

Experimental Manipulation of Action Perception based on Modeling Computations in Visual Cortex

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät

und

der Medizinischen Fakultät

der Eberhard-Karls-Universität Tübingen

vorgelegt

von

Leonid Fedorov

aus St.-Petersburg, Russia

January - 2018

Tag der mündlichen Prüfung: 13 June 2018.....

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel

Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Prof. Dr. / PD Dr. / Dr. Martin Giese.....

2. Berichterstatter: Prof. Dr. / PD Dr. / Dr. Andres Bartels.....

Prüfungskommission: Dr. Prof. Martin Giese.....

Dr. Prof. Andreas Bartels.....

Dr. Prof. Peter Thier.....

Dr. Prof. Hanspeter Mallot.....

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

„.....“

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled
„.....“

, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den

.....

Datum / Date

Unterschrift /Signature

Acknowledgements

I express immense gratitude to my supervisor Martin Giese who gave me an introduction and guidance through the mysterious jungle of vision science. In an interdisciplinary field like this both focus on method and breadth of view are crucial. I was lucky to be able to learn both extremely specialized methods and tools hands-on, and at the same time have access to manifold other approaches and aspects of study at numerous attended scientific meetings. No matter what challenges were upfront, and of what size, when working with Martin I have consistently felt that the door to academic research is open for me.

I learnt (and hope to continue doing so) the way of statistical analysis, the way of scientific writing, the way of in-lab communication and much more from two of my senior coauthors, friends and mentors: Tjeerd Dijkstra and Stephan de la Rosa.

Special ‘Thank you!’ goes to Albert Mukovskiy for a plenty of mentorship, to-the-point technical discussions and frequent advice.

I thank my advisory board members, Andreas Bartels and Peter Thier for timely corrections of the course of my PhD.

When preparing this work I have learnt much from, even when brief – always memorable and useful, encounters with the following scientists at conferences and other meetings, and from discussions of my and their own work (in no particular order): Rufin Vogels, Dzmitry Kaliukhovich, Joris Vangeneugden, Nick Barraclough, Stefano Panzeri, Hongjing Lu, Pierre Kornprobst, Nicolas Rougier, Hugh Wilson, Howard Hock, Theofanis Panagiotaropoulos, Steven Thurman, Gordon Pipa, Vishal Kapoor, Abhilash Dwarakanath and many others.

My colleagues Dominik Endres, Enrico Chiovetto, Nick Taubert and Dong-Seon Chang never hesitated to share insight from their skills, and to provide perspective. For laborious assistance on all institute and graduate school matters I thank Mirjana Angelovska and Tina Lampe.

I would not have been able to carry on all work without constant support from my family: dear Daria, Elena, Andrey and Dmitry.

Summary

Action perception, planning and execution is a broad area of study, crucial for future development of clinical therapies treating social cognitive disorders, as well as for building human-computer interaction systems and for giving foundation to an emerging field of developmental robotics. We took interest in basic mechanisms of action perception, and as a model area chose dynamic perception of body motion.

The focus of this thesis has been on understanding how perception of actions can be manipulated, how to distill this understanding experimentally, and how to summarize via numerical simulation the neural mechanisms helping explain observed dynamic phenomena.

Experimentally we have, first, shown how a careful manipulation of a static object depth cue can in principle modulate perception of actions. We chose the luminance gradient as a model cue, and linked action perception to a perceptual prior previously studied in object recognition – the lighting from above-prior. Second, we have explored the dynamic relationship between representations of actions that are naturally observed in spatiotemporal proximity. We have shown an adaptation aftereffect that may speak of brain mechanisms encoding social interactions.

To qualitatively capture neural mechanisms behind ours and previous findings, we have additionally appealed to the perceptual bistability phenomenon. Bistable perception refers to the ability to spontaneously switch between two perceptual alternatives arising from an observation of a single stimulus. Addition of depth cues to biological motion stimulus resolves depth-ambiguity. To account for neural dynamics as well as for modulation of action percept by light source position, we used a combined architecture with a convolutional neural network computing shading and form features in biological motion stimuli, and a 2-dimensional neural field coding for walking direction and body configuration in the gait cycle. This single unified model matches experimentally observed switching statistics, dependence of recognized walking direction on the light source position, and makes a prediction for the adaptation aftereffect in perception of biological motion.

Contents

Introduction

A.	What is action perception?	1
B.	The focus of this thesis.	1
C.	Brain networks for perceiving actions.	4
D.	Bistability phenomenon.	5
E.	A note on auditory bistability.	7
F.	Three studies modelling aspects of bistability.	8
G.	How to measure bistability.	14
H.	Perceptual adaptation.	17
I.	Neural adaptation.	23
J.	The use of adaptation in study of action perception. ...	25

Declaration of contribution.

Chapter 1. Adaptation aftereffects suggest neural representations for the encoding of social interactions.

Chapter 2. Lighting-from-above prior in biological motion perception.

Chapter 3. Neural model for the influence of shading on the multistability of the perception of body motion.

Introduction

A. What is action perception?

Action perception can be broadly understood as the study of neural and behavioral processes at a time scale of 0.6 - 20 seconds in response to "action" stimuli. Examples of action stimuli are displays of human or animal locomotion, displays of grasping an object or using a tool, or displays of multiple interacting agents. In most studies, the stimuli have been reduced from those observed in nature, so that they are composed of only a limited set of cues (features). These enabled a targeted study of the effects of these cues on perception and on tuning of neurons found to be responding to such stimuli. One frequently studied process, together with its stimulus class, is the recognition of biological motion. Initially, motion and form cues have been addressed the most (Bonda, Petrides, Ostry, & Evans, 1996; Oram & Perrett, 1996; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001), culminating in a phenomenological model (M. A. Giese & Poggio, 2003) incorporating parallel processing via dorsal and ventral pathways of the visual cortex. The outputs of pathways are integrated to model sequence-selective responses of neurons in area STS to full motion patterns. This neurally - plausible architecture has been successfully adopted to aid tasks in computer vision models and systems (Escobar, 2008; Jhuang, 2007; Layher, 2014).

B. The focus of this thesis

Here we focused on modeling dynamic neural processes underlying the perception of locomotion, where a surface shading cue is added to the stimulus of a depth-ambiguous walking figure. Surface shading is a classically studied static object perception cue, that has been adopted also to study perception priors – particularly, the lighting-from-above prior (Adams, Graf, & Ernst, 2004; Brewster, 1844; Kleffner & Ramachandran, 1992; Ramachandran, 1988; Sun & Perona, 1997). We take a point of view that all stages of signal processing should be taken into account, so that an interest in studying, in this example, a lighting-from-above prior, should be start with

understanding computations in the visual cortex that stimuli with luminance gradients induces. In terms of dynamic neural processing of luminance gradients inside the figure contour, several neurally plausible models have been proposed (Grossberg, Kuhlmann, & Mingolla, 2007; Mingolla & Todd, 1986). Brain networks for texture processing and extraction of fine-grained object shape have been partially probed with human imaging studies (Fleming, Holtmann-Rice, & Bulthoff, 2011; Gerardin, Kourtzi, & Mamassian, 2010; Tsutsui, Jiang, Yara, Sakata, & Taira, 2001; Tsutsui, Sakata, Naganuma, & Taira, 2002; Yamane, Carlson, Bowman, Wang, & Connor, 2008). We combined the luminance gradient (a static object cue) with the biological motion display, and discovered the influence of the light source position on the walking direction percept. To explain the surprising combination of these two distinct areas of study (shading cues in object perception and biological motion in action perception), we considered the dynamic stability of action percepts and the shading cues as one way to influence the perceptual state. We unified both in a single model using neurally plausible mechanisms to explain the processing of the shading cue in biological motion (Chapter 2) and the bistability in action perception (Chapter 3). The bistability investigation also triggered our interest in the interplay between perceptual and neural adaptation aftereffects and led to an experimental study delving into an area of social cognition (Chapter 1). The modeling is based on recent electrophysiological studies of coding for such stimuli in macaque visual temporal cortical neurons (Vangeneugden et al., 2011; Vangeneugden, Vancleef, Jaeggli, VanGool, & Vogels, 2009), on psychophysical experiments with such stimuli in humans in other literature, and on human psychophysical experiments performed in this study. Effects represented in the model include multistability, influence of firing rate adaptation on behavioral measurements, and a novel perceptual illusion, also tested experimentally. The model incorporates computations on the same exact stimuli that were used experimentally, and proposes an explanation how stimulus changes drive the observed bistable dynamics at the relevant time scale.

Furthermore, we followed up on the idea that neural-population level adaptation aftereffects in action coding circuits may be observable behaviorally (N. Barraclough

& Jellema, 2011; Ferstl, Bulthoff, & de la Rosa, 2017; Troje, Sadr, Geyer, & Nakayama, 2006). Current approaches deal with coding of single action sequences and their implementation in the brain. We set out to probe the interplay between the coding of multiple action sequences. Particularly, the exact question we formulated showed overlap with social cognition: whether we could speak of dedicated neural processing mechanisms behind social interactions. We describe (psychophysical) evidence for such mechanisms. Namely, the behavioral adaptation aftereffect induced by visually dissimilar, but spatiotemporally proximal actions.

The introduction to this thesis starts with acknowledging the spatially local nature of action processing in the brain by referencing studies on brain networks in action perception, and concluding that we mostly base our understanding on a small part of these networks. We then briefly pay due to the area of bistable perception in general, and further restrict our attention to modelling bistability as a way to qualitatively summarize many empirical accounts. We discuss three modelling studies eliciting aspects of bistable perception in general (and that are applicable to action perception): one using two level of neurons with bistability at both of them, one using psychophysical measurements to restrict the model parameters, and one using a tool similar to ours – a neural field. We highlight methodological challenges in the models that we found useful for our own work. Bistability models feature an adaptive component and can be simulated for exhibiting an adaptation aftereffect as well. Hence, we discuss the empirical basis for adaptation aftereffects in action perception. We review some imaging and behavioral studies that facilitate an adaptation aftereffect to justify why further basic studies are important, as well as why the adaptation aftereffect could be useful.

Working with extremely diverse aspects in perception of action, we have only begun to probe mechanisms in action perception (and not even to account for further stages, such as action planning). Yet, at this point of the journey at which this dissertation is submitted, we propose that both instantaneous processing of external signals and slower dynamic mechanisms, both constrained by brain network properties,

interplay to induce action percepts. The multitude of effects and their complexity requires empirical studies using both behavioral methods, imaging and electrophysiological recordings, and numerical simulation, as opposed to using only a single method of observation dominated by a sole theoretical argument.

.....

C. Brain networks for perceiving actions

An investigation of functional brain networks involved in action perception requires advanced imaging methods in addition to modelling and behavioral methods and thus goes beyond the scope of this thesis. However, a number of key references to imaging studies may be of use for different types of action stimuli, as biological motion perception (Beauchamp, Lee, Haxby, & Martin, 2003; Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Grezes & Decety, 2001; Howard et al., 1996; Saygin, Wilson, Hagler, Bates, & Sereno, 2004), grasping and hand actions, (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Cheng, Meltzoff, & Decety, 2007; Chong, Williams, Cunnington, & Mattingley, 2008; Cunnington, Windischberger, Robinson, & Moser, 2006; Gallagher & Frith, 2004; Wheaton, Thompson, Syngeniotis, Abbott, & Puce, 2004), dynamic face stimuli (Calvert & Campbell, 2003; Grosbras & Paus, 2006; Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005), and social interaction (King-Casas et al., 2005; Montague et al., 2002).

Until now a hierarchical signal processing architecture is one historically developed and well-studied viewpoint in action perception. This architecture was perhaps a starting point, influenced by the success in formalizing basic mechanisms in object recognition (see e.g. (Riesenhuber & Poggio, 1999) and follow ups for a summary). Of course, there is an understanding in the field that further studies need to consider the signal flow of action stimuli through brain networks more completely, and to include recurrent inter-area connections as well (Decety & Grezes, 1999; M. A. Giese & Poggio, 2003; Hamilton & Grafton, 2006; Lestou, Pollick, & Kourtzi, 2008; Van Overwalle & Baetens, 2009; Wurm & Lingnau, 2015). The localized brain region that has been

investigated most in the context of body action processing, is the superior temporal sulcus (STS). In this region one finds single neurons selective for specific actions (Oram & Perrett, 1996; Perrett et al., 1989; Perrett et al., 1985). STS is thought to be a hub of the action recognition network in humans as well (Allison, Puce, & McCarthy, 2000; E. Grossman et al., 2000; E. D. Grossman & Blake, 2002; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Puce & Perrett, 2003). However, interactions between STS and other brain regions are necessary as the action stimulus typically lasts sufficiently long (at least 1 second) to engage other cortical regions and thus inter-area interactions as well. Yet, it is understood how basic body motion processing mechanisms work in a feedforward pass to build up the STS neuronal signaling (M. A. Giese & Poggio, 2003). For perception of causality, we refer to a study of area F5 in monkeys (Caggiano, Fleischer, Pomper, Giese, & Thier, 2016), and for inferring the intention of an action, the inferior parietal and frontal motor cortices are important (Becchio et al., 2012; Fogassi & Luppino, 2005; Hamilton & Grafton, 2006; Iacoboni et al., 2005; Van Overwalle & Baetens, 2009). Our work focuses on using full body motion stimuli exclusively, behavioral measurements and numerical simulation.

D. Bistability phenomenon

Bistable perception refers to interpreting a single stimulus inducing two subjectively distinct perceptual experiences (referred to as 'percepts'), where only a single percept is experienced at a time (R. Blake & Logothetis, 2002; Leopold & Logothetis, 1999; Sterzer & Kleinschmidt, 2007). During continuous observation of a bistable stimulus, the transition between percepts happens spontaneously (Braun & Mattia, 2010; Leopold & Logothetis, 1999). From experimental data, the percept durations are random and appear to be Gamma-distributed (R. R. Blake, Fox, & McIntyre, 1971; J. W. Brascamp, Klink, & Levelt, 2015; Levelt, 1967; Murata, Matsui, Miyauchi, Kakita, & Yanagida, 2003; Walker & Powell, 1979). Variations of the percept duration distribution are possible depending on the experimental paradigm (Lehky, 1995;

Leopold & Logothetis, 1996; Logothetis, Leopold, & Sheinberg, 1996; Moreno-Bote, Shpiro, Rinzel, & Rubin, 2010; Rubin, 2003; J. Zhou, Reynaud, & Hess, 2014; Y. H. Zhou, Gao, White, Merk, & Yao, 2004). While we should be more precise with respect to how the percept durations are measured (this is elaborated below), the finding that the percept durations are Gamma-distributed appears to be observed across a variety of stimuli and types of measurement (Borsellino, De Marco, Allazetta, Rinesi, & Bartolini, 1972; J. W. Brascamp, van Ee, Pestman, & van den Berg, 2005; Walker, 1975; Y. H. Zhou et al., 2004).

There are many types of bistable stimuli that have been commonly used, and we refer to only some of them here. Binocular rivalry is the most prominent example (R. Blake & Logothetis, 2002; Leopold & Logothetis, 1999; van Ee, 2005). A methodologically convenient feature of binocular rivalry is that it enables a percept readout additional to the voluntary report, in case moving gratings of opposite orientation are used as rivalrous displays. Particularly, the opto-kinetic nystagmus (OKN) – which is the eye-movement trace following the perceived grating – is an indication of the percept. This direct read-out using OKN prompted its use in electrophysiological recordings in macaques, as both the trained voluntary report and the OKN trace contain current percept information that can be compared. The kinetic depth effect, such as a rotating sphere or rotating cylinder (Parker & Krug, 2003; Wallach & O'Connell, 1953) is another bistable example, in which a moving point-light display can induce the percept of a voluminous object (sphere or cylinder accordingly), seen rotating in one of the two mutually exclusive directions. A motion quartet is another example of a bistable display, as it is amenable to easy manipulation of perceptual dynamics, and thus featured in development of unique experimental paradigms addressing properties of perceptual hysteresis (Hock, Kelso, & Schoner, 1993; Hock, Schoner, & Giese, 2003) The earliest historical reference, as well as the simplest to demonstrate is the Necker cube ref (Necker, 1832).

In action perception, a point light biological motion display induces bistable perception with respect to the walking direction (de Lussanet & Lappe, 2012), where one way to bias perception towards one of the perceptual alternatives is by changing a depth cue (Jackson & Blake, 2010; Vanrie, Dekeyser, & Verfaillie, 2004). It holds more generally that perception of a bistable display towards one of the alternatives can be biased by a cue, e.g. aspect ratio biases apparent motion (Giese MA, 1995; Hock et al., 1993). This cue can then be used to investigate perceptual stability. As part of this work, we found that one previously well-studied static object perception cue – light source position – can modulate the (ref Chapter 2) walking direction percept.

