Vision as hypothesis testing: The case of biological motion perception

Nikolaus F. Troje

»The question of this meeting is, how we can explain that a simple line or form suddenly transforms into something meaningful, and how this transformation, this becoming alive, can be described, analysed, or even constructed.«

[Abstract]

Biological motion perception refers to the ability to derive a wealth of information from the mere movement of a few isolated light dots that move along with a person to whom they are attached. The appeal of biological motion stems from the discrepancy between the sparseness of the visual stimulus itself and the richness of the information that it appears to provide to our visual system. The enduring interest in this phenomenon is due to the fact that it allows us to study one of the cardinal questions of perception: How does our brain turn generally noisy, incomplete, ambiguous sensory data into a consistent, stable, and predictable model of the world? The answer to this question lies in the fact that perception is not only based on current sensory information alone, but also on expectations based on previously learned knowledge about the statics of the world.

[Zusammenfassung]

Wahrnehmung biologischer Bewegung bezieht sich auf die Fähigkeit, der Bewegung weniger isolierter Lichtpunkte Information über die Person zu entnehmen, welche diese Lichtpunkt bewegt. Die Faszination biologischer Bewegung ergibt sich aus der Diskrepanz zwischen der Spärlichkeit der Darstellung selbst und der Fülle von Information welche sie vermittelt. Das andauernde Interesse an diesem Phänomen hat damit zu tun, dass es uns erlaubt eine der zentralen Fragen über Wahrnehmung schlechthin zu studieren: Wie gelingt es unserem Gehirn verrauschte, unvollständige, mehrdeutige sensorische Daten in ein konsistentes, stabiles und vorhersagbares Weltmodell zu transformieren? Die Antwort zu dieser Frage liegt in der

Tatsache, dass Wahrnehmung nicht nur auf der momentanen sensorischen Information beruht, sondern ebenso auf zuvor gelerntem Wissen über die Statistik der uns umgebenden Welt.

Vision as an inverse problem

Visual perception — like other perceptual processes — is riddled with ambiguities. Our brain has no direct access to the objects and events that form our visual environment — sometimes referred to as the »distal stimulus«. Rather, it has to cope with the information provided by the »proximal stimulus« — the part of the continuous and ever changing stream of light reflected off the objects that surround us which reaches the light-sensitive back of the eye. The unprocessed retinal image is not too useful, to begin with. On the one hand, it is cluttered with irrelevant fluctuations and way too much information. Only a small fraction of the light distribution is directly determined by the qualities of objects and events that are relevant to the observer. On the other hand, it may not contain enough of the relevant information. Even if successfully segregated from the irrelevant parts, the remaining information may be incomplete and insufficient to unambiguously infer the important attributes of objects and events. In general, the mapping from the distal onto the proximal stimulus is a process that involves the loss of information. For instance, consider the projection of a 3D object onto the 2D surface of the retina. The resulting retinal image would be consistent with an infinite number of 3D objects. The projection itself does not provide information about which one of these solutions corresponds to the veridical distal stimulus [Fig.1].

The sampling of the retinal image by means of the pigments contained in the photoreceptors of the eyes, the transduction of the absorbed light into neural responses, the propagation of these signals along the axons of the visual neurons, and the synaptic transmission from one neuron to the next implement additional filters and therefore add ambiguity as well as more noise to the system. The signals that eventually reach our brain have to be considered a vastly degraded representation of the outside world.

This scenario motivates the cardinal question that defines perception itself as well as large parts of the research set out to understand the underlying mechanisms:

How does the brain transform a stream of generally noisy, incomplete, unstructured, ambiguous sensory data into a meaningful, reliable and predictable model of the world?

This question needs to be answered both by the nervous system that is trying to make sense of the initially unstructured flow or sensory information, but it is also the main question of the researcher who tries to understand how computational systems, such as our own brain, another animal's brain, or a computer can possibly approach such a challenging task.

The task that needs to be solved is sometimes formulated in terms of an »inverse problem« [1]. While the mapping from the distal stimulus onto the proximal stimulus is usually well defined and has a unique solution, the same is not the case for the inverse problem: the reconstruction of the distal stimulus from the information contained in the proximal stimulus. The inverse problem is under-constrained and therefore ill posed. Additional constraints are required to arrive at an unambiguous solution.

Over the history of the science of perception, a number of sources for such additional constraints have been proposed. Hermann von Helmholtz [2] formulated perception as an inferential process: Rather than attempting to reconstruct the distal stimulus on the basis of the proximal stimulus, the perception system generates hypotheses about its nature and then tests these hypotheses against the current sensory input. The school of Gestalt psychologists [3] formulated a number of ›laws‹ that reflect plausible regularities of the visual environment in terms of perceptual heuristics. Both approaches suggest the employment of assumptions or ›knowledge‹ about the statistics of the world. This ›knowledge‹ could have been acquired through individual learning in some cases. Alternatively, it could have been collected during the evolution of our sensory and perceptual systems to become established in the physiology of our nervous system. The important point is that it is independent of the current sensory input and can be used as an independent source to constrain inverse problems: If the current sensory input is compatible with more than one solution to the inverse problem, then solutions that are *a priori* unlikely are rejected and more likely solutions are preferred.

