured by calculating the proportion of surface that is pointing up or downwards. It turns out that with observation of a waterfall we still perceive spread, only now on the vertical plane and our metric would fail. It is important that these shape heuristics are invariant to scale, orientation, and can deal with segments of an object as well. This makes it much challenging to come with reliable metrics and requires large engineering efforts.

Neural networks might provide a different informative perspective, where combinations of image filters are specified, getting activated when specific features are dominant. These combinations or connections would be able to change pathways depending on different levels of activation, making the system much more dynamic. We demonstrated that neural networks can be very predictive of perceived viscosity. It is important to replicate human patterns of error; we were not looking for the model that predicts the label best, but searched for a model that can explain human failures: observers misinterpreting the labels. Only then we might be able to derive similar features from an image.

These are large steps to make. A nice intermediate step might be to use human observations of features and try to weigh the importance of these features using characteristics of the image (e.g., in this specific scene irregularity plays an important role because we can measure large contrasts in frequency information). By going through this process, new approaches for identifying and measuring mid-level features might arise.

Once we are able to quantify this type of information from images, it will have large implications for future research and technical applications. Applications that might influence society in a revolutionary manner. It will embed the human quality of generalization and constancy in technology.

References

Adams, W. J., Elder, J. H., Graf, E. W., Leyland, J., Lugtigheid, A. J., and Muryy, A. (2016). The southampton-york natural scenes (syns) dataset: Statistics of surface attitude. *Scientific reports*, 6:35805.

Adelson, E. H. (2000). Lightness perception and lightness illusions. *New Cogn. Neurosci*, 339.

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, 21(24):R978–R983.

Anderson, B. L. and Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of vision*, 9(11):10–10.

Barrow, H. and Tenenbaum, J. (1978). Computer vision systems. *Computer vision systems*, 2.

Bartell, F., Dereniak, E., and Wolfe, W. (1981). The theory and measurement of bidirectional reflectance distribution function (brdf) and bidirectional transmittance distribution function (btdf). In *Radiation scattering in optical systems*, volume 257, pages 154–161. International Society for Optics and Photonics.

Bates, C., Battaglia, P., Yildirim, I., and Tenenbaum, J. B. (2015). Humans predict liquid dynamics using probabilistic simulation. In *CogSci*.

Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332.

Beck, J. and Prazdny, S. (1981). Highlights and the perception of glossiness. *Attention, Perception, & Psychophysics*, 30(4):407–410.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115.

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human perception and performance*, 19(6):1162.

Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence*, 17(1-3):205–244.

Bouguet, J.-Y. (1999). Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. *OpenCV Documents*.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.

Bridson, R. (2015). *Fluid simulation for computer graphics*. CRC Press.

Bülthoff, H. H., Edelman, S. Y., and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 5(3):247–260.

Burge, J. and Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of vision*, 14(2):1–1.

Burge, J. and Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature communications*, 6.

Burnham, K. P. (2002). Information and likelihood theory: a basis for model selection and inference. *Model selection and multimodel inference: a practical information-theoretic approach*, pages 49–97.

Caudek, C. and Domini, F. (1998). Perceived orientation of axis rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2):609.

Chadwick, A. and Kentridge, R. (2015). The perception of gloss: a review. *Vision research*, 109:221–235.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific reports*, 6:27755.

Debevec, P. (2008). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *ACM SIGGRAPH 2008 classes*, page 32. ACM.

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., and Kersten, D. (2011a). Visual motion and the perception of surface material. *Current Biology*, 21(23):2010–2016.

Doerschner, K., Kersten, D., and Schrater, P. R. (2011b). Rapid classification of specular and diffuse reflection from image velocities. *Pattern Recognition*, 44(9):1874–1884.

Doerschner, K., Yilmaz, O., Kucukoglu, G., and Fleming, R. W. (2013). Effects of surface reflectance and 3d shape on perceived rotation axis. *Journal of vision*, 13(11):8–8.

Dong, J. and Chantler, M. (2005). Capture and synthesis of 3d surface texture. *International Journal of Computer Vision*, 62(1-2):177–194.

Dövencioğlu, D. N., Wijntjes, M. W., Ben-Shahar, O., and Doerschner, K. (2015). Effects of surface reflectance on local second order shape estimation in dynamic scenes. *Vision research*, 115:218–230.

Dror, R. O., Willsky, A. S., and Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11–11.

Emrith, K., Chantler, M., Green, P., Maloney, L., and Clarke, A. (2010). Measuring perceived differences in surface texture due to changes in higher order statistics. *JOSA A*, 27(5):1232–1244.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169.

Faul, F. and Ekroll, V. (2012). Transparent layer constancy. *Journal of Vision*, 12(12):7–7.

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., and Wilder, J. (2013). An integrated bayesian approach to shape representation and perceptual organization. In *Shape perception in human and computer vision*, pages 55–70. Springer.

Fleming, R. W. (2012). Human perception: Visual heuristics in the perception of glossiness. *Current Biology*, 22(20):R865–R866.

Fleming, R. W. (2017). Material perception. *Annual review of vision science*, 3(1).

Fleming, R. W. and Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception (TAP)*, 2(3):346–382.

Fleming, R. W., Dror, R. O., and Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of vision*, 3(5):3–3.

Fleming, R. W., Jäkel, F., and Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological science*, 22(6):812–820.

Fleming, R. W., Jensen, H. W., and Bülthoff, H. H. (2004). Perceiving translucent materials. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 127–134. ACM.

Fleming, R. W., Wiebel, C., and Gegenfurtner, K. (2013). Perceptual qualities and material classes. *Journal of vision*, 13(8):9–9.

Foster, D. H. (2011). Color constancy. *Vision research*, 51(7):674–700.

Gentner, D. and Stevens, A. L. (2014). *Mental models*. Psychology Press.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., and Economou, E. (1999). An anchoring theory of lightness perception. *Psychological review*, 106(4):795.

Güçlü, U. and van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27):10005–10014.

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

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, 8(6):280–285.

Hespos, S. J., Ferry, A. L., and Rips, L. J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychological Science*, 20(5):603–611.

Hespos, S. J. and van Marle, K. (2012). Physics for infants: Characterizing the origins of knowledge about objects, substances, and number. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1):19–27.

Ho, Y.-X., Landy, M. S., and Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, 19(2):196–204.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.

Jain, A. and Zaidi, Q. (2011). Discerning nonrigid 3d shapes from motion cues. *Proceedings of the National Academy of Sciences*, 108(4):1663–1668.

Jensen, H. W., Marschner, S. R., Levoy, M., and Hanrahan, P. (2001). A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518. ACM.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732.

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

Khaligh-Razavi, S.-M. and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS computational biology*, 10(11):e1003915.

Khang, B.-G., Koenderink, J. J., and Kappers, A. M. (2007). Shape from shading from images rendered with various surface types and light fields. *Perception*, 36(8):1191–1213.