E. A note on auditory bistability

Another popular setting to investigate bistability is auditory one. In the simplest case, two sequences of tones are concurrently presented to participants binaurally, so that they hear two distinct streams of auditory tones. One tone is considered a single symbol, so that two input sequences become two strings of that single symbol. Participants are probed if they are hearing two or more recurring string patterns, where each pattern recurs a combination of symbols from the two input ones. Just like in visual bistability, only one string pattern is heard at each instant of time. The analysis of such experimental data involves estimation of transition probabilities between discrete perceptual states. We consider it sufficiently distinct from the investigation path we follow in our work, where the questions are asked with a continuous state space in mind, as well as with an ability to continuously morph between stimulation patterns. However, it is worth referring to the literature for one popular auditory line of research by Denham et al. (Farkas et al., 2016; Farkas, Denham, & Winkler, 2018; Winkler, Denham, Mill, Bohm, & Bendixen, 2012), where the above considerations are elaborated in further detail. An auditory setting can be combined with a visual one, and there are studies on crossmodal interaction using

audiovisual bistable perception (Hupe, Joffo, & Pressnitzer, 2008; Sato, Basirat, & Schwartz, 2007).

F. Three studies modelling aspects of bistability

The first study we discuss is the model of binocular rivalry proposed by Hugh Wilson (Wilson, 2003). In that work, there are two competitive levels of neural populations: a monocular level, represented by two populations, and a binocular one – represented by one population. Each of the three populations has two neurons each tuned to one of two orthogonal grating stimuli. In the monocular population, the neuron's firing rate E_{l1} models the response to the grating preferred by the population. Its activity is modulated by the population tuned to the opposite orientation (inhibitory state variable I_{r2}), by the input stimulus V_{l1} and by the firing-rate adaptation of this neuron A_{l1} (at time-scale τ_A an order of magnitude slower than that of the firing rate) via an after-hyperpolarizing potential current. The inhibitory variable of the neuron within a monocular population I_{l1} is driven linearly by an excitatory neuron within same population. The second lower-level monocular population is symmetric. We refer the reader to the original publication for details on parameter values and simulation methodology. The equations of the lower layer are:

$$\begin{aligned}\tau_E \frac{d}{dt} E_{l1} &= -E_{l1} + \frac{100[V_{l1} - gl_{r2}]_+^2}{(10 + H_{l1})^2 + [V_{l1} - gl_{r2}]_+^2}, \\ \tau_I \frac{d}{dt} I_{l1} &= -I_{l1} + E_{l1}, \\ \tau_A \frac{d}{dt} A_{l1} &= -A_{l1} + aE_{l1}, \\ \tau_A &\gg \tau_E.\end{aligned}$$

The single higher-level (binocular) population is virtually same, where the input to a neuron tuned to each of the two gratings is pooled from the two monocular neurons tuned to the same grating. The equations for the higher layer are:

$$\begin{aligned}\tau_E \frac{d}{dt} E_{b1} &= -E_{b1} + \frac{100[c(E_{l1} + E_{r1}) - gI_{b2}]_+^2}{(10 + H_{b1})^2 + [c(E_{l1} + E_{r1}) - gI_{b2}]_+^2}, \\ \tau_I \frac{d}{dt} I_{b1} &= -I_{b1} + E_{b1}, \\ \tau_A \frac{d}{dt} A_{b1} &= -A_{b1} + aE_{b1}, \\ \tau_A &\gg \tau_E.\end{aligned}$$

The rivalry in the model occurs both at a lower (monocular) and at a higher (binocular) level. The model was proposed to resolve the issue of whether the competition between neurons indicative of the experience of rivalry is more prominent in lower visual cortical areas, such as V1, or in higher cortical areas, such as V4. Since then, modifications (J. Brascamp, Sohn, Lee, & Blake, 2013; H. H. Li, Rankin, Rinzel, Carrasco, & Heeger, 2017) of this model, as well as other models (Laing & Chow, 2002; Said & Heeger, 2013; Vattikuti et al., 2016) have been proposed for rivalry. The characteristic qualitative feature is that the bistable dynamics arises at more than one level and at different time scales.

The second study on bistability (Pastukhov et al., 2013) that we discuss here investigated: what qualitative features of neuronal circuitry contribute to bistable perceptual experience? In that and other modelling studies, including the one developed in this thesis (Chapter 3), it is most frequently understood that the right combination of three factors is necessary to exhibit bistable dynamics (Huguet, Rinzel, & Hupe, 2014; Shpiro, Moreno-Bote, Rubin, & Rinzel, 2009). These factors can be represented explicitly as parameters of a model neural circuit. These parameters are: (1) the level of noise and (2) the level of adaptation of neurons coding for target stimuli, and (3) the level of mutual inhibition between neurons coding for rival stimuli. The parameters need to be chosen from a certain range for the circuit to

operate in a bistable regime. Accordingly, the authors experimented with three different bistable displays to collect empirical data constraining their model: a kinetic depth multidot display (Wallach & O'Connell, 1953), a binocular rivalry display with orthogonally oriented green and red gratings ((Meng & Tong, 2004; Wheatstone, 1962), and the Necker cube (Necker, 1832). To match the simulation results to the experimental data, they used summary measures to constrain the parameter space of the model (Laing & Chow, 2002). In particular they used the mean percept duration, dominance duration, the coefficient of variation of the percept duration, the coefficient of correlation with dominance history and the time-constant of dominance history. For the detailed introduction to the last two measures and the reasoning behind them we refer to the earlier paper by (Pastukhov & Braun, 2011). The authors report that using these four measures enabled them to constrain the parameter region more than when only using the dominance duration statistics. They further argued that the perceptual history-dependent statistics are less sensitive to the contribution of the noise and thus constrain the search in parameter space in the dimension given by the noise-modulating constant. Therefore, including history-dependent measures in parameter search helps more precise identification of the realistic operating regime. It is indeed surprising how an approximately similar parameter range was obtained using three diverse bistable displays. The authors propose that their finding generalizes in the sense that this more constrained region should have a similar “shell-like” qualitative shape (as they describe it) for phenomenological models of other bistable stimuli.

The third study we review is the model by Rankin, Meso, Masson, Faugeras and Kornprobst (Rankin, Meso, Masson, Faugeras, & Kornprobst, 2014), who analyzed a model most closely related (in a technical sense) to the one we use here (Chapter 3). The authors perform a bifurcation analysis of a single-dimensional neural field model (Amari, 1977; Wilson & Cowan, 1972, 1973) with the stiffness of the firing rate function (described below) and the adaptation strength as parameters. The characteristic feature of the neural field for competition models is the ability to built in a continuous parameterization of stimulus features. In their case, as the authors

study the bistability induced by a barberpole illdrifting luminance grating (Wade, de Weert, & Swanston, 1984), they use a single (and periodic) feature – the grating orientation. With the neural field, they represent the evolution in time t of the activity of a single neural population $u(v, t)$, where v is said stimulus parameter. The activity (u) is normalized to values in the $[0, 1]$ range. Instantaneous propagation of both the excitatory and the inhibitory connectivity between neurons is described by a convolution operator $w(v)$. In this and some other neural field models (Bressloff & Webber, 2012; Deco & Roland, 2010; Detorakis & Rougier, 2012; Rankin, Tlapale, Veltz, Faugeras, & Kornprobst, 2013), including ours in Chapter 3, the so-called ‘Mexican hat’ kernel is used to model the short-range excitation and long-range inhibition in space, representing the orientation selectivity pattern in visual cortical neurons (Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Somers, Nelson, & Sur, 1995). The authors use adaptation with an-order-of-magnitude slower time scale to induce switching between modelled percepts. The stochastic variable X_t represents an Ornstein-Uhlenbeck process with the same time scale τ_A as the adaption variable $a(v, t)$. It is scaled by a parameter k_X chosen to match the switching time statistic. The model is as follows:

$$\begin{aligned} \tau_u \frac{d}{dt} u(v, t) &= -u(v, t) + S(\lambda[J(v) \star u(v, t) - k_a a(v, t) \\ &\quad + k_X X(v, t) + k_I I(v) - T]), \\ \tau_a \frac{d}{dt} a(v, t) &= -a(v, t) + u(v, t), \\ \tau_a &\gg \tau_u. \end{aligned}$$

The firing rate function $S(x)$ is a sigmoidal nonlinearity that bounds the field input. The explicit input bound is convenient for non-autonomous simulation, where the field input also represents the output of an additional complex signal processing (or statistical) model, as in our case (Chapter 3).

$$S(x) = \frac{1}{1 + \exp(-x)}.$$

However, we and several authors also preferred to use the Heaviside function, which also has the input bounding property and enables a certain simplification in the analysis of the model (Lefebvre, Hutt, Knebel, Whittingstall, & Murray, 2015). Nevertheless, Rankin et al. (Rankin et al., 2014) use the sigmoidal function with its stiffness parameter λ (and the adaptation gain k_a) to illustrate a 2-d bifurcation diagram with respect to its change. What makes the numerical bifurcation analysis (using the numerical continuation methods (E Doedel, 1997) with the AUTO97 package) possible, is the decomposition of field variables in components approximating $J(v)$ ((Curtu R, 2004), following the result by Veltz and Faugeras (Veltz, 2010). There it is shown, when using this approximation, that the remaining orthogonal components decay to zero in infinite time, and thus stationary and oscillatory solutions can be studied using ODE methods:

$$\begin{aligned} J(v) &= J_0 + 2J_1 \cos(v) + 2J_2 \cos(2v), \\ p(v, t) &= \hat{p}_0(t) + \hat{p}_1(t) \cos(v) + \hat{p}_2(t) \sin(v) + \hat{p}_3(t) \cos(2v) \\ &\quad + \hat{p}_4(t) \sin(2v) + \hat{p}_{orthogonal}, \\ a(v, t) &= \hat{a}_0(t) + \hat{a}_1(t) \cos(v) + \hat{a}_2(t) \sin(v) + \hat{a}_3(t) \cos(2v) \\ &\quad + \hat{a}_4(t) \sin(2v) + \hat{a}_{orthogonal}, \end{aligned}$$

Using the decomposition, Rankin et al. exhibit the qualitative difference in main switching mechanism, considering their model can also be used to represent the barberpole illusion (Castet, Charton, & Dufour, 1999). With low contrast in that stimulus, the perceptual switches are argued to be driven by noise, while with high contrast – by adaptation. The main contribution, however, is the use of numerical ODE methods for the neural field model. Not only the barberpole illusion, but also other bistable stimuli with continuous parametrization can be potentially represented by a

1-dimensional field. The Veltz-Faugeras decomposition, together with a demonstration of how an ODE integration and numerical continuation methods could be applied can give way to bifurcation analyses of other interpretable parameters.

In a sense, the basic structure of the neural field bistability model, and especially the practical challenge of parameter choice within the simulation, is the same as in the ODE models, like the one by Pastukhov et al. (Pastukhov et al., 2013), Wilson (Wilson, 2003) or others referenced above. There is mutual inhibition between neurons competing for perceptual dominance, self-induced slow adaptation for oscillation between percepts, and a noise process to model the randomness in percept duration.

The symmetric Mexican hat kernel in a one-dimensional neural field models a winner-takes-all mechanism. Due to long-range inhibitory coupling, the neural field exhibits winner-takes-all dynamics (Chen, McKinstry, & Edelman, 2013; Maass, 2000; Mao & Massaquoi, 2007): field neurons at a sufficiently long distance inhibit each other proportionally to their excitation. Two excited neurons exhibit transient activity right after stimulus onset, but with sufficient time one of them will suppress the other. The time required for suppression is controlled by the field dynamics time constant. The reason to bring this up is that in practice, depending on how long we wait to read out the neuronal activity after stimulus onset, we might consider the activity at readout time as the final “decision” that the neuronal population makes. Presenting the field with input and waiting for its dynamics to evolve into a stationary single bump can model “decision-making” as such. Here we do not discuss the decision making further and refer the reader to the recent paper by Klaes et al. (Klaes, Schneegans, Schoner, & Gail, 2012).

Interestingly, as in the last chapter gives, it is possible to use a two-dimensional neural field, where the Mexican hat kernel can be visualized as a two-dimensional surface. The second dimension there is asymmetric, and exploits the ability to model sequence-selectivity using feed-forward excitation and backward inhibition in a neural circuit (Xie & Giese, 2002; Zhang, 1996). However, the mathematical analysis

of a sequence-selective neural field, has only been performed for a one-dimensional field.

Other modelling applications of neural fields include motion integration (M. A. Giese, 1999), decision making (Klaes et al., 2012), visual working memory (Johnson, Spencer, Luck, & Schoner, 2009) and hallucinatory pattern formation in the visual cortex (Ermentrout & Cowan, 1979).

G. How to measure bistability

Simulation of the models, including ones described in this introduction and the one we used in our chapter (Chapter 3), involves matching experimental data. Human and animal experimental data may come in a different form than the simulation data, however. Because the phenomena underlying bistable perceptual switching are sufficiently complex, it is worth paying attention not only to the generative mechanisms, but also to the "readout functions" that match different data types in order to fit the same phenomenological description. The problem with obtaining the readout time behaviorally is that it may lack precision, because of including the time it takes the participant to report on the percept after seeing it. The problem with simulation is similar to that of the spiking data analysis, in that one has to deal with noisy non-stationary time series. If we want to obtain the percept durations behaviorally, the easiest thing to do is to present the bistable stimulus continuously and instruct participants to report a switch as soon as they experience one. Although there are some methodological considerations (e.g. the fact that the button press time includes the time needed to execute the motor command; here we do not elaborate on this problem further), the instruction would readily give the percept durations. This is not the situation when analyzing extracellular neural recordings (Leopold & Logothetis, 1996)! For example, accepting that we can reliably determine spike density functions of two sites selective for each of the two percepts, we still face, in practice, two noisy non-stationary time series. It is not obvious how to extract the

switching times from these data. We are unaware of any detailed study that addresses the issue of how to extract switching times from noisy spike measurements. However, we expect that the need for time-precise experimental measurements will increase with the ongoing development of neural signaling manipulation techniques such as optogenetics (Boyden, 2011). In the next paragraph we describe an approach that is directly applicable to both spike density data and to simulation data (where the neuronal activities are represented explicitly).

Conceptually, one wants to describe at every point of time, the present ‘operating stage’ of a neuron tuned to a percept, as well as of a neuron tuned to a rivalling percept. There are attempts in literature to link the moment at which a particular percept begins to be experienced to neural activity, when discussing it in terms of visual consciousness (Kang, Petzschner, Wolpert, & Shadlen, 2017). In practice, however one has samples of bivariate neural activity over a fixed time interval, and wants to infer: the point in time at which activity of one neuron became larger than that of the other neuron, and points in time at which each neurons’ activity trended from active to decreasing, from decreasing to completely inactive, and vice-versa. Moreover, one wants to minimize the number of “operating stages” of neural activity detected in the analysis.

It turns out that a recent approach to trend filtering (S. Kim, Koh, K., Boyd, S., Gorinevsky, D., 2009), gives precisely that kind of summary of bistable neural activity data.

Concretely, for an observation vector y (spike train) one obtains a piecewise-linear estimate $\hat{\beta}$ by solving a minimization problem defined as:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \|y - \beta\|_2^2 + \lambda \|D\beta\|_1$$

There, λ is a regularization parameter controlling the smoothness-residual tradeoff, and $\|\cdot\|_1$ denotes the l_1 norm. The difference matrix D is diagonal-constant with first row $[1 - 2 \ 1 \ 0 \dots 0]$ is chosen so that the second term above is minimized when

a triple of consecutive estimate points lie on the same line (same as the original Hodrick-Prescott filtering (HP-filtering; (Hodrick R., 1997)) by which it is inspired). The minimization problem is convex and can be solved in $O(n)$ time (with respect to number of data points) using interior point methods, but unlike the HP-filter, it gives a piecewise linear estimate of the data vector y . One convenience for the bistability data analysis is that we get linear trends, the time-points at which the trend changes (so called kinks), and consequently (not part of the filtering, but after applying it to data from rivaling neurons), can estimate the time at which one neuron stops to dominate the other one. The method itself allows for extension to incorporate periodic components, outlier detection, as well as extension to spatial trend estimates.