Many more recent scholars of the mechanisms underlying perception have elaborated on this general principle [4-6], to name but a few. Quantitative formulations of this principle emerged and have been used extensively in the past. Among the most influential approaches are the simplicity principle [7], regularization approaches e.g. [8], and Bayes'ian methods e.g. [9]. All three are mathematically related and to a large degree equivalent [10,11]

The Bayes'ian approach states that the optimal decision the perceptual system can come up with is based on the distribution of probabilities of distal stimuli, given the proximal stimulus. According to Bayes' rule, this so-called *a posteriori* probability is proportional to the product of the distribution of probabilities of the current proximal stimulus conditional on the choice of distal stimuli and the distribution of *a priori* probabilities of distal stimuli under consideration. Formally:

$$p(w|s) \sim p(s|w) \cdot p(w) \qquad \text{Eq. 1}$$

Here, the letter *w* stands for ›world‹ and therefore for the distal stimulus, and the letter *s* stands for ›sensation‹, and therefore for the proximal stimulus.

Note, that $p(s|w)$ is not a probability distribution in the conventional sense but a *likelihood*, that is, a function of the condition — in this case the manifold of distal stimuli that constitute possible ›worlds‹.

Once the posterior probability distribution has been computed, the question remains of how to derive from it a decision as to how to interpret the proximal stimulus. It seems that human vision often behaves as if it would simply try to maximize the posterior probability. Such an estimator is called a *maximum a posteriori* estimator (MAP). In cases in which the property of interest *w* consists of a continuous parameter the expected value of the posterior probability can be expressed as:

$$\hat{w} = \int w \cdot p(w|s)dw = \int w \cdot p(s|w) \cdot p(w)dw \qquad \text{Eq. 2}$$

An interesting variation of the expected value estimator consists of the case in which the costs of potential decision errors are taken into account. This will have only an effect on the final decision in cases in which some kinds of mistakes are more costly than others. Within decision theory, differential costs (or relative benefits) are then formulated in terms of a *utility function*, which assigns a relative value to each potential outcome w. The decision is then based on maximizing the expected utility:

$$\hat{w} = argmax\, E(u) = \int u(w) \cdot p(w|s)dw = \int u(w) \cdot p(s|w) \cdot p(w)dw \qquad \text{Eq. 3}$$

In this formulation, the optimal decision depends on three factors: the utility function, the likelihood, and the prior probability.

Is there a snow plough in my room?

The following story was told to me by a colleague who at that time had young children. One winter night the two-year old woke up terrified, crying that there was a snow plough in his bedroom. He had seen it roaring by with his own eyes. He had heard it, too. It took quite a while to calm him down.

What had happened was that a noisy truck had driven by the quiet neighbourhood street. Maybe it was in fact a snow plough. The truck's lights had cast the shadow of a chair that happened to be somewhere close to the window onto the wall of the child's bedroom. The shadow was moving along with the truck and perfectly fit the noise that entered the room through the window.

If there had been a real snow plough in the room, it might have cast a shadow on the wall similar to the one that in fact came from the chair. In other words: The likelihood of obtaining this kind of shadow, given that there was a snow plough in the room, is about the same as the probability of obtaining exactly this shadow, given that there was a chair in the room and a truck safely moving past the house.

If an adult observer would have been in the same situation and had to decide if the shadow on the wall was cast by a snow plough in the room or by a piece of furniture *a priori* probabilities would have clearly voted for one over the other solution. For a two-year old, who has not had as many opportunities to learn about the statistics of the world, it may not be that clear that occurrences of snow ploughs in bedrooms are generally extremely rare and therefore extremely unlikely — specifically when compared to the likelihood of occurrences of chairs in bedrooms.

Ambiguous vision

Pretty much any visual stimulus is ambiguous at some level. Take the one from Fig. 2a. It seems to show a cube. The image is certainly consistent with the assumption that it is really showing a cube and there might be good reasons to assume that it is one. However, if shown from a slightly different viewpoint as in Fig. 2b, it becomes obvious that the observer's assumption was wrong. Only presented from a very specific, accidental viewpoint — like the one used in Fig 2a — it generates the same image as a cube would.

Why do we think this is a cube even though it is not? We again employ the Bayes'ian approach from Eq. 3. The prior probability of a regular cube is probably not much higher than the prior probability of an irregular solid. In this case, however, the likelihoods are very different. If this was a completely irregular solid, chances that the observer happened to capture a view in which all edges line up to be parallel in the 2D projection (even though they are not parallel in 3D) are very low. The likelihood, that is, the probability of obtaining this image given that it is an irregular object is very low. The probability that all these lines end up to be parallel in the 2D image, given that this is a cube with parallel lines in 3D is far higher.

An interesting aspect about the particular object depicted in Fig. 2a is, that even if we accept that it is a cube, there are still two possible distal stimuli which could have generated the image we see. Both are adopted by our visual system, and as we look at the cube for an extended period, the two alternate. The observer can either see a cube, in which the lower square is facing the observer, and the upper square is facing away, or we can see a cube in which the upper square is facing us and the lower square is facing away. The two interpretations differ mainly in terms of the viewpoint. In the first case we are looking down at the cube from an elevated position; in the second case we are looking up at the cube from a viewpoint below it. The two interpretations differ in terms of a mirror flip about the projection plane. If a certain corner of the cube is interpreted to be positioned at a certain distance before the projection plane in one interpretation, it will be located at exactly the same distance behind the projection plane in the other interpretation.