Kim, J., Marlow, P. J., and Anderson, B. L. (2012). The dark side of gloss. *Nature neuroscience*, 15(11):1590–1595.

Koenderink, J. J. and Van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564.

Koenderink, J. J., van Doorn, A. J., and Pont, S. C. (2004). Light direction from shad (ow) ed random gaussian surfaces. *Perception*, 33(12):1405–1420.

Kraft, J. M. and Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences*, 96(1):307–312.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2.

Landau, B., Smith, L. B., and Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, 3(3):299–321.

Landy, M. S. and Graham, N. (2004). Visual perception of texture. *The visual neurosciences*, 1:1106.

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision research*, 35(3):389–412.

Liu, J., Dong, J., Cai, X., Qi, L., and Chantler, M. (2015). Visual perception of procedural textures: Identifying perceptual dimensions and predicting generation models. *PloS one*, 10(6):e0130335.

Maloney, L. T. and Wandell, B. A. (1986). Color constancy: a method for recovering surface spectral reflectance. *JOSA A*, 3(1):29–33.

Maloney, L. T. and Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, 3(8):5–5.

Marlow, P. J. and Anderson, B. L. (2015). Material properties derived from three-dimensional shape representations. *Vision research*, 115:199–208.

Marlow, P. J., Kim, J., and Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, 22(20):1909–1913.

Marlow, P. J., Kim, J., and Anderson, B. L. (2017). Perception and misperception of surface opacity. *Proceedings of the National Academy of Sciences*.

Marlow, P. J., Todorović, D., and Anderson, B. L. (2015). Coupled computations of three-dimensional shape and material. *Current Biology*, 25(6):R221–R222.

Motoyoshi, I. and Matoba, H. (2012). Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vision Research*, 53(1):30–39.

Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, 447(7141):206–209.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Nefs, H. T., Koenderink, J. J., and Kappers, A. M. (2006). Shape-from-shading for matte and glossy objects. *Acta psychologica*, 121(3):297–316.

Nicodemus, F. E. (1965). Directional reflectance and emissivity of an opaque surface. *Applied optics*, 4(7):767–775.

Nishida, S. and Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *JOSA A*, 15(12):2951–2965.

Norman, J. F. and Todd, J. T. (1994). Perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2):343.

Nusseck, M., Lagarde, J., Bardy, B., Fleming, R., and Bülthoff, H. H. (2007). Perception and prediction of simple object interactions. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, pages 27–34. ACM.

Oliva, A. and Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, 11(12):520–527.

Olkkonen, M. and Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of vision*, 10(9):5–5.

Ostrovsky, Y., Cavanagh, P., and Sinha, P. (2005). Perceiving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

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

Paulun, V. C., Schmidt, F., van Assen, J. J. R., and Fleming, R. W. (2017). Shape, motion, and optical cues to stiffness of elastic objects. *Journal of vision*, 17(1):20–20.

Pellacini, F., Ferwerda, J. A., and Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 55–64. ACM Press/Addison-Wesley Publishing Co.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4):437–442.

Pentland, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial intelligence*, 28(3):293–331.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision research*, 41(24):3145–3161.

Pont, S. C. and Koenderink, J. J. (2007). Matching illumination of solid objects. *Attention, Perception, & Psychophysics*, 69(3):459–468.

Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nature neuroscience*, 3:1199–1204.

Rips, L. J. and Hespos, S. J. (2015). Divisions of the physical world: Concepts of objects and substances. *Psychological bulletin*, 141(4):786.

Sawayama, M. and Nishida, S. (2015). Visual perception of surface wetness. *Journal of vision*, 15(12):937–937.

Schlüter, N. and Faul, F. (2014). Are optical distortions used as a cue for material properties of thick transparent objects? *Journal of vision*, 14(14):2–2.

Schmid, A. C. and 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.

Schmidt, F. and Fleming, R. W. (2016). Visual perception of complex shape-transforming processes. *Cognitive psychology*, 90:48–70.

Schmidt, F., Paulun, V. C., van Assen, J. J. R., and Fleming, R. W. (2017). Inferring the stiffness of unfamiliar objects from optical, shape, and motion cuesschmidt et al. *Journal of Vision*, 17(3):18–18.

Sharan, L., Liu, C., Rosenholtz, R., and Adelson, E. H. (2013). Recognizing materials using perceptually inspired features. *International journal of computer vision*, 103(3):348–371.

Sharan, L., Rosenholtz, R., and Adelson, E. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, 9(8):784–784.

Sharan, L., Rosenholtz, R., and Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of vision*, 14(9):12–12.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, 50(1):431–445.

Spröte, P., Schmidt, F., and Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific reports*, 6:36245.

Tarr, M. J., Williams, P., Hayward, W. G., and Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature neuroscience*, 1(4):275–277.

Todd, J. T. (2004). The visual perception of 3d shape. *Trends in cognitive sciences*, 8(3):115–121.

Todd, J. T., Norman, J. F., Koenderink, J. J., and Kappers, A. M. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, 26(7):807–822.

Van Assen, J. J. R., Barla, P., and Fleming, R. W. (2018). Visual features in the perception of liquids. *Current Biology*.

Van Assen, J. J. R. and Fleming, R. W. (2016). Influence of optical material properties on the perception of liquids. *Journal of vision*, 16(15):12–12.

Van Assen, J. J. R., Wijntjes, M. W., and Pont, S. C. (2016). Highlight shapes and perception of gloss for real and photographed objects. *Journal of vision*, 16(6):6–6.

Vangorp, P., Laurijssen, J., and Dutré, P. (2007). The influence of shape on the perception of material reflectance. In *ACM Transactions on Graphics (TOG)*, volume 26, page 77. ACM.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological bulletin*, 138(6):1172.

Wijntjes, M. W. and Pont, S. C. (2010). Illusory gloss on lambertian surfaces. *Journal of Vision*, 10(9):13–13.

Xiao, B., Walter, B., Gkioulekas, I., Zickler, T., Adelson, E., and Bala, K. (2014). Looking against the light: How perception of translucency depends on lighting direction. *Journal of vision*, 14(3):17–17.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624.

Zaidi, Q. (2011). Visual inferences of material changes: color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6):686–700.