Linear trends give a more concise summary of behavior of rivaling neurons than a smoothing approach, like e.g. HP-filtering. Yet, they retain the features an experimentalist might be looking for, like the point in time when the neural activity starts to decrease. Because the recordings usually come in multiple trials, it is then possible to consider empirical distributions of the time points of trend change. Same is true for the switching time point.

One possible future prospect in studying visual cortical computation is to distinguish the roles of hierarchical levels in mid- and high-level perceptual dynamics. Electrophysiological recordings from more than one cortical site simultaneously will become one measurement to distinguish those. In a bistable stimulus experiment, one would want to consider stochastic dynamics at multiple sites, and the interaction between them instead of considering a singleton instance of a switching distribution within a circuit. One obvious question would then be – does the switching happen later at one site than at the other. There, more precise quantification of the point in time when one neuron becomes dominant over the other would be of immediate use.

H. Perceptual Adaptation

In our own study (Chapter 1), when studying perception of social interaction, we utilize the well-known psychophysical adaptation paradigm. In it, one can reduce the likelihood of perceiving either of the alternatives from a bistable stimulus by previously showing a participant the stimulus that induces that percept unambiguously. An idea of how behavioral adaptation aftereffects could be, in principle, indicative of neuronal tuning properties, was first put forward by von Bekeesy (Bekeesy, 1929). Exploiting neural adaptation, a multitude of methods have been proposed since then to investigate neural representation of certain stimulus classes. For example, the study of object recognition has enjoyed a multitude of papers exploring the fMRI-adaptation paradigm. Inferior temporal cortex has been a brain area of frequent focus when investigating population codes in object perception (Lehky, Kiani, Esteky, & Tanaka, 2011, 2014; Lehky & Tanaka, 2016). A comprehensive review, as well as a detailed study of adaptation mechanisms in area IT is found in the recent thesis of Kaliukhovich KU Leuven (Kaliukhovich, 2014); or see for example (M. A. Giese, Kuravi, P., Vogels, R., 2016) for a summary of population-level adaptive mechanisms in area IT.

There is also some indication that behavioral action adaptation effects are related to adaptive mechanisms in neuronal response. Whether (and how?) one can relate psychophysical action adaptation aftereffects to neuronal tuning is not completely conclusive at the moment. Yet, there does seem to be an overall link between behaviorally observed and neuronal-level aftereffects. For action perception, an argument that behavioral aftereffects can also elicit properties of neuronal tuning in area STS of primates was put forward by Barraclough (N. Barraclough & Jellema, 2011). In a neuroimaging study Thurman et al. (Thurman, van Boxtel, Monti, Chiang, & Lu, 2016) investigate, using point-light individual action stimuli, that adaptation in area pSTS can correlate with the behavioral aftereffects. Enhancement of action stimuli discrimination through visual learning has been shown to induce plasticity changes that make fMRI-adaptation effects more pronounced in motion-related area

hMT+ as well as pSTS by Jastroff et al. (Jastroff, Kourtzi, & Giese, 2009) – a study we discuss below.

We briefly discuss neural adaptation in general in the next section. After that we discuss several results on fMRI-measured adaptation in studying action perception. But first in this section we give detail on a couple of studies that exploit a purely behavioral adaptation paradigm in action perception. While differing in their research questions from our work, they are methodologically educational in using a psychophysical adaptation paradigm with action stimuli.

To recall, a perceptual action adaptation aftereffect might be measured as follows. We pick two action stimuli that, as reported by participants, unambiguously induce two distinct percepts. That is similar to how we pick the stimuli when measuring bistability. In fact, any bistable display can be used with an adaptation aftereffect measurement. We parameterize the signal space from which the stimuli are chosen in a way that enables to obtain any stimulus signal on a path between the two (treating the initial two stimuli as extreme values on this path). The position on the path is the parameter around which the aftereffect measurement will be designed. This way, how we parameterized and obtained all stimuli along the path between two extreme stimuli then determines what questions we ask. For instance, we may decide to render a volumetric walker stimulus to study biological motion perception (like Vangeneugden et al did in a series of studies (Vangeneugden et al., 2011; Vangeneugden, Peelen, Tadin, & Battelli, 2014; Vangeneugden et al., 2009), though not addressing adaptation in particular). We then may treat the rendering program as a tool that already contains such a parameterization: it takes motion capture data as input, and produces a 3D signal – a 1000ms video sequence of a single gait cycle. We then keep all parameters equal except the one that in our conception (as experiment designers) represents the “walking angle”. Using this parameter – walking angle – we produce two walkers corresponding to two of its values: e.g. -45 degrees of walking direction and 45 degrees of walking direction. As it is possible to generate stimuli with other angles of “walking direction” in the [-45, 45] range, we

will be asking a set of experimental questions on how the “walking direction” is represented in the brain. It is possible to study many properties of neural representation with such stimuli, by taking extracellular in vivo measurements in a local brain area. For example Vangeneugden et al. (Vangeneugden, 2010) studied area pSTS to investigate neural tuning properties with of the walking angle in macaques. With only behavioral measurements this concrete parametrization is problematic because some intermediate signals on the path would also induce stable percepts on their own (say a stimulus showing a 0 degree walking figure would induce its own percept of a 0 degree walking direction). What we want instead for an action adaptation aftereffect measurement is that none of the intermediate parameter values corresponds to an unambiguous percept. Alternatively, we might also take the two -45 and 45 walkers, again as two extreme stimuli signals, and morph the underlying 3D video signal by averaging each corresponding pixel, as such a parameterization will also give all signals on the path between the two extremes. This latter case turns out to be infeasible in practice, however, because none of the intermediate signals on a pixel-level morph path in the stimulus space induces an unambiguous percept.

Then why, do some parametrizations make experimental sense, and some don't? We currently understand that we want two things: first, every morph signal on the path to induce an unambiguous percept at each instant of time (or after a short viewing period); and, second, induced percepts from every point on the path to correspond to one of the initial extreme unambiguous percepts, and not to any other ones.

With two displays of action stimuli inducing stable percepts, as well as with a list of morphs along the parameterization path between them satisfying the two conditions above, we perform the following, rather simple, perceptual manipulation. We first present every action stimulus for a short period of time and ask which of the two extremes is perceived by a participant, thus determining the baseline (non-adapted) perceptual representation. We can also estimate experimentally the likelihood to perceive either of the extremes for each chosen position on the path. It is then simple

to measure an adaptation aftereffect (given an initially arbitrary pair of extremes and their parameterization). We choose that stimulus signal for which the likelihood to see either of the alternatives is equal (called point of subjective equality), which we denote further as a test stimulus (as well as in our study in Chapter 1). Finally, we again measure the likelihood to see either of the percepts from the test stimulus, but now prior to displaying it, we also present one of the two extreme stimuli signals to the participant (which we refer to as an adaptor stimulus). From the resulting data, we infer to have observed a behavioral adaptation aftereffect if the likelihood of perceiving the adaptor-induced percept from the subsequent test stimulus has decreased.

As a clarifying disclaimer – the adaptation aftereffect procedure is also common outside the action perception area, but we still felt it is worth describing it. The reason being that it can, conveniently, also serve as a follow-up constraining measurement to study bistable perception when investigating the same neural substrates, stimuli and participants (or even model animals in some cases).

Even if the mistakes (wrong parameterizations) in the walking direction example might have appeared easy to avoid, and the discussion above somewhat redundant, our ability to generate a list of intermediate stimulus morphs in practice does pose a difficult experimental problem. What we would measure experimentally is constrained by our ability to produce intermediate stimulus morphs satisfying the two properties above. For example, to morph between two arbitrary actions becomes highly complicated depending on what actions are chosen as extremes. Motion capture data (such as post-processed time series of joint angles) for actions of ‘catching’ and ‘taking’ allows simple averaging between the time series to produce the parameterization. But rendering stimuli for biomechanically different actions – given same type of data – becomes a current research problem in the area of computer graphics (working with motion capture data is out of scope of this thesis, yet a somewhat older work on statistical modeling of human interaction data is still illustrative of the point (Taubert, 2012)). For example, in Roether et al. (Roether,

Omlor, Christensen, & Giese, 2009) the investigation of perceiving emotion in action stimuli (while being still much simpler than the more general action morphing problem) already prompted more complex modelling of body movement data (M. Giese, Poggio, T., 2000) . Finally, given video data, instead of motion capture data, the task becomes currently technically unfeasible to best of our knowledge. Therefore, it is the ability to generate action morphs that determines what questions can be asked in studying action perception (with behavioral methods).

Barraclough and Jellema (N. Barraclough & Jellema, 2011) examined whether action adaptation aftereffects occur when the test and adaptor stimuli differ by viewpoint. They use displays of a static walking figure in addition to videos of a walking sequence to measure the aftereffect. In addition to concluding that walking direction inference mechanisms are not sensitive to the identity of the actor, they confirm that perceiving static and dynamic biological motion displays rely on similar mechanisms (broadly speaking). Perhaps more curiously, they show how parameterization by viewpoint is a simple yet effective way to discredit an aftereffect explanation by low-level motion adaptation mechanisms. In the paper, they claim that the walking direction recognition mechanisms are independent of the viewpoint in the test stimulus. The view point in the adaptor and test stimuli was different in one of the experiments. This enabled a direct comparison of aftereffect size when the test and adaptor had either same or opposite view. The authors did not find that the view of the test with respect to the view of the adaptor does has a significant effect on the adaptation aftereffect magnitude. While this does not by itself completely rule out a contribution of lower-level mechanisms to the observed action adaptation aftereffect, they give a direct comparison of aftereffects for stimuli where the low-level motion features are opposite.

Another study supports our argument about how much the morphing ability determines the question. For instance, because it's easy to morph between motion capture signals of the same actions, when the action is performed by actors of two different genders, it is possible to measure an adaptation aftereffect with respect to

such a gender parametrization. See, for example a short note by Jordan et al. (Jordan, Fallah, & Stoner, 2006). A much more extensive study by Troje et al. (Troje et al., 2006) also addressed adaptation derived from gender. However, the interesting part of that study is that the authors demonstrate an investigation of the concrete parameters of the adaptation paradigm: that is the adaptor duration, the interval of time between the adaptor and test (which they call the 'storage interval', or the interstimulus interval in other studies) and the test stimulus duration. Indeed, the data on the duration of the adaptation aftereffect is a reasonably constraining addition for finding the right simulation parameters in bistability models that include an adaptation effect. With action stimuli such as biological motion and hand actions, as well as for facial actions the increase of adaptor duration leads to an increase of the aftereffect size (though not going above a certain bound the exact value of which will depend on the experimental paradigm), while the decrease of the measured aftereffect size can be triggered by increasing the storage interval (N. E. Barraclough, Keith, Xiao, Oram, & Perrett, 2009; Becchio et al., 2012; Ghuman, McDaniel, & Martin, 2010; Hershenson, 1993; Leopold, Rhodes, Muller, & Jeffery, 2005; Magnussen & Johnsen, 1986; Rhodes, Jeffery, Clifford, & Leopold, 2007; Troje et al., 2006). In the work by Troje et al. (Troje et al., 2006) the authors do manage to find similar trends when varying the adaptor duration between 3.5 seconds and 14 seconds, and the storage interval between 0 seconds (it always must be a bit bigger than zero to avoid apparent motion effect to carry over from adaptor to test, but in some cases it can be very small to be treated as effectively zero at this time scale, especially when conceiving a numerical simulation after a psychophysics study) and 2.4 seconds. These time intervals are relatively big, so that the aftereffect measurements only speak about the dynamics of the behavioral readout of the underlying action processing by the brain. The manipulation of adaptor, interstimulus and test durations is illustrative in the sense of being strongly constraining for models of underlying dynamics, -- potentially even applicable to a completely different type of recording, e.g. to a BOLD signal or an extracellular potential signal.

With respect to the perceptual action adaptation aftereffect, we should also mention that it could be a useful marker in distinguishing autism spectrum disorder (Pavlova, 2012; Pellicano, Jeffery, Burr, & Rhodes, 2007) in humans using face (Pellicano et al., 2007) or biological motion stimuli (van Boxtel & Lu, 2013).

I. Neural Adaptation

The general term “adaptation” refers to a reduction of a measured spatially localized neuronal signal over time. This may refer to a reduced firing rate of an extracellularly measured spiking activity, to a decrease in a haemodynamic response (Clifford et al., 2007; Grill-Spector, 2006; Grill-Spector & Malach, 2001; Krekelberg, Boynton, & van Wezel, 2006), or similar phenomenon observed when working with other signals at different spatial and temporal scales (for instance using EEG and MEG measurement techniques (Ghuman, Bar, Dobbins, & Schnyer, 2008; Gilbert, Gotts, Carver, & Martin, 2010)). A standard experimental paradigm to induce adaptation is the repeated presentation of same stimulus, which is referred to as “repetition suppression” (Kohn, 2007; Wark, Lundstrom, & Fairhall, 2007; Webster, 2011).

One use of the paradigm has been to decode the neural representations of a certain stimuli class, as it differentiates between stimulus parameter modifications in repeated presentation that do or do not lead to adaptation. For example, static object representations in inferior temporal cortex have been studied with adaptation paradigms using both fMRI paradigms and electrophysiological recordings (Grill-Spector et al., 1999; Kaliukhovich & Vogels, 2011; Vuilleumier, Henson, Driver, & Dolan, 2002). There, with the firing rate signal, the time scale of response rate reduction is reported to be approximately 100-250 ms (M. A. Giese, Kuravi, P., Vogels, R., 2016; Kuravi & Vogels, 2017), which is smaller than the typical action stimulus presentation time. This already suggests that these paradigms could not directly be carried over from object perception to action perception.

When collecting fMRI response data in a repetition suppression paradigm, one may also see how the response decreases with the number of repetitions (Malach, 2012). Like in neural adaptation, this decrease is itself also limited so that it can saturate after six to eight stimulus repetitions (Grill-Spector & Malach, 2001). Important to keep in mind is that the nature of the relationship between fMRI adaptation and the neural adaptation is not completely understood. First, the detailed effects of neuronal spiking on the BOLD response remains to be studied (Bartels, Logothetis, & Moutoussis, 2008; D. S. Kim et al., 2004; Logothetis & Wandell, 2004; Sirotin, Cardoso, Lima, & Das, 2012). Second, it is not clear whether locally-assessed fMRI adaptation does not incorporate influences from other brain areas (Kohn & Movshon, 2003). Yet, while it is debated whether fMRI-adaptation can be interpreted in same way as adaptation observed electrophysiologically (Weigelt, Muckli, & Kohler, 2008), repetition suppression still gives an attractive paradigm to study neural representations. We reference its use in investigating perception of actions later.

While detailed quantitative properties of adaptation may vary between experimental paradigms, of principal interest to us is the qualitative reduction in neural response induced by repetition suppression.

Mechanistic explanations of adaptation include the decrease of firing rate (Miller & Desimone, 1994), the decrease in number of neurons responding to the stimulus ((Desimone, 1996; L. Li, Miller, & Desimone, 1993), or the decrease in response latency (James & Gauthier, 2006); also see (Gotts, Chow, & Martin, 2012; Grill-Spector, Henson, & Martin, 2006) for an elaborate discussion of different adaptation mechanisms).

Out of those three above, in modelling bistable perception (Chapter 3) we take the first point of view – when the neuronal response amplitude decreases already after several hundreds of milliseconds of stimulus repetition, and this reduction lasts up to, coarsely, 10 seconds. To clarify, we do not know of a detailed electrophysiological study that confirms that the firing rate adaptation in a certain proportion of biological motion stimuli sensitive neurons area pSTS will last for up to 10 seconds. The

important note here, instead, is the rough match of the particular type of adaptation time scale (500ms – 10s) to the common stimuli presentation time scale within experiments using that stimuli class (i.e. action stimuli). For instance, a recent investigation by Kuravi et al. (Kuravi, Caggiano, Giese, & Vogels, 2016) probed for repetition-suppression-induced adaptation using 1.4 second hand action stimulus using single-unit and multi-unit electrophysiology in STS, where they demonstrate the decrease in both firing rate and LFP activity.