The bistability caused by the depth-ambiguity of this rendering was first described in 1832 by a Swiss crystallographer named Luis Albert Necker [12] and has since adopted his name. Note that if we rotate the Necker cube from one position into the next, the direction of rotation depends on whether we see the upper or the lower square to be facing us. In other words, once an observer indicates whether a rotating Necker cube is perceived to be spinning or counter-clockwise we know which of the two depth interpretations he is currently experiencing.

Conducting such experiments demonstrates that the two possible Necker cube percepts do not occur at exactly the same frequency. It turns out that the interpretation that involves looking down at the cube from above occurs more frequently than the interpretation in which we look at the cube from below. How strong the dominance of the viewing-from-above bias is depends on how it is measured. If presented over longer periods the cube is seen in 60% of the cases as viewed from above [13]. For short presentations, this preference can become very pronounced, approaching almost a 100% viewing-from-above bias [14].

Other biases have been described in the literature that demonstrate how we use prior knowledge and expectations about the regularities of the word to shift the percept of an ambiguous stimulus in one or the other direction. Fig. 3 shows an image which some observers would interpret as a dent and others as a bump [15]. If we see it as bump, we also see the light that causes the shading coming from the lower right side. If you see it as a dent, we see the light coming from above. If we turn the page on which it is printed upside down, we can still see it as a dent or as a bump, but now the bump is associated with light coming from above and the dent is associated with light coming from below.

If we systematically measure the frequency with which observers experience the different interpretations, we find that while the percept is certainly ambiguous and bistable, observers have systematic preferences (or more precisely: The observers' visual systems have systematic preferences). On the one hand, we seem to give precedence to the assumption that the light is coming from above rather than from below. On the other hand, we seem to prefer convex interpretations (bump) over concave interpretations (dent). If we turn the page upside down, one of the two theoretically possible interpretations is a bump lit from above. Our bias for seeing this interpretation is so strong that it is hardly ever flipping into the other interpretation, a dent lit from below. If we keep the page right side up, the two biases work against each other. If the light-from-above bias dominates, we see it that way, but the shape becomes concave. If the convexity bias dominates, we see a bump, but the light now comes from the non-preferred direction.

Where do these biases come from? Light usually does come from above. The sun, and the sky which scatters the sunlight, are above us and they are the main sources of light in our world. Sometimes light may come from below, but only in rare and generally artificial cases. In Bayes'ian terms, we can say that the prior probability of light coming from above is much higher than the prior probability that light comes from below. The likelihood for the two interpretations is the same, however. Provided that the light is coming from below and the shape of the object is concave, the probability of ending up with the image of Fig. 3 is the same as the probability of getting this image, given that light comes from above and the shape is convex.

The convexity bias also seems to correlate with the statistics of the world. Objects tend to be convex. The most prototypical shape is a sphere, which is entirely convex. Deforming a sphere to craft other shapes may involve the introduction of concavities, but each concavity comes along with even more pronounced convexities. Unless we are looking at an object from inside the object, we are looking more often at convex surfaces than at concave ones. Again, speaking in Bayes'ian terms, the *a priori* probability for convex surfaces is larger than the one for concave surfaces.

Maybe even more important is the fact that any feature on a convex part of a body is more likely to be seen than a feature on a concave surface. Among all possible viewpoints from which objects can be observed, many result in self-occlusion of a given feature — the more concave the location of the feature, the more often that will happen. The likelihood $p(s|w)$ (Eq. 1) that assumes convexity is larger than the likelihood that assumes concavity.

Bistable figures like the Necker cube or the shaded surfaces of Fig. 3 receive on-going attention from vision researchers. The reason is that they take us back to the central question of perceptual science mentioned above: Why does the world around us feel so stable and reliable, given what we know about the limited reliability of sensory information?

The world probably did not feel as reliable at the time we were born and it took our visual system substantial time and experience to make up its mind about all the potential ambiguities that characterize proximal stimulation. We had to learn our priors. My colleague's two-year old boy hadn't completed this process yet. But once he is grown, he will be done with it for the most part. Bistable figures are so interesting because they provide rare cases on which our brain still has to work. For some bistable stimuli we (well, that is, our visual system) will never find final solutions. But for others that might happen. If we are able to tame them such that we can study them under controlled laboratory conditions, we might be able to study perceptual disambiguation as it unfolds in front of our eyes.

Perception of biological motion

The ability to retrieve a wealth of information from just a few moving dots is a wonderful demonstration of the effectiveness of perceptual organization. The different phenomena involved with biological motion provide a rich field to study the mechanisms underlying perceptual processes.

The term *biological motion* was first introduced by the Swedish psychologist Gunnar Johansson almost 40 years ago [16]. Johansson created visual stimuli which demonstrate the incredible ability of the human visual system to perceptually organize a few moving light dots into the compelling percept of a person in action. His early displays were created by asking a person to wear dark clothing attached with small light bulbs. Later he used reflective material. In both cases, he then filmed them in a dark room while they were performing simple actions, such as walking or dancing. The result was a set of movies in which the actor's body itself was entirely invisible. All that could be seen were the light dots (or reflective patches) and their motion as bright dots against a dark background.