Appendix A

Chapter 2

A.1 Supplemental videos

Video A.1

The six different viscosities, related to Figure 2.2.
http://www.janjaap.info/dissertation/video_a1.mov

Video A.2

The nine different optical materials, related to Figure 2.2.
http://www.janjaap.info/dissertation/video_a2.mov

A.2 Remaining rating results for all four variations

A.3 Full data set of the naming experiment

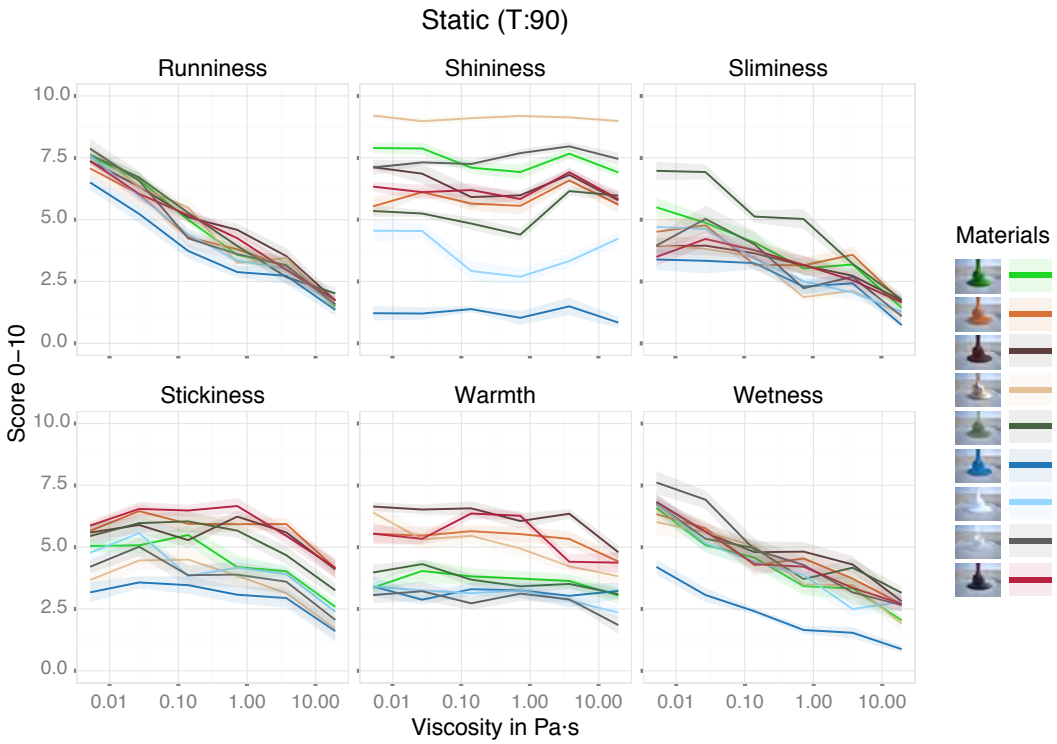


Figure A.1: Showing the liquid property rating results with static stimuli. Error envelopes represent standard error of the mean.

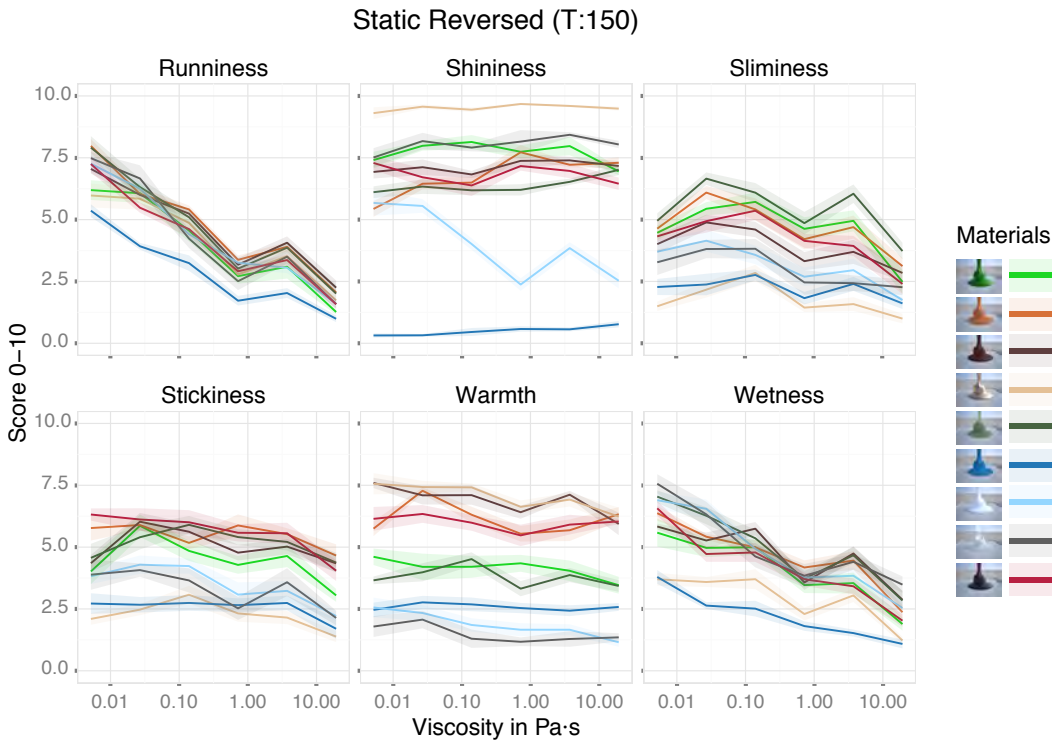


Figure A.2: Showing the liquid property rating results with static stimuli of the reversed condition. Error envelopes represent standard error of the mean.

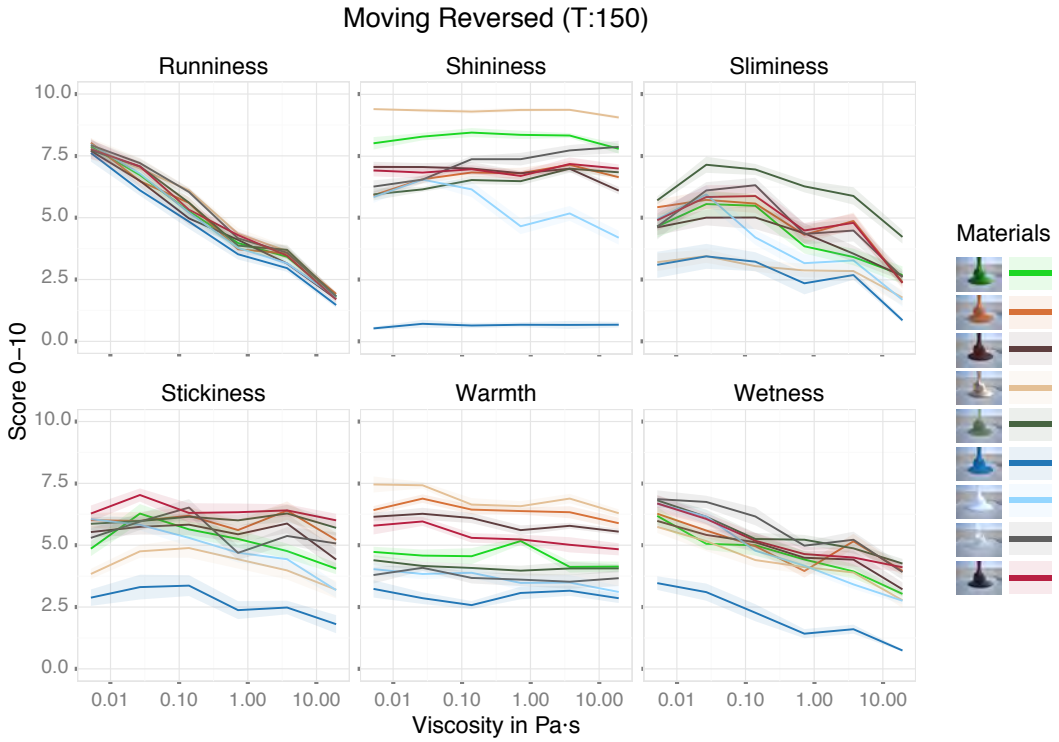


Figure A.3: Showing the liquid property rating results with moving stimuli of the reversed condition. Error envelopes represent standard error of the mean.