J. Use of adaptation in study of action perception

It is interesting that changes of the neural representation of actions at a much larger visual learning time scale (one to few days) can be employed to affect the representation at the adaptation time scale. For example, Jastorff et al (Jastorff et al., 2009) have studied whether visual learning of point-light action stimuli enhances discriminability of similar movements. The authors use spatiotemporal morphing of action patterns to generate point-light displays with different degrees of discriminability between the pairs of patterns (refer to e.g. (M. A. Giese & Lappe, 2002) for more detail on the relationship between perceptual similarity of spatiotemporal morphs). One of their experiments shows emergence of observable fMRI-adaptation through neural plasticity induced in participants by repetitive execution of discrimination task. Before long-term plasticity-related processes are triggered (that is – before training simply put; refer to (Berninger & Bi, 2002; Bi & Poo, 2001; Fremaux & Gerstner, 2015; Fusi, Drew, & Abbott, 2005; Messinger, Squire, Zola, & Albright, 2001, 2005; Yang & Calakos, 2013) for information on learning-related neural plasticity time scales and mechanisms, though not specifically in the action perception context), the authors probe how much adaptation can be observed in biological motion related areas FBA and pSTS (E. D. Grossman & Blake, 2002; Michels, Lappe, & Vaina, 2005; Peelen, Wiggett, & Downing, 2006; Peuskens, Vanrie, Verfaillie, & Orban, 2005), as well in lower-level motion related areas hMT+ (Born &

Bradley, 2005; Herrington, Nymberg, Faja, Price, & Schultz, 2012; Sani et al., 2010). After training, they repeated the same measurement. By comparing the pre- and post-training adaptation statistics, they inferred an increase in fMRI-selective adaptation in biological-motion-related areas. More surprisingly, they inferred the appearance of fMRI-selective adaptation effects in low-level motion areas after training (with none present in pre-training measurement). We find it to be a potentially curious practical trick: perhaps visual training for discrimination can help to observe the fMRI-selective adaptation aftereffects, when the chosen stimuli patterns are not initially producing the aftereffect. With respect to the full paper contribution itself, the authors relate functional changes in action processing in different brain areas to the behavioral performance.

Grossman et al. (E. D. Grossman, Jardine, & Pyles, 2010) addressed what action stimulus parameters the neural representations of actions in area STS might be invariant to? Using an fMRI-adaptation paradigm and point-light action stimuli, they expectedly observed specificity with respect to the action viewpoint in extrastriate hMT+. Yet, they concluded (partial) invariance to position, point of view and size of the action stimuli, when analyzing the haemodynamic response functions from pSTS. Hence, this particular study of invariance to action stimuli features using of fMRI-adaptation is consistent with previous physiological findings (Jellema & Perrett, 2006). Of course, the method enabled the authors to look at other brain regions as well, besides the hMT+ and pSTS areas. However, they duly point out that the conclusions from fMRI-adaptation in humans about those two areas are backed by both physiological studies and the anatomical homology of monkey and human visual cortices. Hence, invariance/sensitivity properties in other areas that they (can technically) compute is not necessarily conclusive without similar support. Of course, such limitation of the fMRI-adaptation method is constraining to studies with other sensorimotor stimuli classes, beyond action perception, as well. Yet, it is particularly critical for further study of perception of actions, because at the action stimulus time scale (0.6-20 seconds), we do indeed want to look at brain networks and inter-architectural interactions holistically, beyond feature sensitivities of singleton areas.

In a more recent study (Thurman et al., 2016) have directly assessed the BOLD signal decrease in the repetition suppression paradigm using point-light walker stimuli in both human middle temporal complex (hMT+) and posterior STS. Thurman et al use a 25 seconds interval for an initial adaptor duration phase before every block, during which they display a point-light biological motion stimulus (it was a runner, a walker or parameterized morph between them). In the repetitive adaptation trials, just like in our action adaptation paradigm (Chapter 1), it was sufficient to use a 4.5 second adaptation block. What could be more critical for the technical execution of the paradigm, as well as for potential ecologically valid simulations of it (using neural modelling techniques as in Chapter 3), -- is the use of a variable (3 to 7 second) interstimulus interval (ISI). The authors refer to (Dale, 1999; Serences, 2004) as an improvement to extracting event-related responses from BOLD signal data.

The question the authors of the above article intended to ask, however, is of utmost importance to our own study (Chapter 1). That is: when using repetition suppression with action stimuli, can one relate adaptation in posterior STS at a neural population level to behaviorally observed perceptual adaptation? If the answer is positive, then it might be possible to develop paradigms inferring properties of neural representation using fast-to-execute psychophysical methods. To be clear, previous probes of fMRI-adaptation aftereffects in pSTS have already successfully used repetition suppression to study neural representation of action stimuli at a sub-voxel resolution (Grill-Spector & Malach, 2001; Krekelberg et al., 2006; Webster, 2011). Yet, they have not directly tested how much the behavioral correlate of the aftereffect (i.e. perceptual adaptation) is representative of its fMRI-adaptation counterpart.

In hMT+, similarly to the conclusions from the Jastorff et al. study (Jastorff et al., 2009) that we discussed above, Thurman et al. confirm location-specific adaptation effects . This conclusion is in line with the general understanding of feedforward biological motion perception theory (M. A. Giese & Poggio, 2003). Further, using a measure of neural aftereffect strength, and of behavioral aftereffect strength, Thurman et al find a positive correlation within the right pSTS. They did not, unfortunately, confirm that

aftereffects are significantly correlated when looking at the left pSTS. Moreover, they note that the behavioral responses were collected prior to the fMRI scan sessions.

Further analysis in the paper (Thurman et al., 2016) focuses on attempting to elicit fMRI-adaptation aftereffects as a marker for autism traits. This is indeed an important objective, as well a potentially promising research venue. More generally speaking, from biological motion stimuli one may expand to study perception of social interactions (as we attempt to do in our Chapter 1). Subsequently, one may attempt to develop fast-to-execute behavioral procedures to measure indications of Autism Spectrum Disorder (ASD), and even therapies in the future. An elaborate discussion of Thurman's results as well of ASD studies is out of scope of this thesis, however. For completeness, we give the references for the reader who may want to look at use of biological motion stimuli in ASD studies: (Pavlova, 2012; Pellicano et al., 2007; van Boxtel & Lu, 2013).

As it stands, complete empirical evidence relating fMRI-adaptation in pSTS to behavioral adaptation aftereffects induced when using action stimuli is lacking. This might perhaps be complicated by the long time scale of an action display that enables the stimulation signal to traverse the whole brain before the corresponding BOLD signal is read out. Despite this, we are not prevented from mining further patterns using purely behavioral methods. We hope this thesis will support basic mechanistic understanding of the relationship of the neural and behavioral aftereffects in action perception in the future.

- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nat Neurosci*, 7(10), 1057-1058. doi:10.1038/nn1312
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn Sci*, 4(7), 267-278.
- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern*, 27(2), 77-87.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr Biol*, 16(18), 1818-1823. doi:10.1016/j.cub.2006.07.060
- Barracough, N., & Jellema, T. (2011). Visual aftereffects for walking actions reveal underlying neural mechanisms for action recognition. *Psychol Sci*, 22(1), 87-94. doi:10.1177/0956797610391910

- Barraclough, N. E., Keith, R. H., Xiao, D., Oram, M. W., & Perrett, D. I. (2009). Visual adaptation to goal-directed hand actions. *J Cogn Neurosci*, *21*(9), 1806-1820. doi:10.1162/jocn.2008.21145
- Bartels, A., Logothetis, N. K., & Moutoussis, K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends Neurosci*, *31*(9), 444-453. doi:10.1016/j.tins.2008.06.004
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci*, *15*(7), 991-1001. doi:10.1162/089892903770007380
- Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: from mirroring to mentalizing. *Neuroimage*, *61*(1), 240-248. doi:10.1016/j.neuroimage.2012.03.013
- Bekesy, G. v. (1929). Theorie des Hörens: über die Bestimmung des einem reinen Tonempfinden entsprechenden Erregungsgebietes der Basilar membran vermittelt Ermüdungserscheinungen. *Physikalische Zeitschrift*, *30*, 115-125.
- Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci U S A*, *92*(9), 3844-3848.
- Berninger, B., & Bi, G. Q. (2002). Synaptic modification in neural circuits: a timely action. *Bioessays*, *24*(3), 212-222. doi:10.1002/bies.10060
- Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu Rev Neurosci*, *24*, 139-166. doi:10.1146/annurev.neuro.24.1.139
- Blake, R., & Logothetis, N. (2002). Visual competition. *Nat Rev Neurosci*, *3*(1), 13-21. doi:10.1038/nrn701
- Blake, R. R., Fox, R., & McIntyre, C. (1971). Stochastic properties of stabilized-image binocular rivalry alternations. *J Exp Psychol*, *88*(3), 327-332.
- Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci*, *16*(11), 3737-3744.
- Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annu Rev Neurosci*, *28*, 157-189. doi:10.1146/annurev.neuro.26.041002.131052
- Borsellino, A., De Marco, A., Allazetta, A., Rinesi, S., & Bartolini, B. (1972). Reversal time distribution in the perception of visual ambiguous stimuli. *Kybernetik*, *10*(3), 139-144.
- Boyden, E. S. (2011). A history of optogenetics: the development of tools for controlling brain circuits with light. *F1000 Biol Rep*, *3*, 11. doi:10.3410/B3-11
- Brascamp, J., Sohn, H., Lee, S. H., & Blake, R. (2013). A monocular contribution to stimulus rivalry. *Proc Natl Acad Sci U S A*, *110*(21), 8337-8344. doi:10.1073/pnas.1305393110
- Brascamp, J. W., Klink, P. C., & Levelt, W. J. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Res*, *109*(Pt A), 20-37. doi:10.1016/j.visres.2015.02.019
- Brascamp, J. W., van Ee, R., Pestman, W. R., & van den Berg, A. V. (2005). Distributions of alternation rates in various forms of bistable perception. *J Vis*, *5*(4), 287-298. doi:10.1167/5.4.1
- Braun, J., & Mattia, M. (2010). Attractors and noise: twin drivers of decisions and multistability. *Neuroimage*, *52*(3), 740-751. doi:10.1016/j.neuroimage.2009.12.126
- Bressloff, P. C., & Webber, M. A. (2012). Neural field model of binocular rivalry waves. *J Comput Neurosci*, *32*(2), 233-252. doi:10.1007/s10827-011-0351-y
- Brewster, D. (1844). On the conversion of relief by inverted vision. *Edinburgh Philosophical Transactions*, *15*(4), 657-662. doi:<https://doi.org/10.1017/S0080456800030234>

- Caggiano, V., Fleischer, F., Pomper, J. K., Giese, M. A., & Thier, P. (2016). Mirror Neurons in Monkey Premotor Area F5 Show Tuning for Critical Features of Visual Causality Perception. *Curr Biol*, *26*(22), 3077-3082. doi:10.1016/j.cub.2016.10.007
- Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J Cogn Neurosci*, *15*(1), 57-70. doi:10.1162/089892903321107828
- Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb Cortex*, *15*(8), 1243-1249. doi:10.1093/cercor/bhi007
- Castet, E., Charton, V., & Dufour, A. (1999). The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Res*, *39*(5), 915-932.
- Chen, Y., McKinstry, J. L., & Edelman, G. M. (2013). Versatile networks of simulated spiking neurons displaying winner-take-all behavior. *Front Comput Neurosci*, *7*, 16. doi:10.3389/fncom.2013.00016
- Cheng, Y., Meltzoff, A. N., & Decety, J. (2007). Motivation modulates the activity of the human mirror-neuron system. *Cereb Cortex*, *17*(8), 1979-1986. doi:10.1093/cercor/bhl107
- Chong, T. T., Williams, M. A., Cunnington, R., & Mattingley, J. B. (2008). Selective attention modulates inferior frontal gyrus activity during action observation. *Neuroimage*, *40*(1), 298-307. doi:10.1016/j.neuroimage.2007.11.030
- Clifford, C. W., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: neural, psychological and computational aspects. *Vision Res*, *47*(25), 3125-3131. doi:10.1016/j.visres.2007.08.023
- Cunnington, R., Windischberger, C., Robinson, S., & Moser, E. (2006). The selection of intended actions and the observation of others' actions: a time-resolved fMRI study. *Neuroimage*, *29*(4), 1294-1302. doi:10.1016/j.neuroimage.2005.09.028
- Curtu R, E. B. (2004). Pattern Formation in a Network of Excitatory and Inhibitory Cells with Adaptation. *SIAM J. Appl. Dyn. Syst.*, *3*(3), 191-231. doi:10.1137/030600503
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Hum Brain Mapp*, *8*(2-3), 109-114.
- de Lussanet, M. H., & Lappe, M. (2012). Depth perception from point-light biological motion displays. *J Vis*, *12*(11). doi:10.1167/12.11.14
- Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn Sci*, *3*(5), 172-178.
- Deco, G., & Roland, P. (2010). The role of multi-area interactions for the computation of apparent motion. *Neuroimage*, *51*(3), 1018-1026. doi:10.1016/j.neuroimage.2010.03.032
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A*, *93*(24), 13494-13499.
- Detorakis, G. I., & Rougier, N. P. (2012). A neural field model of the somatosensory cortex: formation, maintenance and reorganization of ordered topographic maps. *PLoS One*, *7*(7), e40257. doi:10.1371/journal.pone.0040257
- E Doedel, A. C., TF Fairgrieve, YA Kuznetsov, B Sandstede and X Wang. (1997). *AUTO97: Continuation and bifurcation software for ordinary differential equations (with HOMCONT)*. Retrieved from
- Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. *Biol Cybern*, *34*(3), 137-150.
- Escobar, M., Kornprobst P. (2008). Action recognition with a bioinspired feedforward motion processing model: the richness of center-surround interactions. *Computer Vision - ECCV 2008*, *2*, 186-199. doi:10.1007/978-3-540-88688-4

- Farkas, D., Denham, S. L., Bendixen, A., Toth, D., Kondo, H. M., & Winkler, I. (2016). Auditory Multi-Stability: Idiosyncratic Perceptual Switching Patterns, Executive Functions and Personality Traits. *PLoS One*, *11*(5), e0154810. doi:10.1371/journal.pone.0154810
- Farkas, D., Denham, S. L., & Winkler, I. (2018). Functional brain networks underlying idiosyncratic switching patterns in multi-stable auditory perception. *Neuropsychologia*, *108*, 82-91. doi:10.1016/j.neuropsychologia.2017.11.032
- Ferstl, Y., Bulthoff, H., & de la Rosa, S. (2017). Action recognition is sensitive to the identity of the actor. *Cognition*, *166*, 201-206. doi:10.1016/j.cognition.2017.05.036
- Fleming, R. W., Holtmann-Rice, D., & Bulthoff, H. H. (2011). Estimation of 3D shape from image orientations. *Proc Natl Acad Sci U S A*, *108*(51), 20438-20443. doi:10.1073/pnas.1114619109
- Fogassi, L., & Luppino, G. (2005). Motor functions of the parietal lobe. *Curr Opin Neurobiol*, *15*(6), 626-631. doi:10.1016/j.conb.2005.10.015
- Fremaux, N., & Gerstner, W. (2015). Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. *Front Neural Circuits*, *9*, 85. doi:10.3389/fncir.2015.00085
- Fusi, S., Drew, P. J., & Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron*, *45*(4), 599-611. doi:10.1016/j.neuron.2005.02.001
- Gallagher, H. L., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia*, *42*(13), 1725-1736. doi:10.1016/j.neuropsychologia.2004.05.006
- Gerardin, P., Kourtzi, Z., & Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proc Natl Acad Sci U S A*, *107*(37), 16309-16314. doi:10.1073/pnas.1006285107
- Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proc Natl Acad Sci U S A*, *105*(24), 8405-8409. doi:10.1073/pnas.0710674105
- Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Curr Biol*, *20*(1), 32-36. doi:10.1016/j.cub.2009.10.077
- Giese M A, S. G., Hock H S. (1995). *The Dynamics of the Perceptual Organization in Apparent Motion*. Paper presented at the Neural Networks: Artificial Intelligence and Industrial Applications.
- Giese, M., Poggio, T. (2000). Morphable Models for the Analysis and Synthesis of Complex Motion Patterns. *International Journal of Computer Vision*, 59-73. doi:10.1023/A:1008118801668
- Giese, M. A. (1999). *Dynamic Neural Field Theory for Motion Perception*: Springer US.
- Giese, M. A., Kuravi, P., Vogels, R. (2016). *Phenomenological Model for the Adaptation of Shape-Selective Neurons in Area IT*. Paper presented at the Artificial Neural Networks and Machine Learning
- Giese, M. A., & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Res*, *42*(15), 1847-1858.
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci*, *4*(3), 179-192. doi:10.1038/nrn1057
- Gilbert, J. R., Gotts, S. J., Carver, F. W., & Martin, A. (2010). Object repetition leads to local increases in the temporal coordination of neural responses. *Front Hum Neurosci*, *4*, 30. doi:10.3389/fnhum.2010.00030
- Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition priming and repetition suppression: Multiple mechanisms in need of testing. *Cogn Neurosci*, *3*(3-4), 250-259. doi:10.1080/17588928.2012.697054

- Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum Brain Mapp*, *12*(1), 1-19.
- Grill-Spector, K. (2006). Selectivity of adaptation in single units: implications for fMRI experiments. *Neuron*, *49*(2), 170-171. doi:10.1016/j.neuron.2006.01.004
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci*, *10*(1), 14-23. doi:10.1016/j.tics.2005.11.006
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187-203.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)*, *107*(1-3), 293-321.
- Grosbras, M. H., & Paus, T. (2006). Brain networks involved in viewing angry hands or faces. *Cereb Cortex*, *16*(8), 1087-1096. doi:10.1093/cercor/bhj050
- Grossberg, S., Kuhlmann, L., & Mingolla, E. (2007). A neural model of 3D shape-from-texture: multiple-scale filtering, boundary grouping, and surface filling-in. *Vision Res*, *47*(5), 634-672. doi:10.1016/j.visres.2006.10.024
- Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *J Cogn Neurosci*, *12*(5), 711-720.
- Grossman, E. D., & Blake, R. (2002). Brain Areas Active during Visual Perception of Biological Motion. *Neuron*, *35*(6), 1167-1175.
- Grossman, E. D., Jardine, N. L., & Pyles, J. A. (2010). fMR-Adaptation Reveals Invariant Coding of Biological Motion on the Human STS. *Front Hum Neurosci*, *4*, 15. doi:10.3389/neuro.09.015.2010
- Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *J Neurosci*, *26*(4), 1133-1137. doi:10.1523/JNEUROSCI.4551-05.2006
- Herrington, J., Nymberg, C., Faja, S., Price, E., & Schultz, R. (2012). The responsiveness of biological motion processing areas to selective attention towards goals. *Neuroimage*, *63*(1), 581-590. doi:10.1016/j.neuroimage.2012.06.077
- Hershenson, M. (1993). Linear and rotation motion aftereffects as a function of inspection duration. *Vision Res*, *33*(14), 1913-1919.
- Hock, H. S., Kelso, J. A., & Schoner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *J Exp Psychol Hum Percept Perform*, *19*(1), 63-80.
- Hock, H. S., Schoner, G., & Giese, M. (2003). The dynamical foundations of motion pattern formation: stability, selective adaptation, and perceptual continuity. *Percept Psychophys*, *65*(3), 429-457.
- Hodrick R., P., E. (1997). Postwar U.S. business cycles: An empirical investigation. *J. Money, Credit, and Banking*, *29*(1), 1-16.
- Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., & Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Curr Biol*, *6*(8), 1015-1019.
- Huguet, G., Rinzal, J., & Hupe, J. M. (2014). Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J Vis*, *14*(3), 19. doi:10.1167/14.3.19
- Hupe, J. M., Joffo, L. M., & Pressnitzer, D. (2008). Bistability for audiovisual stimuli: Perceptual decision is modality specific. *J Vis*, *8*(7), 1 1-15. doi:10.1167/8.7.1

- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol*, 3(3), e79. doi:10.1371/journal.pbio.0030079
- Jackson, S., & Blake, R. (2010). Neural integration of information specifying human structure from form, motion, and depth. *J Neurosci*, 30(3), 838-848. doi:10.1523/JNEUROSCI.3116-09.2010
- James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Hum Brain Mapp*, 27(1), 37-46. doi:10.1002/hbm.20165
- Jastorff, J., Kourtzi, Z., & Giese, M. A. (2009). Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci*, 29(44), 14026-14038. doi:10.1523/JNEUROSCI.3070-09.2009
- Jellema, T., & Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia*, 44(9), 1535-1546. doi:10.1016/j.neuropsychologia.2006.01.020
- Jhuang, H., Serre, T., Wolf, L., Poggio, T. (2007). A biologically inspired system for action recognition. *IEEE 11th International Conference on Computer Vision*, 1-8.
- Johnson, J. S., Spencer, J. P., Luck, S. J., & Schoner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychol Sci*, 20(5), 568-577. doi:10.1111/j.1467-9280.2009.02329.x
- Jordan, H., Fallah, M., & Stoner, G. R. (2006). Adaptation of gender derived from biological motion. *Nat Neurosci*, 9(6), 738-739. doi:10.1038/nn1710
- Kaliukhovich, D. A. (2014). *Mechanisms of adaptation in macaque inferior temporal cortex*. (PhD), KU Leuven.
- Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb Cortex*, 21(7), 1547-1558. doi:10.1093/cercor/bhq207
- Kang, Y. H. R., Petzschner, F. H., Wolpert, D. M., & Shadlen, M. N. (2017). Piercing of Consciousness as a Threshold-Crossing Operation. *Curr Biol*, 27(15), 2285-2295 e2286. doi:10.1016/j.cub.2017.06.047
- Kim, D. S., Ronen, I., Olman, C., Kim, S. G., Ugurbil, K., & Toth, L. J. (2004). Spatial relationship between neuronal activity and BOLD functional MRI. *Neuroimage*, 21(3), 876-885. doi:10.1016/j.neuroimage.2003.10.018
- Kim, S., Koh, K., Boyd, S., Gorinevsky, D. (2009). 11 Trend Filtering. *SIAM Review*, 51(2), 339-360.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308(5718), 78-83. doi:10.1126/science.1108062
- Klaes, C., Schneegans, S., Schoner, G., & Gail, A. (2012). Sensorimotor learning biases choice behavior: a learning neural field model for decision making. *PLoS Comput Biol*, 8(11), e1002774. doi:10.1371/journal.pcbi.1002774
- Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Percept Psychophys*, 52(1), 18-36.
- Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *J Neurophysiol*, 97(5), 3155-3164. doi:10.1152/jn.00086.2007
- Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, 39(4), 681-691.
- Krekelberg, B., Boynton, G. M., & van Wezel, R. J. (2006). Adaptation: from single cells to BOLD signals. *Trends Neurosci*, 29(5), 250-256. doi:10.1016/j.tins.2006.02.008

- Kuravi, P., Caggiano, V., Giese, M., & Vogels, R. (2016). Repetition suppression for visual actions in the macaque superior temporal sulcus. *J Neurophysiol*, *115*(3), 1324-1337. doi:10.1152/jn.00849.2015
- Kuravi, P., & Vogels, R. (2017). Effect of adapter duration on repetition suppression in inferior temporal cortex. *Sci Rep*, *7*(1), 3162. doi:10.1038/s41598-017-03172-3
- Laing, C. R., & Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J Comput Neurosci*, *12*(1), 39-53.
- Layher, G., Giese, M., Neumann, H. (2014). Learning Representations of Animated Motion Sequences - A Neural Model. *Topics in Cognitive Science*, *6*(1), 170-182. doi:10.1111/tops.12075
- Lefebvre, J., Hutt, A., Knebel, J. F., Whittingstall, K., & Murray, M. M. (2015). Stimulus statistics shape oscillations in nonlinear recurrent neural networks. *J Neurosci*, *35*(7), 2895-2903. doi:10.1523/JNEUROSCI.3609-14.2015
- Lehky, S. R. (1995). Binocular rivalry is not chaotic. *Proc Biol Sci*, *259*(1354), 71-76. doi:10.1098/rspb.1995.0011
- Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2011). Statistics of visual responses in primate inferotemporal cortex to object stimuli. *J Neurophysiol*, *106*(3), 1097-1117. doi:10.1152/jn.00990.2010
- Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2014). Dimensionality of object representations in monkey inferotemporal cortex. *Neural Comput*, *26*(10), 2135-2162. doi:10.1162/NECO_a_00648
- Lehky, S. R., & Tanaka, K. (2016). Neural representation for object recognition in inferotemporal cortex. *Curr Opin Neurobiol*, *37*, 23-35. doi:10.1016/j.conb.2015.12.001
- Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*(6565), 549-553. doi:10.1038/379549a0
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn Sci*, *3*(7), 254-264.
- Leopold, D. A., Rhodes, G., Muller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proc Biol Sci*, *272*(1566), 897-904. doi:10.1098/rspb.2004.3022
- Lestou, V., Pollick, F. E., & Kourtzi, Z. (2008). Neural substrates for action understanding at different description levels in the human brain. *J Cogn Neurosci*, *20*(2), 324-341. doi:10.1162/jocn.2008.20021
- Levelt, W. J. (1967). Note on the distribution of dominance times in binocular rivalry. *Br J Psychol*, *58*(1), 143-145.
- Li, H. H., Rankin, J., Rinzel, J., Carrasco, M., & Heeger, D. J. (2017). Attention model of binocular rivalry. *Proc Natl Acad Sci U S A*, *114*(30), E6192-E6201. doi:10.1073/pnas.1620475114
- Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol*, *69*(6), 1918-1929. doi:10.1152/jn.1993.69.6.1918
- Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature*, *380*(6575), 621-624. doi:10.1038/380621a0
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu Rev Physiol*, *66*, 735-769. doi:10.1146/annurev.physiol.66.082602.092845
- Maass, W. (2000). On the computational power of winner-take-all. *Neural Comput*, *12*(11), 2519-2535.
- Magnussen, S., & Johnsen, T. (1986). Temporal aspects of spatial adaptation. A study of the tilt aftereffect. *Vision Res*, *26*(4), 661-672.

- Malach, R. (2012). Targeting the functional properties of cortical neurons using fMR-adaptation. *Neuroimage*, *62*(2), 1163-1169. doi:10.1016/j.neuroimage.2012.01.002
- Mao, Z. H., & Massaquoi, S. G. (2007). Dynamics of winner-take-all competition in recurrent neural networks with lateral inhibition. *IEEE Trans Neural Netw*, *18*(1), 55-69. doi:10.1109/TNN.2006.883724
- Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J Vis*, *4*(7), 539-551. doi:10.1167/4.7.2
- Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proc Natl Acad Sci U S A*, *98*(21), 12239-12244. doi:10.1073/pnas.211431098
- Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2005). Neural correlates of knowledge: stable representation of stimulus associations across variations in behavioral performance. *Neuron*, *48*(2), 359-371. doi:10.1016/j.neuron.2005.08.035
- Michels, L., Lappe, M., & Vaina, L. M. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *Neuroreport*, *16*(10), 1037-1041.
- Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146), 520-522.
- Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biol Cybern*, *53*(3), 137-151.
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., . . . Fisher, R. E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage*, *16*(4), 1159-1164.
- Moreno-Bote, R., Shpiro, A., Rinzel, J., & Rubin, N. (2010). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J Vis*, *10*(11), 1. doi:10.1167/10.11.1
- Murata, T., Matsui, N., Miyauchi, S., Kakita, Y., & Yanagida, T. (2003). Discrete stochastic process underlying perceptual rivalry. *Neuroreport*, *14*(10), 1347-1352. doi:10.1097/01.wnr.0000077553.91466.41
- Necker, L. A. (1832). Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science*, *1*(5), 329-337. doi:doi:10.1080/14786443208647909
- Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol*, *76*(1), 109-129. doi:10.1152/jn.1996.76.1.109
- Parker, A. J., & Krug, K. (2003). Neuronal mechanisms for the perception of ambiguous stimuli. *Curr Opin Neurobiol*, *13*(4), 433-439.
- Pastukhov, A., & Braun, J. (2011). Cumulative history quantifies the role of neural adaptation in multistable perception. *J Vis*, *11*(10). doi:10.1167/11.10.12
- Pastukhov, A., Garcia-Rodriguez, P. E., Haenicke, J., Guillamon, A., Deco, G., & Braun, J. (2013). Multi-stable perception balances stability and sensitivity. *Front Comput Neurosci*, *7*, 17. doi:10.3389/fncom.2013.00017
- Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb Cortex*, *22*(5), 981-995. doi:10.1093/cercor/bhr156
- Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, *49*(6), 815-822. doi:10.1016/j.neuron.2006.02.004

- Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Curr Biol*, *17*(17), 1508-1512. doi:10.1016/j.cub.2007.07.065
- Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb Cortex*, *15*(12), 1866-1876. doi:10.1093/cercor/bhi064
- Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., . . . Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J Exp Biol*, *146*, 87-113.
- Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., . . . Jeeves, M. A. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Behav Brain Res*, *16*(2-3), 153-170.
- Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *Eur J Neurosci*, *21*(10), 2864-2875. doi:10.1111/j.1460-9568.2005.04106.x
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci*, *18*(6), 2188-2199.
- Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci*, *358*(1431), 435-445. doi:10.1098/rstb.2002.1221
- Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, *331*(6152), 163-166. doi:10.1038/331163a0
- Rankin, J., Meso, A. I., Masson, G. S., Faugeras, O., & Kornprobst, P. (2014). Bifurcation study of a neural field competition model with an application to perceptual switching in motion integration. *J Comput Neurosci*, *36*(2), 193-213. doi:10.1007/s10827-013-0465-5
- Rankin, J., Tlapale, E., Veltz, R., Faugeras, O., & Kornprobst, P. (2013). Bifurcation analysis applied to a model of motion integration with a multistable stimulus. *J Comput Neurosci*, *34*(1), 103-124. doi:10.1007/s10827-012-0409-5
- Rhodes, G., Jeffery, L., Clifford, C. W., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Res*, *47*(17), 2291-2296. doi:10.1016/j.visres.2007.05.012
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci*, *2*(11), 1019-1025. doi:10.1038/14819
- Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *J Vis*, *9*(6), 15 11-32. doi:10.1167/9.6.15
- Rubin, N. (2003). Binocular rivalry and perceptual multi-stability. *Trends Neurosci*, *26*(6), 289-291; discussion 287-289. doi:10.1016/S0166-2236(03)00128-0
- Said, C. P., & Heeger, D. J. (2013). A model of binocular rivalry and cross-orientation suppression. *PLoS Comput Biol*, *9*(3), e1002991. doi:10.1371/journal.pcbi.1002991
- Sani, L., Ricciardi, E., Gentili, C., Vanello, N., Haxby, J. V., & Pietrini, P. (2010). Effects of Visual Experience on the Human MT+ Functional Connectivity Networks: An fMRI Study of Motion Perception in Sighted and Congenitally Blind Individuals. *Front Syst Neurosci*, *4*, 159. doi:10.3389/fnsys.2010.00159
- Sato, M., Basirat, A., & Schwartz, J. L. (2007). Visual contribution to the multistable perception of speech. *Percept Psychophys*, *69*(8), 1360-1372.
- Saygin, A. P., Wilson, S. M., Hagler, D. J., Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J Neurosci*, *24*(27), 6181-6188. doi:10.1523/JNEUROSCI.0504-04.2004

- Serences, J. T. (2004). A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. *Neuroimage*, *21*(4), 1690-1700. doi:10.1016/j.neuroimage.2003.12.021
- Shapiro, A., Moreno-Bote, R., Rubin, N., & Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *J Comput Neurosci*, *27*(1), 37-54. doi:10.1007/s10827-008-0125-3
- Sirotin, Y. B., Cardoso, M., Lima, B., & Das, A. (2012). Spatial homogeneity and task-synchrony of the trial-related hemodynamic signal. *Neuroimage*, *59*(3), 2783-2797. doi:10.1016/j.neuroimage.2011.10.019
- Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci*, *15*(8), 5448-5465.
- Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proc Natl Acad Sci U S A*, *104*(1), 323-328. doi:10.1073/pnas.0609006104
- Sun, J., & Perona, P. (1997). Shading and stereo in early perception of shape and reflectance. *Perception*, *26*(4), 519-529.
- Taubert, N., Christensen, A., Endres, D., Giese, M. (2012). *Online simulation of emotional interactive behaviors with hierarchical Gaussian process dynamical models*. Paper presented at the ACM Symposium on Applied Perception.
- Thurman, S. M., van Boxtel, J. J., Monti, M. M., Chiang, J. N., & Lu, H. (2016). Neural adaptation in pSTS correlates with perceptual aftereffects to biological motion and with autistic traits. *Neuroimage*, *136*, 149-161. doi:10.1016/j.neuroimage.2016.05.015
- Troje, N. F., Sadr, J., Geyer, H., & Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *J Vis*, *6*(8), 850-857. doi:10.1167/6.8.7
- Tsutsui, K., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J Neurophysiol*, *86*(6), 2856-2867.
- Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, *298*(5592), 409-412. doi:10.1126/science.1074128
- Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci U S A*, *98*(20), 11656-11661. doi:10.1073/pnas.191374198
- van Boxtel, J. J., & Lu, H. (2013). Impaired global, and compensatory local, biological motion processing in people with high levels of autistic traits. *Front Psychol*, *4*, 209. doi:10.3389/fpsyg.2013.00209
- van Ee, R. (2005). Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Res*, *45*(1), 29-40. doi:10.1016/j.visres.2004.07.039
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009
- Vangeneugden, J. (2010). *Action representation by macaque visual temporal cortical neurons*. (PhD), KU Leuven.
- Vangeneugden, J., De Maziere, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *J Neurosci*, *31*(2), 385-401. doi:10.1523/JNEUROSCI.2703-10.2011
- Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *J Neurosci*, *34*(2), 574-585. doi:10.1523/JNEUROSCI.4032-13.2014