When looking at a still frame of one of these movies, the individual dots appear meaningless and their arrangement seems to be arbitrary. However, as soon as the pictures are set into motion, a vivid percept of the person performing the action is elicited in the observer. The motion seems to connect the individual dots into a coherent shape, replacing the missing sticks that connect the dots. This articulated shape can then be tracked as it deforms to generate action. It has been shown that the reconstruction of the underlying structure of the actor's body and the kinematics of its motion is sufficient to recognize familiar individuals, to understand sophisticated actions, and infer emotions, intentions, and even attempted deception for reviews, see [17,18]. For a demonstration of the phenomenon, see [19].

In Fig. 4a we show a still frame in which we connected the dots with lines in order to simulate the structure that would be missing in a still point-light depiction, but would be present in an animated version of it.

The figure shows an array of 15 white dots against a black background. However, what the observer sees (and what he would be seeing if the lines were not drawn, but the dots were moving along with the person's joints) is much more than just 15 moving dots. One cannot help but seeing a person that walks and that is probably heading towards you.

That is in spite of the fact that, in analogy to our discussion of the Necker cube, the image of Fig. 4a does not have to be the depiction of a human figure. The image (or the animation) itself is perfectly compatible with other shapes, too. Both the stick figures shown in Fig. 4b and 4c would result in the same projection — the one of Fig. 4a — if rotated 10 degrees clockwise.

For the same reasons mentioned when we discussed the Necker cube, it makes perfect sense to assume that we are looking at a human figure here, even though we can never be absolutely sure: It is highly unlikely that a non-human shape projected onto a 2D surface would accidentally create an image identical to the projection of a true human shape.

Another analogy with the Necker cube also holds true. Even if we assume that Fig. 4a is a non-accidental view of a human walker, there still remains an ambiguity. The display is perfectly compatible with two kinds of interpretations which both involve the shape of a normal human walker. It could show a walker that is moving towards the observer, but it could also show a walker that is facing away from the observer — presenting us his back. As for the Necker cube, the two interpretations differ in a flip about the projection plane.

If the stick-figure had been rendered with a camera that had been pointing down on it from a slightly elevated angle, the mirror flip about the projection plane would also induce a change in perceived camera elevation. The camera would now look from below at the figure. Also, if the stick-figure were spinning about a vertical axis (or, for that matter, about any axis within the image plane) the perceived spinning direction would reverse along with a flip in perceived depth. It should be noted that the handedness of the figure would flip as well. What used to be the figures right hand, now becomes the left hand, and vice versa. As long as the figure is bilaterally symmetric, the flip in handedness goes unnoticed as was the case for the Necker cube, too, but if the figure adopts an asymmetrical postures, this might play a role [14].

The facing-the-viewer bias

Biological motion point-light walkers and stick figures feature another interesting bias which cannot readily be reduced to any of the other biases that we have encountered so far – at least not on first sight. Most observers will strongly prefer an interpretation in which the figure is facing them over an interpretation in which they see the walker going away. The phenomenon has first been described by Vanrie and Verfaille [20]. A number of studies followed up on this first observation in an attempt to formulate an explanation for it and a few interesting ideas have been pursued.

The first one was first implied by Vanrie and Verfaille [20] who discussed possible differences in biological and social relevance between the two percepts. There are two kinds of errors an observer can make when deciding if a person is walking towards or away from the observer. One can falsely assume that someone who is in reality approaching is walking away, and one can falsely assume that someone who is in reality walking away is approaching. Depending on the situation the costs of making the first error might be significantly larger than the costs of making the second error. The consequences of mistaking someone who is leaving for someone who is approaching may be negligible, but not being prepared for a looming encounter has the potential of serious trouble. Being faced with an ambiguous situation, we might rather decide to act as if the other person is facing us.

So far, no study has explicitly tested this hypothesis and it may prove hard to do that in a well-controlled way. However, a study by Brooks et al. [21] shows that the facing bias is much more pronounced when a point-light walker displays clearly male traits, and weaker, absent, or even reversed when the walker appears to be female. This finding seems to be consistent with the above theory: If the main cost of a potential error is that the observer is not prepared to respond to a potential threat, then it seems plausible that the facing bias is particularly pronounced when dealing with a big, strong male character, as compared to a slim, potentially less threatening female person. In terms of Bayes'ian decision theory, this argument rests on the utility function which introduces the asymmetry that results in the facing-the-viewer bias.

While this idea is certainly appealing, the data provided by Brooks et al. [21] leave room for other explanations, too. Schouten et al. [22], in an attempt to replicate the above data with a larger group of participants, found a dissociation between the apparent gender of the walker and the facing bias it elicited. The authors manipulated point-light displays to appear either male or female in two different ways. Half of their participants were presented with walkers in which only the geometry of their bodies, but not their movements, clearly indicated their sex. The other half of the participants were presented with walkers in which the motion itself was indicative of the walker's sex, but the geometry of the bodies was kept neutral. Both manipulations resulted in the same degree of perceived gender, but only the manipulation of body geometry resulted in the expected change in the degree of facing bias. Manipulation of the kinematics of the motion had a small effect which pointed in the opposite direction: Female kinematics generated a larger facing bias than male kinematics.

In the same study, the authors presented observers with either only the upper half or only the lower half of the point-light figure. They found that both for male and female walkers, just seeing the lower body induced a very pronounced facing bias, whereas just showing the upper body induced a strong bias toward seeing the figure receding. The effect of body half was much stronger than the one resulting from a manipulation of the gender of the walker. Also, the gender dependency was more pronounced for the upper body than for the lower body.