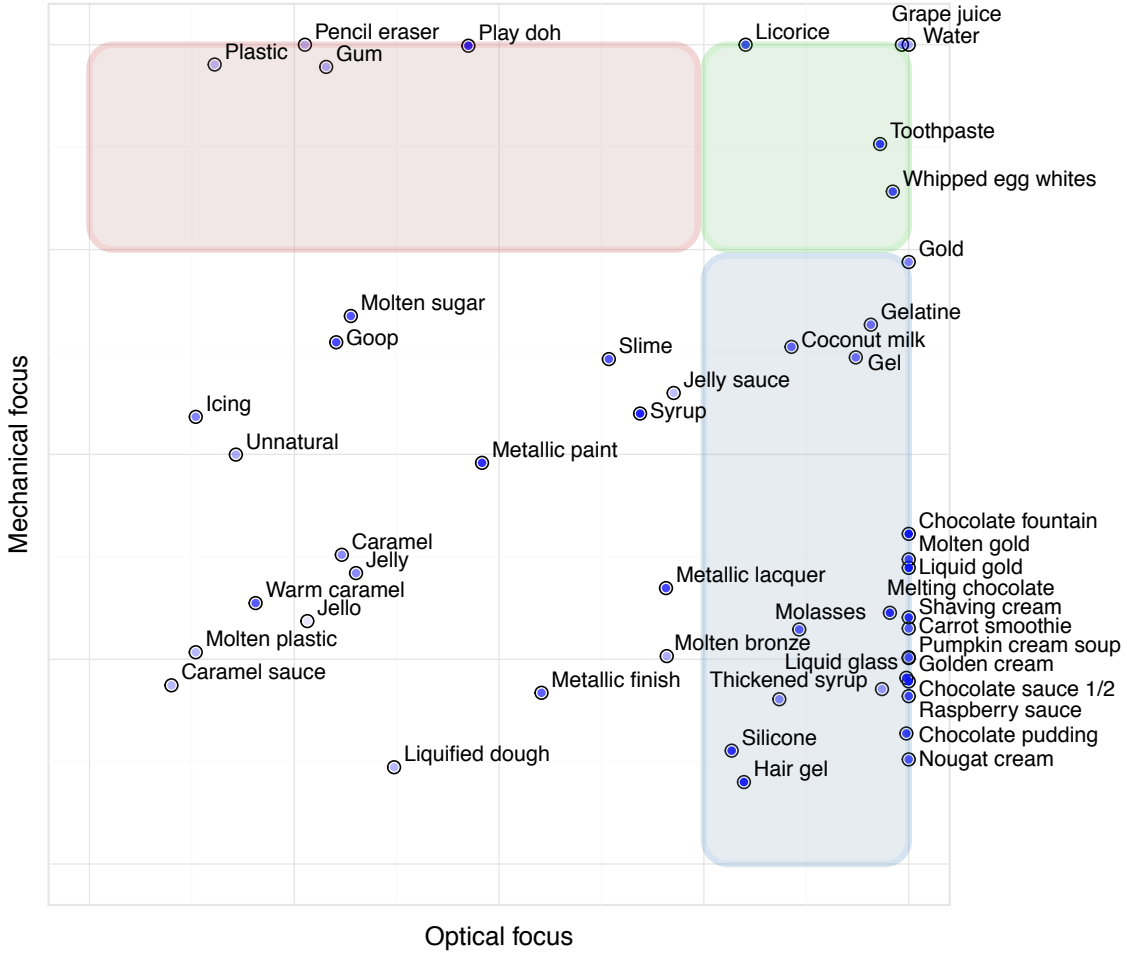
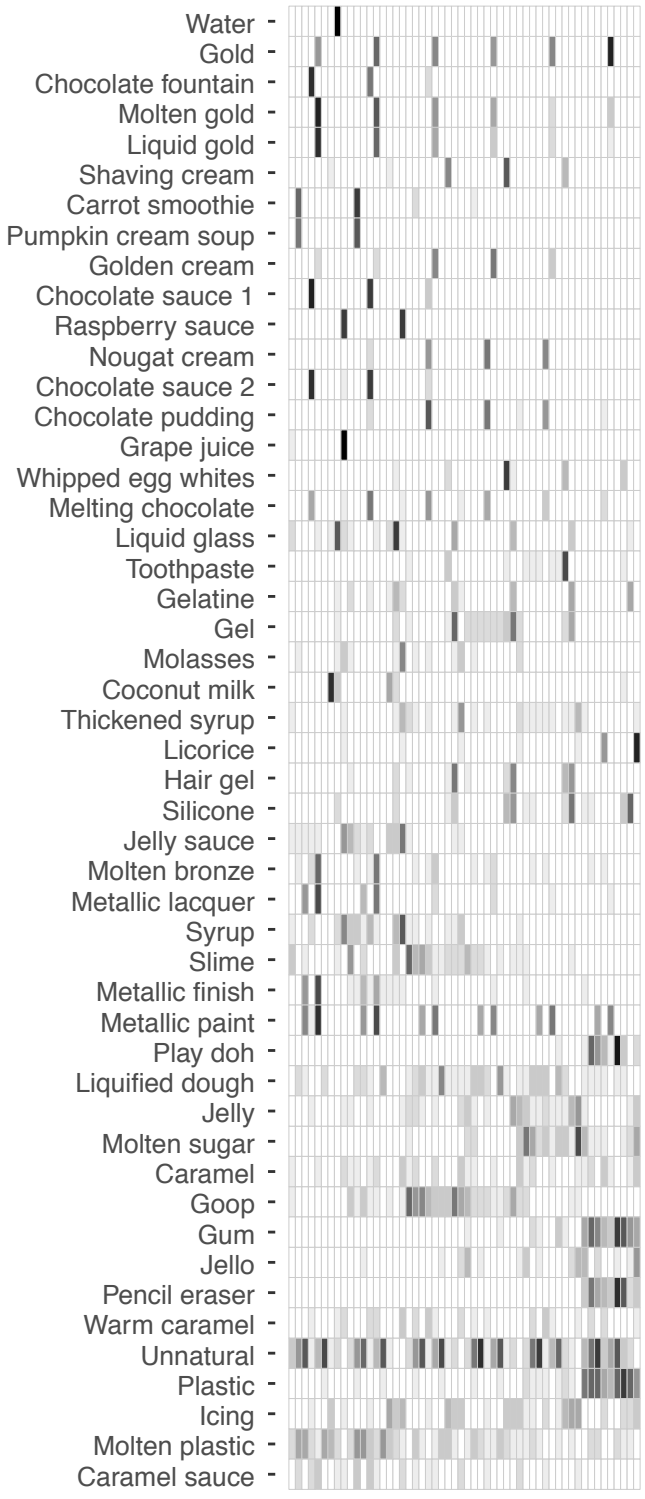