- Vangeneugden, J., Vancleef, K., Jaeggli, T., VanGool, L., & Vogels, R. (2009). Discrimination of locomotion direction in impoverished displays of walkers by macaque monkeys. *J Vis*, *10*(4), 22 21-19. doi:10.1167/10.4.22
- Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception*, *33*(5), 547-560.
- Vattikuti, S., Thangaraj, P., Xie, H. W., Gotts, S. J., Martin, A., & Chow, C. C. (2016). Canonical Cortical Circuit Model Explains Rivalry, Intermittent Rivalry, and Rivalry Memory. *PLoS Comput Biol*, *12*(5), e1004903. doi:10.1371/journal.pcbi.1004903
- Veltz, R., Faugeras, O. (2010). Local/Global Analysis of the Stationary Solutions of Some Neural Field Equations. *SIAM J. Appl. Dyn. Syst.*, *9*(3), 954-998. doi:10.1137/090773611
- Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci*, *5*(5), 491-499. doi:10.1038/nn839
- Wade, N. J., de Weert, C. M., & Swanston, M. T. (1984). Binocular rivalry with moving patterns. *Percept Psychophys*, *35*(2), 111-122.
- Walker, P. (1975). The subliminal perception of movement and the 'suppression' in binocular rivalry. *Br J Psychol*, *66*(3), 347-356.
- Walker, P., & Powell, D. J. (1979). The sensitivity of binocular rivalry to changes in the nondominant stimulus. *Vision Res*, *19*(3), 247-249.
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *J Exp Psychol*, *45*(4), 205-217.
- Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Curr Opin Neurobiol*, *17*(4), 423-429. doi:10.1016/j.conb.2007.07.001
- Webster, M. A. (2011). Adaptation and visual coding. *J Vis*, *11*(5). doi:10.1167/11.5.3
- Weigelt, S., Muckli, L., & Kohler, A. (2008). Functional magnetic resonance adaptation in visual neuroscience. *Rev Neurosci*, *19*(4-5), 363-380.
- Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage*, *22*(1), 277-288. doi:10.1016/j.neuroimage.2003.12.043
- Wheatstone, C. (1962). On some remarkable and hitherto unobserved phenomena of binocular vision. *Optom Wkly*, *53*, 2311-2315.
- Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proc Natl Acad Sci U S A*, *100*(24), 14499-14503. doi:10.1073/pnas.2333622100
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*, *12*(1), 1-24. doi:10.1016/S0006-3495(72)86068-5
- Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*(2), 55-80.
- Winkler, I., Denham, S., Mill, R., Bohm, T. M., & Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philos Trans R Soc Lond B Biol Sci*, *367*(1591), 1001-1012. doi:10.1098/rstb.2011.0359
- Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *J Neurosci*, *35*(20), 7727-7735. doi:10.1523/JNEUROSCI.0188-15.2015
- Xie, X., & Giese, M. A. (2002). Nonlinear dynamics of direction-selective recurrent neural media. *Phys Rev E Stat Nonlin Soft Matter Phys*, *65*(5 Pt 1), 051904. doi:10.1103/PhysRevE.65.051904
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci*, *11*(11), 1352-1360. doi:10.1038/nn.2202

- Yang, Y., & Calakos, N. (2013). Presynaptic long-term plasticity. *Front Synaptic Neurosci*, 5, 8. doi:10.3389/fnsyn.2013.00008
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, 16(6), 2112-2126.
- Zhou, J., Reynaud, A., & Hess, R. F. (2014). Real-time modulation of perceptual eye dominance in humans. *Proc Biol Sci*, 281(1795). doi:10.1098/rspb.2014.1717
- Zhou, Y. H., Gao, J. B., White, K. D., Merk, I., & Yao, K. (2004). Perceptual dominance time distributions in multistable visual perception. *Biol Cybern*, 90(4), 256-263. doi:10.1007/s00422-004-0472-8

Declaration of contribution

This thesis comprises three manuscripts two of which are published and one is submitted. The author of the manuscripts is the first author of all three manuscripts. The following list gives an overview of the contributions of the author of this thesis, and states the contributions of each manuscript co-author respectively.

1. L Fedorov, D-S Chang, M Giese, H Bülthoff, S de la Rosa (submitted). **Adaptation aftereffects suggest neural representations for the encoding of social interactions.** Full text provided in Chapter 1. LF designed, implemented and performed all experiments, analyzed data and wrote the submitted manuscript. SR conceived initial study and designed all experiments, supported data analysis and wrote the submitted manuscript. DSC conceived initial study, collected motion capture data and prepared initial manuscript draft. MG supported design of Experiments 2 and 3 and writing of submitted manuscript.
2. L Fedorov, T Dijkstra, M Giese (2018). **Lighting-from-above-prior in biological motion perception.** *Scientific Reports. 8: 1507.* Full text provided in Chapter 2. LF performed the experiments, analyzed the data, implemented the model and wrote the paper. TD analyzed the data and wrote the manuscript. MG designed the research, supported the data analysis and wrote the paper.
3. L Fedorov, J Vangeneugden, M Giese (2016). **Neural model for the influence of shading on the multistability of the perception of body motion.** *NCTA Neural Computation Theory and Applications. Conf. Proceedings: 1-8.* The paper received Best Student Paper award. Full text provided in Chapter 3. LF implemented the model and simulations and wrote the paper. MG implemented part of model and wrote the paper. JV collected and provided behavioral dataset.

Parts of this work were also presented at the following conferences.

1. L Fedorov, J Vangeneugden, M Giese. Neurodynamical model for visual action recognition. **Bernstein Conference for Computational Neuroscience 2014.**(poster)
2. L Fedorov, J Vangeneugden, M Giese. Perception of biological motion depends on lighting-from-above prior. **ECVP 2014.**(poster)
3. L Fedorov, D Endres, J Vangeneugden, M Giese. Neurodynamical model for the multi-stable perception of biological motion. **VSS 2014.**(poster)
4. L Fedorov, M Giese. Shading cues in the perception of biological motion: a neural model and a new illusion. **ECVP 2015** (talk)
5. L Fedorov, M Giese. Neural model of biological motion recognition based on shading cues. **Organization of Computational Neuroscience meeting 2015** (poster)
6. L Fedorov, M Giese. Lighting-from-above prior in the perception of biological motion: new illusion and a neural model. **VSS 2015** (poster)

7. L Fedorov, M Giese. Model for the integration of form and shading cues in multi-stable body motion perception. **ECVP 2016** (*talk*)
8. D-S Chang, L Fedorov, M Giese, H Bülthoff, S de la Rosa. Adaptation aftereffects in perception of complementary actions. **ECVP 2016** (*talk*)
9. L Fedorov, M Giese. An integrated model for the shading and silhouette cues in the perception of biological motion. **VSS 2016** (*poster*)
10. L Sting, L Fedorov, T Dijkstra, H Hock, M Giese. Dynamics of multistable biological motion perception. **VSS 2017** (*poster*)
11. L Fedorov, J Vangeneugden, M Giese. Neural model for the influence of shading on the multistability of the perception of body motion. **IJCCI/NCTA 2016** (*talk*)
12. L Fedorov, M Giese. Silhouette and shading cues control multi-stability in action recognition. **Organization for Computational Neuroscience meeting 2016** (*poster*)
13. L Fedorov, T Dijkstra, L Sting, H Hock, M Giese. Modeling of the perceptual dynamics of the perception of body motion. **Organization for Computational Neuroscience meeting 2017** (*poster*)

Chapter 1. Adaptation Aftereffects suggest neural representations for the encoding of social interactions.

(submitted)

Title: "Adaptation aftereffects reveal neural representations for encoding of contingent social actions"

Short title: "Representation of contingent social actions"

*Authors: Leonid A Fedorov^{*1,3}, Dong-Seon Chang^{2,3}, Martin A Giese^{1,3}, Heinrich Bülthoff^{2,4} and Stephan de la Rosa^{*2}*

Classification: Social Sciences & Biological Sciences (Major category)

Psychological and Cognitive Sciences (Minor category)

Author Affiliation:

¹ Section for Computational Sensomotorics, Dept. Cognitive Neurology, CIN & HIH, UKT, University of Tübingen, 25 Otfried-Müller Strasse, 72076 Tübingen, Germany

² Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38-44, 72076 Tübingen, Germany

³ International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Österbergstrasse 3, 72076 Tübingen, Germany

⁴Department of Brain and Cognitive Engineering, Korea University, 136-713 Seoul, Korea

^{*} Equal contribution to the study

Corresponding authors: Heinrich Bülthoff & Stephan de la Rosa; Social and Spatial Cognition Group, Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Spemannstrasse 36, 72076 Tübingen, Germany; tel: +49 7071 601 606; email: hnb@tuebingen.mpg.de & delarosa@tuebingen.mpg.de

Keywords: perception, vision, social interaction, action recognition, adaptation, aftereffects, neural representation

Abstract

A hallmark of human social behavior is the effortless ability to relate one's own actions to that of the interaction partner, e.g. when stretching out one's arms to catch a tripping child. What are the neural substrates that support this indispensable human skill? Here we examined the neural substrates of a candidate process underlying the ability to relate actions to each other, namely the recognition of temporal-spatial contingencies between actions (e.g. a 'giving' that is followed by 'taking'). We used a behavioral adaptation paradigm to examine the response properties of neural mechanisms at a behavioral level. In contrast to the common view that action sensitive neural processes are primarily selective for one action (i.e. primary action, e.g. 'throwing'), we demonstrate that these processes also exhibit sensitivity to the matching contingent action (e.g. 'catching'). Control experiments demonstrate that the sensitivity of action recognition processes to contingent actions cannot be explained by lower-level visual features. Moreover, we show that action recognition processes are only sensitive to contingent actions but not to non-contingent actions demonstrating their *selective* sensitivity to contingent actions. Our findings, therefore, provide first evidence for the selective neural coding of action contingencies by action sensitive processes and demonstrate how the link between individual actions in social interactions can be established at a neural level.

Significance Statement

Why is it so easy for humans to interact with each other? In social interactions humans coordinate their action with each other non-verbally. For example, dance partners need to relate their actions to each other in order to coordinate their movements. The underlying neurocognitive mechanisms supporting this ability are surprisingly poorly understood. We show for the first time that human brain possesses neural mechanisms that are sensitive to pairs of matching actions that make up a social interaction. These findings provide novel insights into the neural architecture that helps humans to relate actions with each. This capability is essential for social interaction and its understanding will aide future development of therapies to treat social cognitive disorders.

Results

Adaptation effects have been demonstrated to be a powerful tool in the examination of response properties of neural processes at the behavioral level. Specifically, it is believed that behavioral adaptation is able to selectively target the neural mechanisms and tuning characteristics of perceptual processes across the cortical hierarchy (1-4). This is so because the repetitive sensory stimulation during adaptation results in transient response decrease in the neuronal populations involved in the processing of the stimulus (5-12). As a result, the perception of a subsequently presented ambiguous test stimulus is altered for a short period. We have previously shown that adaptation to an action (e.g. 'throwing') transiently changes the visual percept of a subsequently presented ambiguous action that contains visual elements of two actions (e.g. a morph between a 'throwing' and a 'giving' - see Supplementary Movie 1; or a Supplementary Movie 2 for 'catching-taking' pair)(13, 14). For example, adapting to a 'throwing' action transiently causes participants' to report an ambiguously looking test action (e.g. a 'throwing-giving' morph) to look more like 'giving' and vice versa.

In the current study, we used this paradigm to examine the neural basis of the ability to relate one action, e.g. 'throwing', to a naturally occurring spatio-temporal proximate actions (contingent actions), e.g. 'catching'. Specifically, we were interested in whether action recognition mechanisms are sensitive to both their primary action and to the related contingent action. To this end, we used two actions, namely 'throwing' and 'giving', to examine the sensitivity of action sensitive processes to contingent actions - 'catching' and 'taking'. Note that in a normal version of this adaptation paradigm, adapting to a 'throwing' action causes participants to perceive an ambiguous action morph between 'throwing' and 'giving' more likely as 'giving'. We reasoned that if 'throwing' action sensitive processes are also sensitive to the contingent action (i.e. 'catching'), then the 'catching' adaptor should also induce an adaptation aftereffect in action recognition processes sensitive to a 'throwing' action. Consequently, adaptation to 'catching' should lead participants to perceive a 'throwing-giving' test morph action more likely as 'giving'. Hence, one would expect that a 'catching' adaptor causes a similar, though somewhat smaller, adaptation effect than a 'throwing' adaptor, if action recognition processes are also sensitive to the contingent action. We refer to these adaptation effects induced by the contingent action hereafter as cross-adaptation effects. The presence of significant cross-adaptation effects provides direct evidence for action recognition processes being sensitive to their contingent action.

We used two social interactions, namely ‘throwing’ and ‘catching’ and ‘giving’ and ‘taking’ (see Fig 2A and Supplementary Movies 1 and 2) to create ambiguous test stimuli. Specifically, we morphed (see Methods) between ‘throwing’ and ‘giving’ (i.e. along the ‘throwing-giving’ morph axis), and between ‘catching’ and ‘taking’ (i.e. along the ‘catching-taking’ morph axis). We probed the adaptation effects for *normal* adaptation conditions and *cross-adaptation* conditions. In the normal adaptation conditions, the test and adaptor stimuli were taken from the same morph axis, e.g. using ‘giving’ as the adaptor stimuli and ‘throwing-giving’ morph as the test stimuli. In the cross-adaptation conditions, the adaptor and test were taken from action-contingent morph axes, e.g. using ‘taking’ as the adaptor and ‘throwing-giving’ as the test. We hypothesize, that behavioral adaptation effects in the cross condition indicate the sensitivity of action recognition processes to contingent actions.

In line with this hypothesis, our results show a significant cross adaptation effect as indicated by an one-sample t-test ($t(24) = 5.83$, Cohen’s $d = 1.17$, $p < 0.001$). Despite their sensitivity to contingent actions, action recognition processes are more sensitive to their primary action as suggested by the significantly larger adaptation effect in the normal than in the cross adaptation condition (paired t-test: $t(24) = 6.26$, Cohen’s $d = 1.48$, $p < 0.001$) (see Fig 2B). Overall, these results provide strong evidence for action-sensitive neural units being partially activated by contingent actions that naturally occur in the spatiotemporal proximity of the primary action.

The cross-adaptation effects are particularly surprising as the adaptor and test stimuli were perceptually very dissimilar. This makes the contribution of low-level adaptation effects to the cross-adaptation effect unlikely. Nevertheless, it is still possible that low-level adaptation effects, e.g. adaptation to joint angle changes, might have caused the cross-adaptation effect. In a control experiment we assessed the contribution of these lower-level adaptation effects in the cross-adaptation condition. Specifically, we retained local joint angle movements while impairing the holistic percept of the action by remapping the joint angle movements of the arms onto the legs and vice versa. This manipulation left the overall available movement information of the stimulus largely unaffected (see Fig 3A and Supplementary Movie 3). Yet, participants were unable to identify these scrambled actions as meaningful ones, let alone as the original actions. We reasoned that if low-level visual cues were the sole contributors to the cross-adaptation effect, then adaptation to scrambled actions should

lead to very similar cross-adaptation effects as adaptation to non-scrambled actions. Participants were adapted with non-scrambled and scrambled actions as adaptors in both the cross-adaptation and the normal condition and always probed with non-scrambled morphed test stimuli. The most important finding of this experiment was that in the cross-adaptation condition the scrambled adaptors, which contained very similar movement cues to the non-scrambled adaptors, significantly reduced the size of the adaptation effect compared to the non-scrambled adaptors (paired t-test $t(24) = 3.56$, Cohen's $d = 1.17$, $p = 0.001$). Specifically, the scrambled adaptors in the cross-adaptation condition did not induce significant adaptation effects ($t(24) = 1.6$, Cohen's $d = 0.32$, $p = 0.12$) showing that action recognition processes are little or not sensitive at all to scrambled contingent actions (Fig 3B). In addition, we replicated the cross-adaptation effect of the previous experiment with non-scrambled adaptors (Fig 3B, $t(24)=4.35$, Cohen's $d = 0.87$, $p < 0.001$). This control experiment demonstrates that cross-adaptation effects cannot be explained by low-level visual cues alone.