Schouten et al.'s [22] first experiment shows that it is not gender *per se* that determines whether we interpret an otherwise ambiguous point-light walker as approaching or receding. Rather it is related to some attributes of the walker which are inherent to the structural differences between male and female bodies. What are these differences?

Schouten's et al.'s [22] second experiment offers an interesting hypothesis which relates to the above mentioned convexity bias — preference of an observer to perceive a surface as convex rather than concave. When looking only at the lower body, the strongest convexity/concavity in the direction the line of sight is provided by the knee joints. If these joints are seen as a convexity, the walker is approaching, and if seen as a concavity, the walker is receding. The opposite may occur with the upper half of the body — at least for the caricatured versions of female walkers that were used in the experiments. Women tend to hold their elbows closer to their bodies than men which causes them to point backwards. Seen from behind the elbow joint is convex and seen from the front it is concave. That is somewhat different in prototypical men and certainly in the exaggerated depiction used in the experiments. Men have a tendency to point the elbows laterally outwards. The bending of the arm therefore lies in the fronto-parallel plane. The elbows are still convex, if seen from the side, but there is little convexity/concavity with respect to the line of sight if we look at the figure from the front or from behind.

If perception of local surface attitude in depth-ambiguous situations is influenced by a preference for convex curvatures, we would expect exactly the results which Schouten et al. report: The lower body is perceived to be facing the viewer and the upper body has a tendency to face away, particularly for female walkers. We would also expect an overall effect of gender for full body displays driven by the differences of the upper body.

In terms of Bayes'ian decision theory — as explained above — the convexity bias can be explained both on the level of the prior as well as on the level of the likelihood term of Eq. 3. For reasons which are beyond the scope of this chapter, we assume that it is in fact the likelihood of image formation rather than the larger prior probability that determines human visual perception of ambiguous point-light figures.

How about the prior probability of approaching versus receding people? Maybe, people in our immediate visual environment are more likely to face us than to face away? That might in fact be the case, specifically in diadic situations or and in small groups. No data exist that demonstrate that the statistics of relative orientation of people in small groups truly form the prior used by the decision system that causes the facing bias, but a preliminary study from our own lab provides some interesting insight [23].

This study demonstrated that an initially moderate facing-the-viewer bias experienced when first encountering depth-ambiguous point-light figures grows much stronger with increasing exposure to point-light displays. Initially, we were surprised to find that undergraduate students who participated in our experiments never showed the same extent of facing bias as the members of our laboratory experienced themselves [24] [Fig. 5a]. In an attempt to rule out other causes for this pronounced difference, we conducted an experiment in which naïve undergraduate students were tested with respect to their facing bias first at the beginning of the school year, and then again 10 weeks later towards the end of the term. Between these two main sessions, we asked them to work on a number of biological motion related tasks for about an hour every week. All tasks were entirely unrelated to the one used to measure their facing bias. Among other tasks, they had to detect biological motion in noise, they had to indicate the apparent gender of a point-light figure or they had to determine which kind of action they performed. The purpose of the weekly sessions was to provide them with ample exposure to the displays.
The difference in the amount of facing bias between the two main sessions was not quite as large as the one between naïve undergraduate students and long term graduate students of the BioMotion lab, but still impressive [Fig. 5b]. At the end of the experiment, the participants demonstrated a much stronger facing bias that in the initial test. What is the origin of this experience dependent increase in the facing-the-viewer bias?

What seems to be happening here is that observers are modifying the prior distributions which contribute to their decisions based on the experience they had during the 12 weeks of this experiment. That, in itself, is nothing surprising. It makes sense to re-learn priors based on individual experience. It has been shown in other contexts that this happens even within the course of much shorter experiments, e.g. [25,26], and the underlying mechanism is considered an integral ingredient of Bayes'ian learning. What is surprising is that this happens even in a case in which the experience is not based on any kind of ›ground truth‹. All of the point-light stimuli that they had been confronted with over the course of the experiment were ambiguous with respect to their depth. The only ›evidence‹ for a higher prevalence of point-light figures facing the viewer as compared to point-light figures facing away was based on the observers' experience which in itself is determined by the initially moderate, but nevertheless dominating, facing bias.

The experience of seeing more approaching than receding walker seems to reinforce itself, shaping the initially biased prior distribution into more extreme directions. Such a drift in the prior probability distribution is only possible if the true experience becomes inflated at least by a small degree. For instance, experiencing 60% of the walkers facing the viewer would then result in the use of a prior that peaks at a somewhat higher value, say at 62%. This in turn enters the decision making process and therefore the subjective experience, which will now be at 62%. Under conditions in which the experience is determined directly by the statistics of a visual environment that can be assumed to be independent of the observer, such a small inflation would not bear a significant consequence as it will be kept at limit by the increasing discrepancy between experience and the current prior. But in the absence of an independently existing ›ground truth‹ the mechanism causes self-reinforcement which only ends when the initially bistable percept stabilizes on one of the possible percepts.

Conclusion

Perceptual decision making is a complex process, despite that fact that our mature perceptual experience seems to be rather unambiguous. Subjectively, the inferential nature of perceptual processes is not apparent. Nevertheless, the noisy and generally ambiguous nature of natural visual stimuli requires such processes, and controlled experiments in which the reliability of certain stimulus properties is manipulated demonstrate that our perceptual system follows strategies which can be well described in terms of Bayes'ian decision theory.