Figure A.4: Showing the data of Figure A.5 in 2D space. The intensity of the dots represents the confidence ratings. The names with high mechanical focus are in the red area, the names with high optical focus are in the blue area and the names with both high optical and mechanical focus are in the green area.



10	10	6.12
10	7.35	6.81
10	4.03	8.69
10	3.72	7.81
10	3.61	8.62
10	3.01	8.31
10	2.88	7.44
10	2.52	7.38
10	2.51	6.06
10	2.23	8.5
10	2.05	7.88
10	1.28	7.56
9.97	2.27	8.44
9.97	1.59	8
9.92	10	6.56
9.81	8.21	8
9.77	3.07	8.5
9.67	2.13	6.56
9.65	8.79	8.31
9.54	6.58	7.31
9.35	6.18	7.31
8.66	2.86	7.62
8.57	6.31	7.5
8.42	2.01	6.88
8.01	10	7.81
7.99	1	8.62
7.84	1.38	8.38
7.13	5.75	5.88
7.05	2.54	6.12
7.04	3.37	7.69
6.72	5.5	8.25
6.34	6.16	7.44
5.52	2.09	7.31
4.79	4.9	8
4.62	9.99	8.31
3.72	1.18	6
3.25	3.55	6.69
3.19	6.69	7.44
3.08	3.77	6.62
3.01	6.37	7.62
2.89	9.73	6.19
2.66	2.96	5.31
2.63	10	5.94
2.03	3.18	7.38
1.79	5	6.12
1.53	9.76	6
1.3	5.46	6.75
1.3	2.58	6
1	2.18	6

Optical focus
Mechanical focus
Confidence rating

Figure A.5: Showing the raw data of the name matching experiment.

Appendix B

Chapter 3

B.1 Supplemental videos

Video B.1

Showing the different stimuli of the matching experiments, related to Figure 3.1.
http://www.janjaap.info/dissertation/video_b1.mov

Appendix C

Chapter 4

C.1 Supplemental videos

Video C.1

Stimuli overview, related to Figures 4.1 and 4.2.
http://www.janjaap.info/dissertation/video_c1.mov

Video C.2

Trial examples experiments, related to Figures 4.1 and 4.2.
http://www.janjaap.info/dissertation/video_c2.mov

C.2 Supplemental figures

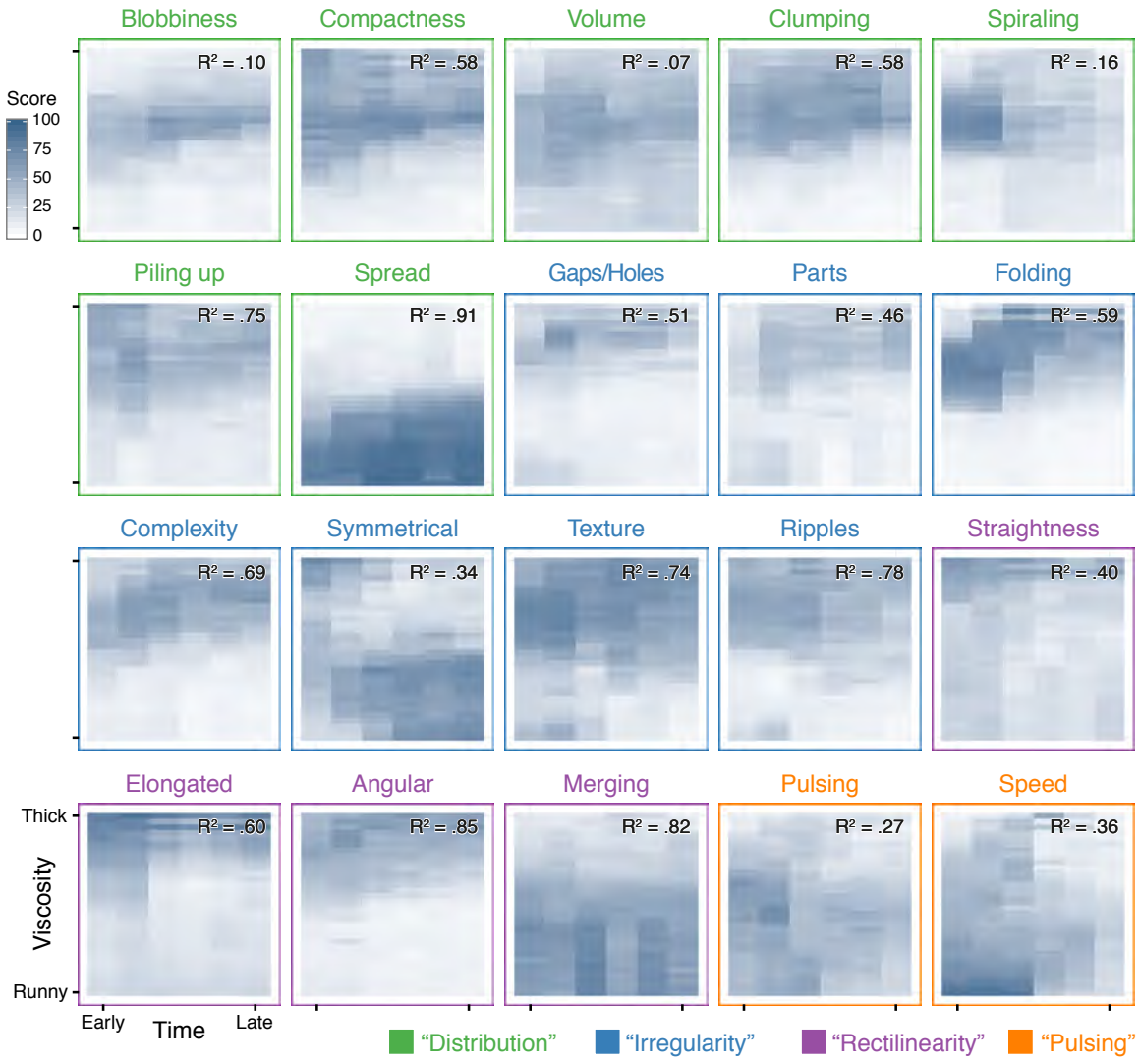


Figure C.1: Mean shape feature ratings of Experiment 3 for all twenty shape features, colour coded by the factors for which they have the largest weights. Y-axis: viscosity (32 viscosities, from runny 0.001 Pa·s to thick 80.30 Pa·s); X-axis: time (six time periods of 1.67 seconds or 10 seconds divided by six). The shape judgments generally varied in complex (often non-monotonic) ways as a function of viscosity and time period.