Experiments 1 and 2 demonstrate the sensitivity of recognition processes to contingent actions. However, they do not show the *selective* sensitivity of action recognition processes to contingent actions (i.e. contingency specificity). For example, it is possible that action recognition processes are sensitive to both contingent and non-contingent actions. In a third experiment we examined the specificity of the cross-adaptation effect to contingent actions. We used two novel actions as adaptors that were non-contingent to the morphed actions. Namely, we used the lead and the follow action from a salsa-dancing couple as non-contingent adaptor actions (Fig 4A and Supplementary Movie 4) to probe cross-adaptation effects with the action morphs of Experiment 1 and 2. Because these adaptor actions were non-contingent to the actions of the test stimulus, the direction of their perceptual effect on the test stimuli is not known a-priori. Hence, we calculated the difference between these two non-contingent conditions post-hoc in such a way that their difference resembled an action adaptation effect (i.e. their difference was positive rather than negative). Note, that by doing so, we choose the most conservative approach that is least likely to produce a significant reduction of the cross-adaptation effect in the non-contingent compared to the contingent condition. In addition, we also replicated the cross-adaptation with the contingent actions as adaptors from Experiment 1. These non-

contingent adaptors significantly reduced the cross-adaptation effect compared to the contingent adaptors ($t(23) = 2.41$, Cohen's $d = 0.77$, $p = 0.02$) demonstrating that action recognition processes are sensitive to action contingency. The contingent condition provided evidence for a cross-adaptation effect, $t(23) = 5.26$, Cohen's $d = 1.07$, $p < 0.001$. In summary, Experiment 3 demonstrates the specificity of cross adaptation effects to the contingent action.

Discussion

Here we provide first evidence of neural coding for action contingencies. In three experiments, we show a significant cross-adaptation aftereffect, demonstrating that mechanisms of action perception are sensitive to both the primary action and its contingent action. Importantly, the sensitivity to the contingent action was observed despite large visual dissimilarities between the adaptor and test stimuli. This renders the explanation unlikely that the sensitivity of action recognition mechanisms to the contingent action is mediated by low-level visual cues such as joint angles or local motion cues. A more explicit test of this explanation involved the distortion of the action semantics while retaining many of the low-level motion cues by swapping the arm and leg movements in Experiment 2. In this experiment only a weak cross adaptation effect was found, while the low-level motion cues were very similar to Experiment 1, suggesting that a holistic interpretation of the body motion stimuli was essential for the e cross-adaptation effects. Taken together these results strongly suggest that cross-adaptation effects rely very little or not at all on overall available low-level visual information. It rather seems that features that are processed later in the action recognition hierarchy mediate cross-adaptation effects. Importantly, Experiment 3 shows that these features are specific to the probed social interaction as the cross-adaptation effect is significantly decreased when non-contingent actions are used for adaptation. This rules out generic explanations for the cross-adaptation effect that are not social interaction-specific. In summary, we show first evidence for the selective encoding of action contingencies of naturally observed interactions by neural action representations.

The ability to recognize action contingencies is essential for several important social cognitive functions. For example, contingent actions are better detected in noise than non-contingent actions during social interaction recognition (15-18). The ability to take advantage of action contingencies in these detection tasks seem to be related to the overall social functioning. High functioning autistic

persons have been shown to take less advantage of action contingencies in order to improve the detection of social interactions in noise compared to healthy controls (19). This deficit in people with autistic traits cannot be attributed to a general lack of recognizing contingencies. Children with autism focus on non-social contingencies when observing biological motion patterns (20) suggesting that people with autistic traits are able to detect physical contingencies. The selective impairment of using social contingencies in autistic people indicates that action contingencies are an integral part for normal social functioning.

What possible neural architecture gives rise to action recognition mechanisms being sensitive to contingent actions? Action recognition has been associated with activation in superior temporal sulcus (STS) (21-23). Converging evidence, imaging(24) and physiological(7) studies show that visual adaptation with the action stimuli transiently changes the neural response properties in area STS. In line with this, behavioral adaptation aftereffect and the neurophysiological adaptation have been linked(25, 26). We believe that the prolonged presentation of an action transiently inhibits the neural population sensitive to the adapted action that causes the adaptation effect. When an ambiguous test action is shown that contains features of both the adapted and non-adapted action, the non-adapted population exhibit a relatively stronger response compared to the non-adapted population. As a result, the ambiguous test stimulus is reported more often as the non-adapted stimulus. Within this framework the sensitivity of action recognition mechanisms to action contingencies might arise from associative learning in the following way. In social interactions, individual actions frequently co-occur in close temporal proximity resulting in observable statistical regularities between the actions (27). These statistical regularities could result in correlated activation between the underlying neural action recognition populations. Associative learning causes this correlated activity to manifest over time by selectively strengthening the connections between these neural populations (28, 29). It could be these strengthened connections that cause the activation of both the 'catching' and the neural 'throwing' action recognition processes during the observation of a 'catching' action. Accordingly, the repeated observation of a 'catching' action might therefore induce an adaptation effect in the 'throwing' neural

population, hence, giving rise to the cross-adaptation effect. However further experimentation is needed to determine the precise neural architecture of the effect.

We observed that cross adaptation effects are independent on the natural temporal order in which two actions occur. Specifically, cross-adaptation effects did not statistically differ when the ‘throwing’ adaptor preceded a ‘catching-taking’ action morph (correct temporal order) and when the ‘catching’ adaptor preceded a ‘throwing-giving’ action morph (reversed temporal order). Hence action recognition mechanisms provide information about the upcoming action from the preceding contingent action, when action are observed in the correct temporal order. This observation is in line with previous notions of predictive coding in the perception of action contingencies (16, 17, 30, 31). On the other hand, action mechanisms also provide information about the contingent action that preceded the seen action. In this sense action recognition mechanisms are postdictive. Postdiction is useful to recover the cause of an action if the preceding action has been missed, for example, due to occlusion of the actor or the observer looking away. Our results therefore suggest a mechanism that allows a social interaction to be recognized even if only the final action of the social interaction is seen.

Finally, we would like to point out the robustness of the cross-adaptation effect. We were able to replicate this effect in three independent experiments with each effect associated with a large effect size of at least Cohen’s $d = 0.87$.

In conclusion, we provide first evidence for the selective neural encoding of contingencies occurring between natural social interactions. The encoding of contingencies cannot be explained by the overall available low-level visual information. Moreover, the encoding of action contingencies is specific to the interaction. We suggest that such a mechanism is critical for humans to ‘see’ relationships between otherwise independent social actions.

Methods

Stimuli and Apparatus.

We recorded the action stimuli from real life dyadic interactions using two motion capture suits each equipped with 17 inertial motion trackers distributed over the whole body (MOVEN Motion Capture Suit from XSense, Netherlands). Two actors were standing facing each other and carried out the actions starting from a neutral pose. We recorded three different sets of interactions between the actors: Interaction 1 ('giving-taking') consisted of one person giving a small bag to another person who was taking it, Interaction 2 ('throwing-catching') consisted of one person throwing a small bag to another person who was catching it, and Interaction 3 ('salsa dancing') consisted of two people dancing salsa together as partners. From these three interactions, six actions were recorded: 'giving', 'taking', 'throwing', 'catching', 'leading salsa', 'following salsa'. The recorded action stimuli contained the information about the change of three-dimensional (3D) spatial coordinates of 22 body joints over time. All actions were processed into short movies of standardized lengths lasting 1.2 seconds. We also generated ambiguous action morphs between actions originating from Interaction 1 or Interaction 2, whereas actions were either morphed between "initiating" actions ('giving-throwing' morphs) or "responding" actions ('catching-taking' morphs). The morphs were calculated by the weighted average of local joint angles between two actions (we used the same procedure as in (13, 14)). The morph levels chosen as test stimuli were determined individually in practice trials (see Procedure). All stimuli were presented with an augmented reality setup where participants could see the actions in 3D and carried out by a real life size avatar (height = 1.73 m) rendered as a human female figure in Experiment 1, and as a stick figure for Experiment 2 and Experiment 3. All actions were presented with the avatar facing the viewer at a fixed distance of 2.3 meters from the screen to the motion-tracked glasses. In Experiment 1 with the human-like avatar, the avatar kept a neutral facial expression. The setup was programmed and controlled with the Unity game engine (San Francisco, USA), and the animated avatar stimuli were acquired from Rocketbox (Hannover, Germany). The stimuli were projected using back-projection technique enabled by a Christie Mirage S+3K stereo projector (Kitchener, Canada) with a refresh rate of 115 Hz, and all participants wore Nvidia 3D Vision Pro shutter glasses (Santa Clara, USA) synchronized to the display in order to perceive the stimuli in 3D. An ART Smarttrack system (Weilheim, Germany) was used to track the position of the head as well as the position of the hands of

participants, in order to update the 3D visual scene in response to the viewpoint changes and enable action execution and task responses of the participants by using their hands.

Procedure.

Practice Trials: At the beginning of all experiments, participants put on a motion-tracked 3D-goggles and the hand-tracker. Then they were shown how to answer which actions they perceived in the experiment by moving their tracked hands and touching one of the two virtual 3D-buttons appearing mid-air labelled with the respective names of the actions ('giving', 'throwing', 'taking', 'catching'). Participants learned to categorize ambiguous actions to an action either belonging to Interaction 1 ('giving', 'taking') or Interaction 2 ('throwing', 'catching'), depending on the morphs of ambiguous actions which were either "initiating" or "responding" actions from Interaction 1 and Interaction 2 ('giving-throwing' or 'catching-taking' action morphs). They were able to practice this repeatedly while different morph weights between the actions were presented in ascending and descending manners in order to determine the point of overall ambiguous perception (PSE) for each participant. All morph levels were presented twice.

Baseline Condition: After the morph weight was determined in the practice trials, participants were presented three repetitions of each test morph stimuli in the absence of adaptor stimuli (baseline). The participants who showed inconsistent responses in these three iterations of the same action morphs were considered as ineligible for the experiment and excluded prior to participation in main experimental phase. We determined the baseline perception for each action, and for each participant. Once determined, the morph weights in the test stimuli were kept identical for each participant across all experimental blocks during the main experimental phase. The total number of trials in the baseline condition consisted of 42 trials (2 action morphs \times 7 morph levels \times 3 repetitions) for each participant. Same baseline and practice phases were used in all 3 experiments as a preliminary step to familiarize participants with the setup.

(Main) Adaptation Experiment 1: The main adaptation experiment consisted of eight experimental blocks, a fully crossed design with the factors adaptor stimuli (4 \times) and test morphs (2 \times). The order of experimental blocks was completely balanced across all participants. The adaptor stimuli consisted of the four recorded actions ('giving', 'throwing', 'taking', and 'catching') each presented in separate

experimental blocks. The test stimuli were morphed actions either between “initiating” actions or between “responding” actions from Interaction 1 and Interaction 2 (‘giving–throwing’ or ‘catching–taking’ action morphs). The test stimuli consisted of a set of seven different morph weights with equal morph distances between each other and each morphed action was shown three times in each adaptor condition resulting in twenty-one test stimuli presentations in each experimental block. The order of test stimuli presentation within each experimental block was completely randomized. Each experimental block of trials consisted of an initial adaptation phase where the same adaptor stimulus was repeatedly presented 30 times with an inter-stimulus interval (ISI) of 250 ms. The main action categorization phase directly followed, where the adaptor stimulus was repeatedly presented 4 times (same ISI of 250 ms) before each trial, and then the test stimuli appeared for a 2-Alternatives-Forced-Choice (2AFC) task where participants had to judge which action they perceived (e.g. Did you see ‘giving’ or ‘throwing’). The ISI between the adaptor stimuli and test stimuli were 400 ms. The next trial started as soon as the participants recorded their answer by moving their tracked hands towards the virtual button. The total number of trials for the main adaptation Experiment 1 was 168 trials (8 experimental blocks × 7 morph levels × 3 repetitions of test stimuli) for each participant. Participants took about 60 minutes to finish the whole experiment, taking approximately 5 minutes per experimental block.

Adaptation Experiment 2: The second adaptation experiment consisted of sixteen experimental blocks, a fully crossed design with the factors adaptor action (4×), test morphs (2×) and whether the adaptor action was scrambled or not (2×). The order of experimental blocks was completely balanced across all participants. The adaptor stimuli consisted of the four recorded actions (‘giving’, ‘throwing’, ‘taking’, and ‘catching’) each presented in separate experimental block, and four same actions in which the joint angles were permuted between the arms and the legs (scrambled). The test stimuli were, same as in Experiment 1, the morphed actions either between “initiating” actions or between “responding” actions from Interaction 1 and Interaction 2 (‘giving–throwing’ or ‘catching–taking’ action morphs). The total number of trials for the main adaptation Experiment 2 was 288 trials (16 experimental blocks × 6 morph levels × 3 repetitions of test stimuli) for each participant. Participants took about 115 minutes to finish the whole experiment, taking approximately 5 minutes per experimental block.

Adaptation Experiment 3: The second adaptation experiment consisted of twelve experimental blocks, a fully crossed design with the factors adaptor action (6×), test morphs (2×) and whether the adaptor action was scrambled or not (2×). The order of experimental blocks was completely balanced across all participants. The adaptor stimuli consisted of the six recorded actions ('giving', 'throwing', 'taking', 'catching', 'leading salsa', 'following salsa') each presented in separate experimental block. The test stimuli were, same as in Experiments 1 and 2, the morphed actions either between "initiating" actions or between "responding" actions from Interaction 1 and Interaction 2 ('giving-throwing' or 'catching-taking' action morphs). The total number of trials for the main adaptation Experiment 3 was 216 trials (12 experimental blocks × 6 morph levels × 3 repetitions of test stimuli) for each participant. Participants took about 80 minutes to finish the whole experiment, taking approximately 5 minutes per experimental block.

Participants

Twenty-five (n=25) volunteers participated in Experiment 1, a distinct group of twenty-five (n=25) volunteers participated in Experiment 2, and a distinct group of twenty-four (n=24) volunteers participated in Experiment 3. All participants were compensated by 8 EURO per hour. After the experiments, they were debriefed and informed about the study. Exclusion criteria applied to participants who could not perceive the action stimuli with the 3D-goggles, who showed inconsistent responses after three iterations in the process determining the point of overall ambiguous perception for the two morphed actions (i.e. point of subjective equality; PSE), or who reported to have had extensive trainings in sports involving throwing or catching actions (e.g. experts in baseball, basketball, juggling).

Consent.

Psychophysical experiments were performed with informed consent of participants. All participants were informed about the purpose of the experiment before signing an informed consent. All participants were naïve concerning the hypotheses of the experiment. The experiment was conducted in line with the Declaration of Helsinki and in accordance with the recommendations of the ethics board of the University of Tübingen (Germany).