Biological motion perception, that is, the ability of the human visual system to derive information about a moving person, his or her actions, personality, emotions, and intentions from just a few moving dots, provides a wonderful playground to study the processes underlying inferential perceptual decision making. Particularly interesting are stimuli for which even the mature, adult brain has not yet settled in on a final decision, but can still be observed in going through that process of coming up with one.

Here, we used the inherent depth-ambiguity of point-light displays to show that all essential ingredients of Bayes'ian decision making can play a role: The differential utility (or its inverse, the cost) of the various possible decisions; the likelihood, that is, the probability of arriving at the stimulus that we are seeing as a function of the nature of the distal stimulus; and the prior probability which describes the observation independent probabilities of the distal stimuli under consideration.

Captions:

Fig. 1: The two 3D shapes on the left side produce the same 2D image on the retina of the eye depicted in the centre – namely a straight line as depicted on the right. Without additional information, the brain connected to the eye has no way to decide exactly which shape it is looking at.

Fig. 2: The object on the left (a) looks like the projection of a cube. The image is in fact perfectly consistent with that assumption. However, other objects could as well result in the same image. For instance, the irregular object shown on the right (b), if rotated 10 degrees about a vertical axis, would also project the same image.

Fig. 3: The above picture can be perceived either as a convex structure (bump) illuminated from the lower right, or as a concave structure (dent) illuminated from the upper left. If turned upside down the two, theoretically possible percepts are a bump lit from above or a dent lit from below.

Fig. 4: The figure on the left (a) seems to show the projection of a human stick figure. In fact, if we were turning the one depicted in the centre (b) 10 deg clockwise, we would arrive at the view we see on the left. However the same is true for the figure on the right (c). If turned 10 degrees, we would also get the drawing on the left.

Fig. 5: The diagram on the left side (a) shows the difference in the degree of facing-the-viewer bias between observers which had little or no experience with point-light displays (shaded bar, first year undergraduate students), and students who were working with them on a daily basis (dark bar, members of the BioMotion lab) [24]. The diagram on the right (b) illustrates the difference in the degree of facing-the-viewer bias before and after familiarizing participants of the study systematically with biological motion displays [23]. The units on the ordinate are relative units. Error bars represent standard errors of the mean.

Endnotes:

1 Pizlo, Z. Perception viewed as an inverse problem. *Vision Res* **41**, 3145-3161 (2001).

2 Helmholtz, H. L. in *Handbuch der physiologischen Optik* (eds A. Gullstrand, J. von Kries, & W. Nagel) (L. Voss, 1867/1910).

3 Koffka, K. *Principles of Gestalt psychology*. (Harcourt, Brace & World, 1995).

4 Rock, I. *The logic of perception*. (MIT Press, 1983).

5 Barlow, H. Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research* **30**, 1561-1571 (1990).

6 Gregory, R. L. *Eye and brain: The psychology of seeing*. (Princeton University Press, 1997).

7 Pomerantz, J. R. & Kubovy, M. Theoretical approaches to perceptual organization: Simplicity and Likelihood Principles. *Organization* **36**, 3 (1986).

8 Tikhonov, A. & Arsenin, V. Y. Methods for solving ill-posed problems. *Scripta Series in Mathematics (Scripta Mathematics, New York, 1977)* (1979).

9 Kersten, D. High-Level Vision as Statistical Inference. *The new cognitive neurosciences*, 353 (1999).

10 Mumford, D. in *Perception as Bayesian inference* (eds D. C. Knill & W. Richards) 25-62 (Cambridge University Press, 1999).

11 Poggio, T., Torre, V. & Koch, C. Computational vision and regularization theory. *Nature* **317**, 314-319 (1985).

12 Necker, L. A. Observations on some remarkable phenomena seen in Switzerland; and an optical phenomenon which occurs on viewing of a crystal or geometrical solid. *Philosophical Magazine* **1**, 329-343 (1832).

13 Sundareswara, R. & Schrater, P. R. Perceptual multistability predicted by search model for Bayesian decisions. *Journal of Vision* **8**, 12 11-19 (2008).

14 Troje, N. F. & McAdam, M. The viewing-from-above bias and the silhouette illusion. *i-Perception* **1**, 143-148 (2010).

15 Ramachandran, V. Perception of shape from shading. *Nature* **331**, 163 - 166 (1988).

16 Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics* **14**, 201-211 (1973).

17 Blake, R. & Shiffrar, M. Perception of human motion. *Annual Review of Psychology* **58**, 47-73 (2007).

18 Troje, N. F. in *The Senses: A Comprehensive Reference* (ed A. Basbaum) 231-238 (Elsevier, 2008).

19 Troje, N. F. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *Journal of Vision* **2**, 371-387 (2002). For an animated demonstration of biological motion, see: http://www.biomotionlab.ca/Demos/BMLwalker.html.

20 Vanrie, J., Dekeyser, M. & Verfaillie, K. Bistability and biasing effects in the perception

of ambiguous point-light walkers. *Perception* **33**, 547-560 (2004).

21 Brooks, A. *et al.* Correlated changes in perceptions of the gender and the orientation of ambiguous biologicalmotion figures. *Current Biology* **18**, R728-729 (2008).