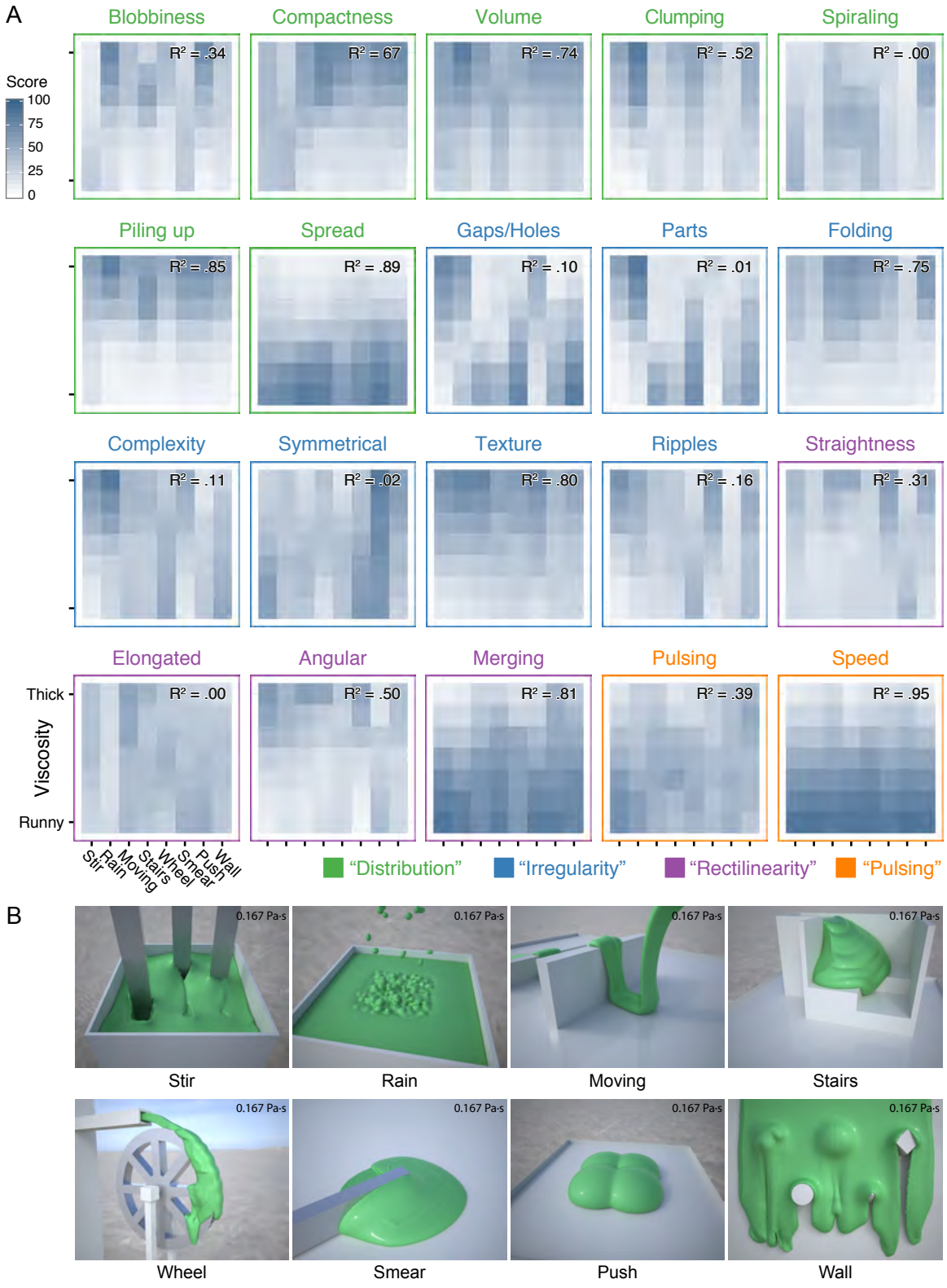


Figure C.2: (A) Mean shape feature ratings of Experiment 4 for all twenty shape features, colour coded by the factors for which they have the largest weights. Y-axis: viscosity (7 viscosities, from runny 0.004 Pa·s to thick 7.74 Pa·s); X-axis: eight different scenes. The shape judgements vary considerably on scene-by-scene basis, indicating that our simulated scenes capture a wide range of different liquid behaviours. (B) The eight scenes with an intermediate viscosity (0.167 Pa·s).

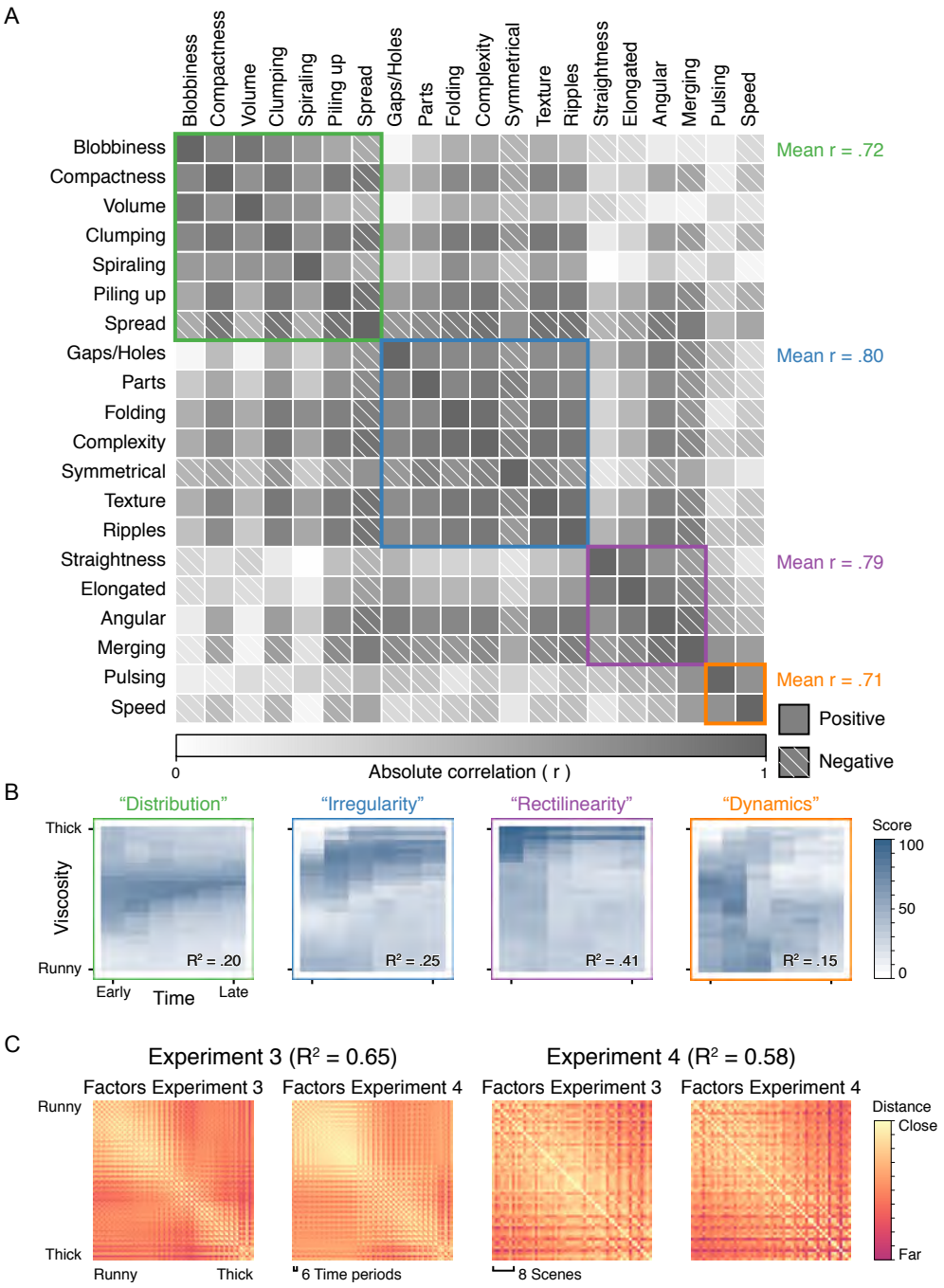


Figure C.3: (A) Correlations between perceptual feature ratings from Experiment 3; gray shade indicates absolute magnitude of correlation, stripes indicate negative correlations. Coloured frames indicate the four factors; *r*-values indicate mean absolute correlations between features within each factor, the overall mean absolute correlation is $r = 0.57$. (B) The four factors that result from applying the factor loadings to the twenty perceptual features. R^2 values indicate linear regression between each factor on its own and the viscosity ratings from Experiment 1. The complementary nature of the factors means that each on its own predicts only a small proportion of the variance, but combined in a multiple linear regression, they explain 97% of the variance in the viscosity ratings. (C) Representational Dissimilarity Matrices (RDMs) for the two regression models, derived from the Pouring scene (left) and the 8 scenes (right). In each case we apply factors derived from Experiment 3 and Experiment 4 and quantify how similar these factor spaces are. Colours represent the Euclidean distance between the corresponding pair of stimuli in the respective factor-space representation. The R^2-score indicates the explained variance between the lower triangles of the two matrices (i.e., diagonal and upper triangle excluded from analysis).

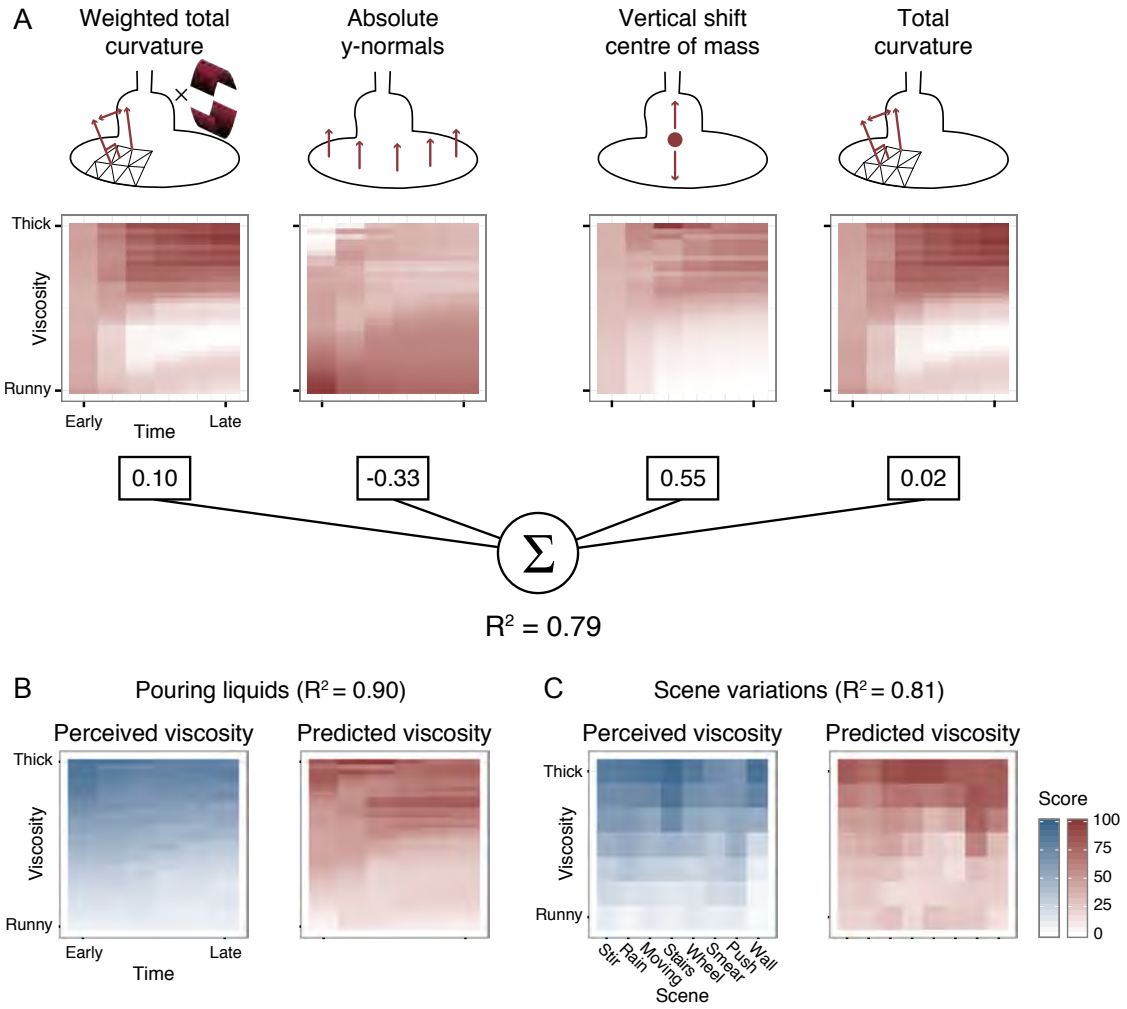


Figure C.4: (A) Schematics of the four mesh measurements used in the 3D stimulus-computable model, along with values for all stimuli in the pouring scene (Experiments 1 and 3). A regression model using the weights below each mesh measurement predicts 79% of the variance in viscosity ratings across all scenes. (B) Human viscosity ratings (blue) and mesh model predictions (mauve) for Experiment 1 (pouring liquids). R^2-score indicates regression for a model fit only to these data. (C) Human viscosity ratings (blue) and mesh measurements predictions (mauve) for Experiment 3 (eight scenes). R^2-score indicates regression for a model fit only to these data.