22 Schouten, B., Troje, N. F. & Verfaille, K. The facing bias in biological motion perception: Structure, kinematics, and body parts. *Attention, Perception & Psychophysics.* **73**, 130-143 (2011).

23 Troje, N. F. & Davis, M. Bootstrapping a prior? Effects of experience on the facing bias in biological motion perception. *Journal of Vision* **11**, 692 (2011).

24 Troje, N. F. Perceptual biases in biological motion perception and other depth-ambiguous stimuli. *Journal of Vision* **10**, 792 (2010).

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

26 Flanagan, J. R., Bittner, J. P. & Johansson, R. S. Experience can change distinct size-weight priors engaged in lifting objects and judging their weights. *Current Biology* **18**, 1742-1747 (2008).

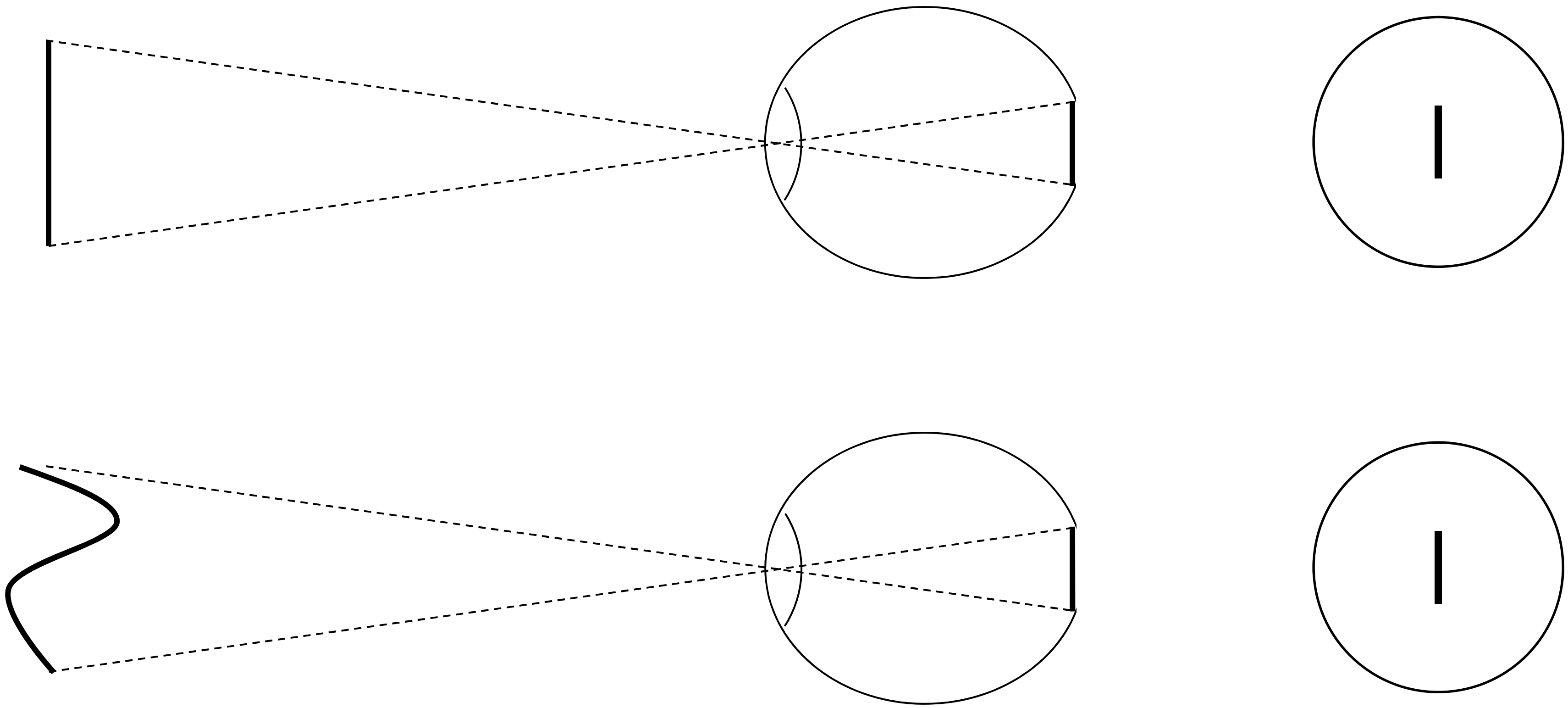

Fig. 1: The two 3D shapes on the left side produce the same 2D image on the retina of the eye depicted in the centre – namely a straight line as depicted on the right. Without additional information, the brain connected to the eye has no way to decide exactly which shape it is looking at.

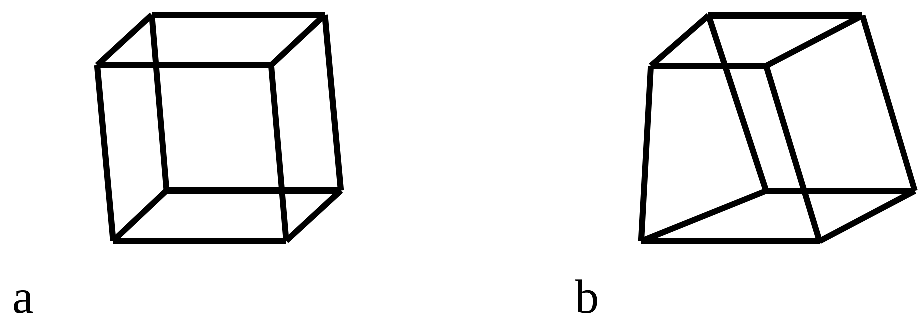

Fig. 2: The object on the left (a) looks like the projection of a cube. The image is in fact perfectly consistent with that assumption. However, other objects could as well result in the same image. For instance, the irregular object shown on the right (b), if rotated 10 degrees about a vertical axis, would also project the same image.

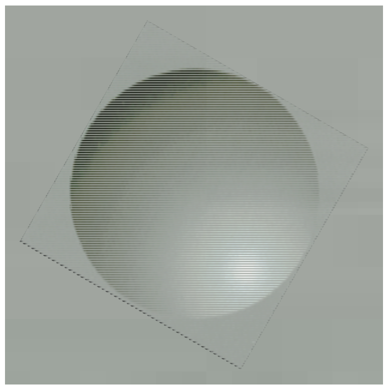

Fig. 3: The above picture can be perceived either as a convex structure (bump) illuminated from the lower right, or as a concave structure (dent) illuminated from the upper left. If turned upside down the two, theoretically possible percepts are a bump lit from above or a dent lit from below.

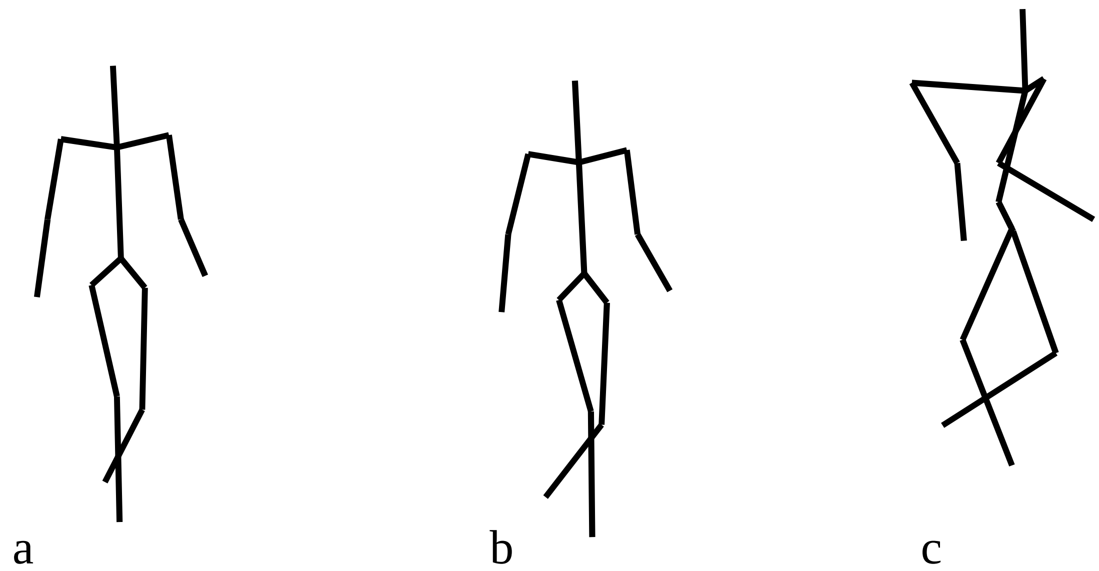

Fig. 4: The figure on the left (a) looks like the projection of a human stick figure. In fact, if we were turning the one depicted in the center (b) 10 deg clockwise, we would arrive at the view we see on the left. However the same is true for the figure on the right (c). If turned 10 degrees clockwise, we would also get the drawing on the left.

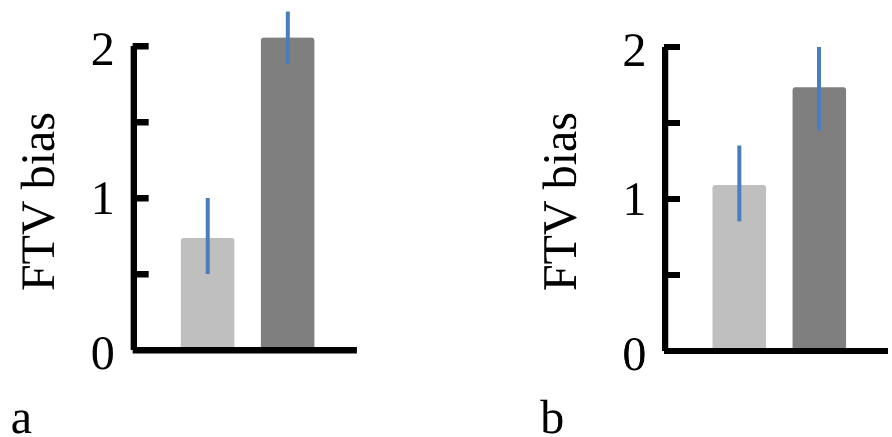

Fig. 5: The diagram on the left side (a) shows the difference in the degree of facing-the-viewer bias between observers which had little or no experience with point-light displays (shaded bar, first year undergraduate students), and students who were working with them on a daily basis (dark bar, members of the BioMotion lab). The diagram on the right (b) illustrates the difference in the degree of facing-the-viewer bias before and after familiarizing participants of the study systematically with biological motion displays. The units on the ordinate are relative units. Error bars represent standard errors of the mean.