C.3 Supplemental tables

Shape Feature	Description
Symmetrical	How symmetrical the shape is
Compactness	How tightly arranged the shape is
Clumping	How much the shape features distinct clumps
Folding	How much does the shape fold back on itself
Pulsing	How much does shape change in a rhythmical repeating way
Merging	How the shape absorbs into itself
Speed	How quickly the shape changes or moves
Spiraling	How much does the shape change in a spiral movement or form
Elongated	How much is the shape stretched out in a single direction
Texture	How much does the surface have variations rather than being smooth
Blobbiness	How rounded or bulbous is the shape
Piling up	How piled up is the shape
Straightness	How much does the shape contain straight features
Complexity	How complex is the shape i.e. not simple
Spread	How spread out is the shape
Ripples	How much does the shape feature ripples
Gaps/Holes	How much does the shape shows gaps or holes
Parts	How much does the shape consist of multiple parts rather than one single part
Angular	How sharp or angular are the features of the shape
Volume	How voluminous is the shape

Table C.1: First column showing the twenty different shape features used in Experiment 3 and 4. Second column showing the corresponding description given as additional information to make it easier to rate the corresponding feature. The separation in the middle of the table shows the two different feature groups, each observer only rated ten features out one of these two groups.

Appendix D

Chapter 5

D.1 Supplemental videos

Video D.1

Stimuli overview, related to Figures 5.1.
http://www.janjaap.info/dissertation/video_d1.mov

D.2 Supplemental figures

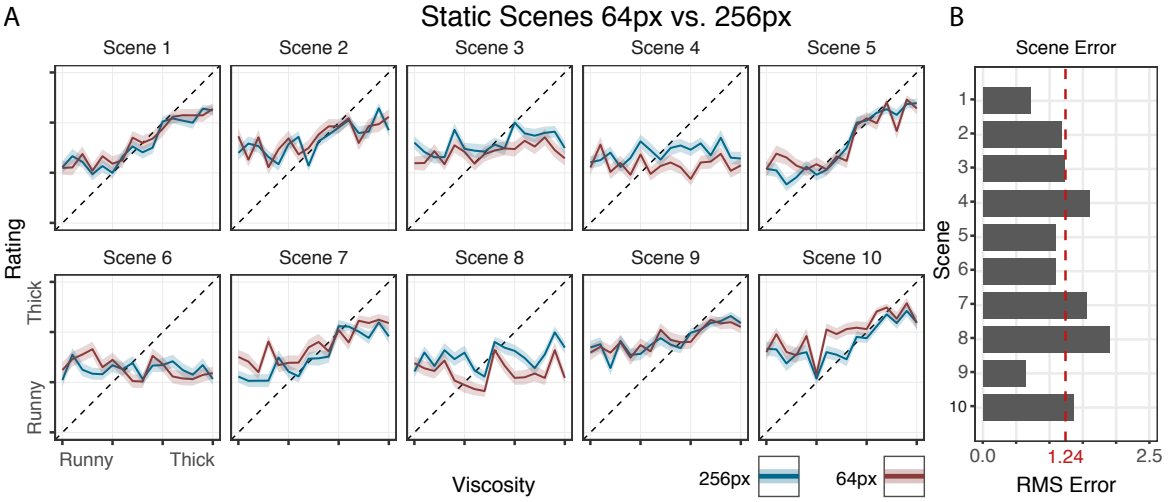


Figure D.1: (A) Viscosity ratings for the 10 different scenes with static stimuli of a 64px size (red) and 256px size (blue). The x-axis shows the tested viscosity steps (1-16). The y-axis shows the perceived viscosity. The error ribbons show the standard error of the mean (SEM). The dotted line shows the physical truth. (B) The x-axis shows the Root Mean Square Error for each of the 10 scenes on the y-axis. This is the error between the two conditions (64px - 256px). The dotted line shows the mean error across scenes.

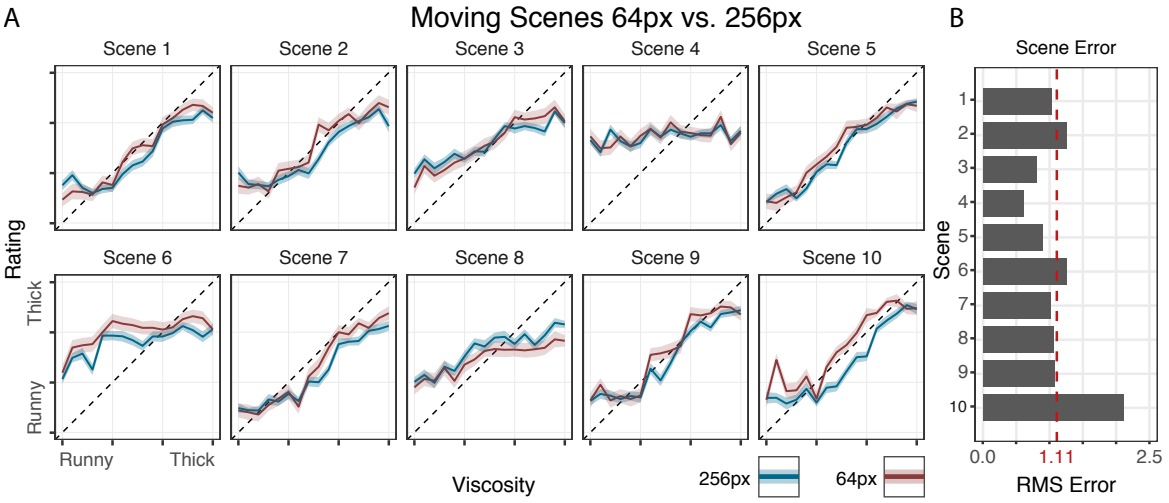


Figure D.2: (A) Viscosity ratings for the 10 different scenes with moving stimuli of a 64px size (red) and 256px size (blue). The x-axis shows the tested viscosity steps (1-16). The y-axis shows the perceived viscosity. The error ribbons show the standard error of the mean (SEM). The dotted line shows the physical truth. (B) The x-axis shows the Root Mean Square Error for each of the 10 scenes on the y-axis. This is the error between the two conditions (64px - 256px). The dotted line shows the mean error across scenes.

Statement

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ in carrying out the investigations described in the dissertation.

Gießen, Tuesday 30th January, 2018

_____

Jan Jaap van Assen