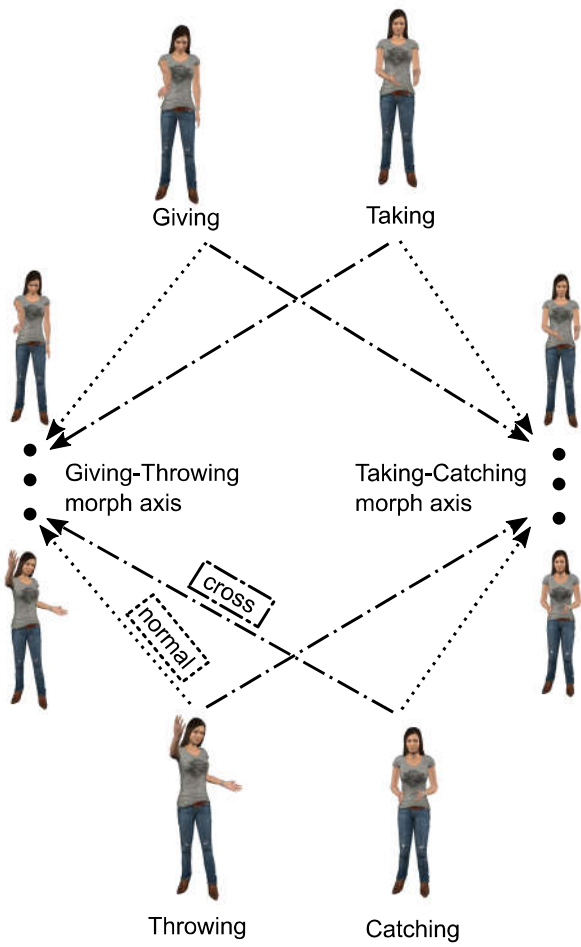
References

1. de la Rosa S, Ekramnia M, & Bulthoff HH (2016) Action Recognition and Movement Direction Discrimination Tasks Are Associated with Different Adaptation Patterns. *Front Hum Neurosci* 10:56.
2. Clifford CW, et al. (2007) Visual adaptation: neural, psychological and computational aspects. *Vision Res* 47(25):3125-3131.
3. Lawson RP, Clifford CW, & Calder AJ (2009) About turn: the visual representation of human body orientation revealed by adaptation. *Psychol Sci* 20(3):363-371.
4. Bekesy Gv (1929) Theorie des Hörens: über die Bestimmung des einem reinen Tonempfinden entsprechenden Erregungsgebietes der Basilmembran vermitteltst Ermüdungserscheinungen. . *Physikalische Zeitschrift* 30:115–125.
5. Grill-Spector K & Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107(1-3):293-321.
6. Webster MA (2011) Adaptation and visual coding. *J Vis* 11(5).
7. Kuravi P, Caggiano V, Giese M, & Vogels R (2016) Repetition suppression for visual actions in the macaque superior temporal sulcus. *J Neurophysiol* 115(3):1324-1337.
8. Kuravi P & Vogels R (2017) Effect of adapter duration on repetition suppression in inferior temporal cortex. *Sci Rep* 7(1):3162.
9. Vogels R (2016) Sources of adaptation of inferior temporal cortical responses. *Cortex* 80:185-195.
10. Kaliukhovich DA, De Baene W, & Vogels R (2013) Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *J Cogn Neurosci* 25(5):777-789.
11. Kaliukhovich DA & Vogels R (2012) Stimulus repetition affects both strength and synchrony of macaque inferior temporal cortical activity. *J Neurophysiol* 107(12):3509-3527.
12. Kaliukhovich DA & Vogels R (2011) Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb Cortex* 21(7):1547-1558.
13. de la Rosa S, Ferstl Y, & Bulthoff HH (2016) Visual adaptation dominates bimodal visual-motor action adaptation. *Sci Rep* 6:23829.
14. Ferstl Y, Bulthoff H, & de la Rosa S (2017) Action recognition is sensitive to the identity of the actor. *Cognition* 166:201-206.
15. Manera V, et al. (2012) Are you approaching me? Motor execution influences perceived action orientation. *PLoS One* 7(5):e37514.
16. Manera V, Schouten B, Verfaillie K, & Becchio C (2013) Time will show: real time predictions during interpersonal action perception. *PLoS One* 8(1):e54949.
17. Manera V, Becchio C, Schouten B, Bara BG, & Verfaillie K (2011) Communicative interactions improve visual detection of biological motion. *PLoS One* 6(1):e14594.
18. Neri P, Luu JY, & Levi DM (2006) Meaningful interactions can enhance visual discrimination of human agents. *Nat Neurosci* 9(9):1186-1192.
19. von der Luhe T, et al. (2016) Interpersonal predictive coding, not action perception, is impaired in autism. *Philos Trans R Soc Lond B Biol Sci* 371(1693).
20. Klin A, Lin DJ, Gorrindo P, Ramsay G, & Jones W (2009) Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459(7244):257-261.
21. Vaina LM, Solomon J, Chowdhury S, Sinha P, & Belliveau JW (2001) Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci U S A* 98(20):11656-11661.
22. Grossman ED, Jardine NL, & Pyles JA (2010) fMR-Adaptation Reveals Invariant Coding of Biological Motion on the Human STS. *Front Hum Neurosci* 4:15.
23. Pelphrey KA, Morris JP, Michelich CR, Allison T, & McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. *Cereb Cortex* 15(12):1866-1876.
24. Jastorff J, Kourtzi Z, & Giese MA (2009) Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci* 29(44):14026-14038.

25. Thurman SM, van Boxtel JJ, Monti MM, Chiang JN, & Lu H (2016) Neural adaptation in pSTS correlates with perceptual aftereffects to biological motion and with autistic traits. *Neuroimage* 136:149-161.
26. Barraclough NE, Keith RH, Xiao D, Oram MW, & Perrett DI (2009) Visual adaptation to goal-directed hand actions. *J Cogn Neurosci* 21(9):1806-1820.
27. Heyes C (2010) Where do mirror neurons come from? *Neurosci Biobehav Rev* 34(4):575-583.
28. Cooper RP, Cook R, Dickinson A, & Heyes CM (2013) Associative (not Hebbian) learning and the mirror neuron system. *Neurosci Lett* 540:28-36.
29. Keyzers C & Perrett DI (2004) Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci* 8(11):501-507.
30. Neri P (2009) Wholes and subparts in visual processing of human agency. *Proc Biol Sci* 276(1658):861-869.
31. Manera V, Del Giudice M, Bara BG, Verfaillie K, & Becchio C (2011) The second-agent effect: communicative gestures increase the likelihood of perceiving a second agent. *PLoS One* 6(7):e22650.

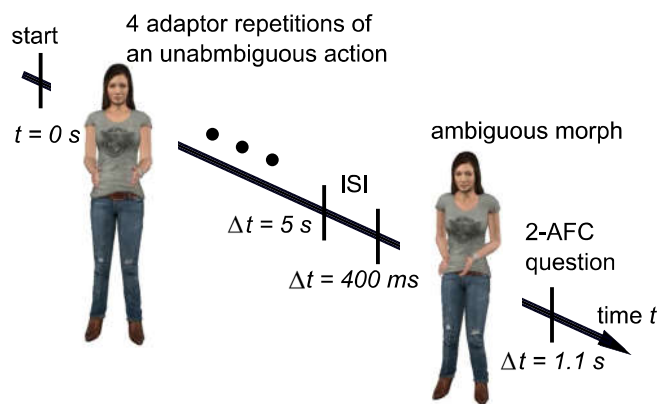
A

Design of Experiment 1



B

Adaptation Experiment Scheme



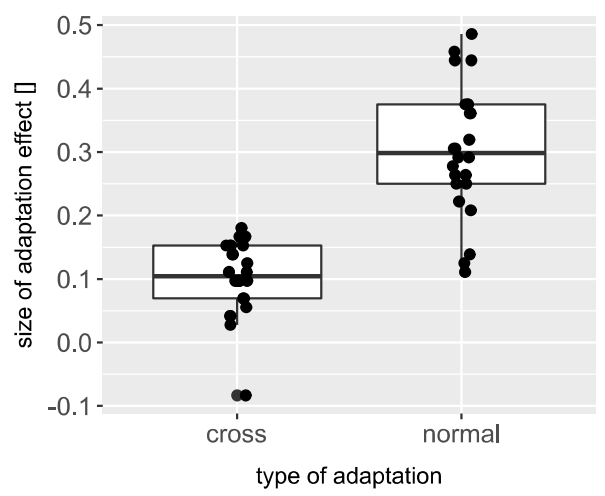
A

THROWING-GIVING

CATCHING-TAKING

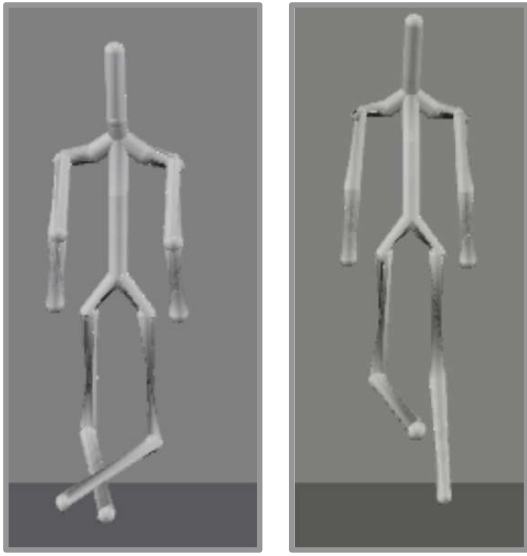


B

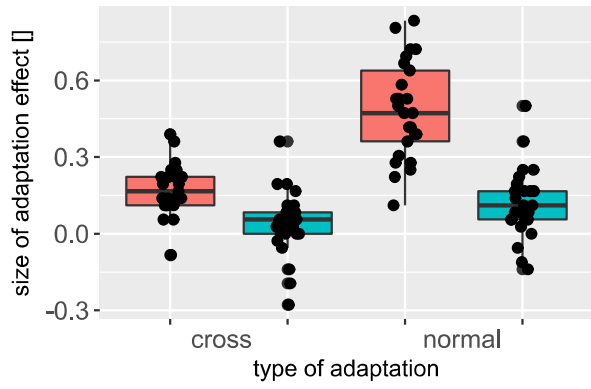


A

SCRAMBLED TAKING SCRAMBLED GIVING

**B**

■ non-scrambled ■ scrambled



A

SALSA FOLLOWER



SALSA LEADER

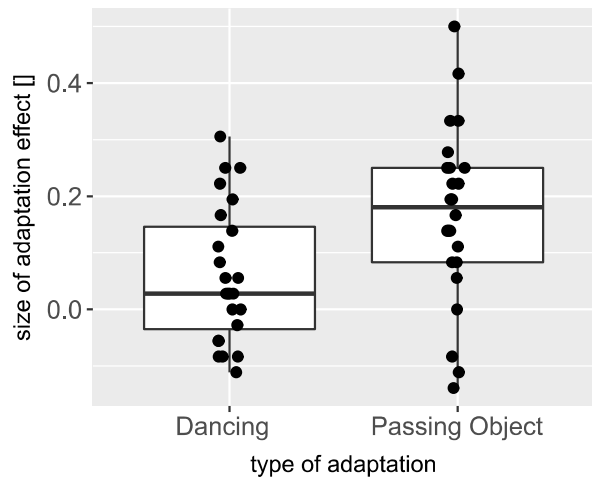
**B**

Figure Captions.

Fig. 1. Design of main cross-adaptation experiment (Experiment 1) and general experimental scheme for eliciting adaptation aftereffect. (A) Design of Experiment 1. One of the four clearly distinguishable (unmorphed) actions are chosen as an adaptor, and one of the two morphs actions are chosen as a test. The morphs are created between Throwing and Giving – as one morph axis, and between Catching and Taking – as another morph axis. When the adaptor action is chosen from the same morph axis as the test- we call it a normal condition. When the adaptor action is chosen from the different morph axis than the test – we call it a cross condition. (B) Scheme of a single adaptation trial. The adaptor is presented repetitively 4 times within a duration of approximately 5 seconds. Then an inter-stimulus interval (ISI) is introduced when nothing is displayed to a participant. Finally, the test is presented – once, before the 2-AFC question is given to a participant. A participant must choose to which action the test was more similar (the choice is always between 2 actions corresponding to the test's morph axis definition). The next trial follows the participant's answer immediately. The diagram shows a single trial in one of condition blocks. At the start of the very start of the block (before the first trial) the number of adaptor presentations is, instead much bigger – in our case 26 (see Methods for more details).

Fig. 2. Stimuli snapshots and results for Experiment 1. (A) Snapshots of test stimuli from Experiment 1. The left snapshot is one frame from the morph between Throwing and Giving, and the right snapshot – between Catching and Taking. (B) Results of Experiment 1. Boxplots of the sizes of adaptation aftereffect for cross and normal adaptation conditions. The size of adaptation aftereffect is defined as the difference between probabilities to perceive from test action the same action as the as the preceding adaptor was. Boxes indicate the ranges of the data (middle 50% interquartile range (IQR)). Black thick lines within the boxes indicate the medians. Whiskers mark intervals of 1.5 times the IQR ranges, and the dots indicate outliers that do not fall within these intervals.

Fig. 3. Stimuli snapshots and results for Experiment 2. (A) Snapshots of adaptor stimuli from Experiment 2. The left snapshot is one frame from the scrambled Taking adaptor, and the right snapshot – from the scrambled Giving adaptor. (B) Results of Experiment 2. Boxplots of the sizes of adaptation aftereffect for cross and normal, as well as scrambled cross and normal adaptation conditions.

Fig. 4. Stimuli snapshots and results for Experiment 3. (A) Snapshots of adaptor stimuli from Experiment 3. The left snapshot is one frame from the Salsa follower adaptor, and the right snapshot – from the Salsa leader adaptor. (B) Results of Experiment 2. Boxplots of the sizes of adaptation aftereffect for dancing and Passing Object conditions. In both the Dancing and Passing Object conditions the morph axis were, as in all other experiments, Throwing-Giving and Catching-Taking morphs. But the adaptor in the Dancing condition was the Salsa dance – which is semantically mismatching the test. The data in Passing Object condition is pooled across all four cross-adaptation conditions (see Text and Methods for more details).

Supplemental Movie Captions.

Supplemental Movie 1. Demonstration of a sequence of action avatars, starting from an unmorphed 'catching' action, and with each consecutive one being morphed with a 'taking' action to a larger degree than a previous one, and ending with the unmorphed 'taking' action. Unmorphed 'catching' and 'taking' avatars were used as two of the four adaptor stimuli in Experiment 1. One of the morphed avatars (the one determined to be equally likely perceived as 'catching' or 'taking' for each participant individually) was used as a test stimulus.

Supplemental Movie 2. Demonstration of a sequence of action avatars, starting from an unmorphed 'throwing' action, and with each consecutive one being morphed with a 'giving' action to a larger degree than a previous one, and ending with the unmorphed 'giving' action. Unmorphed 'throwing' and 'giving' avatars were used as two of the four adaptor stimuli in Experiment 1. One of the morphed avatars (the one determined to be equally likely perceived as 'throwing' or 'giving' for each participant individually) was used as a test stimulus.

Supplemental Movie 3. Demonstration of (stick-figure) avatars with scrambled joint positions that were used as control adaptor stimuli in Experiment 2. The movie shows three repetitions of each of the four scrambled avatars ('throwing', 'giving', 'taking', 'catching').


Supplemental Movie 4. Demonstration of (stick-figure) avatars from 'salsa dancing' social interaction that were used as control adaptor stimuli in Experiment 3. The movie shows three repetitions of each of the two avatars ('salsa follower', 'salsa leader', 'taking', 'catching').

Chapter 2. Lighting-from-above prior in biological motion perception.

(published as

L Fedorov, T Dijkstra, , M Giese (2018). Lighting-from-above-prior in biological motion perception. Scientific Reports. 8: 1507)

SCIENTIFIC REPORTS



OPEN

Lighting-from-above prior in biological motion perception

Leonid A. Fedorov^{1,3}, Tjeerd M. H. Dijkstra^{1,2} & Martin A. Giese^{1,3}

The visual system is able to recognize body motion from impoverished stimuli. This requires combining stimulus information with visual priors. We present a new visual illusion showing that one of these priors is the assumption that bodies are typically illuminated from above. A change of illumination direction from above to below flips the perceived locomotion direction of a biological motion stimulus. Control experiments show that the underlying mechanism is different from shape-from-shading and directly combines information about body motion with a lighting-from-above prior. We further show that the illusion is critically dependent on the intrinsic luminance gradients of the most mobile parts of the moving body. We present a neural model with physiologically plausible mechanisms that accounts for the illusion and shows how the illumination prior might be encoded within the visual pathway. Our experiments demonstrate, for the first time, a direct influence of illumination priors in high-level motion vision.

The perception of body motion is dependent on a variety of cues, including 2D form and motion^{1,2}, but also on other cues which help to disambiguate the three-dimensional structure of the body, such as disparity³⁻⁶. While natural body motion stimuli often specify many cues for the disambiguation of the three-dimensional body structure, it has been shown that humans effortlessly recognize three-dimensional body motion even from strongly impoverished two-dimensional stimuli⁷. This requires the combination of ambiguous stimulus information with perceptual priors that are encoded by the visual system. The exact nature of such priors for the recognition of three-dimensional body motion remains largely unknown.

We present a new perceptual illusion that implies that the perceived locomotion direction of body motion stimuli critically depends on the prior assumption that such bodies typically are illuminated from above. Such 'lighting-from-above priors' have been previously found for the perception of static shapes⁸⁻¹⁷. However, the influence of illumination direction and shading on body motion perception has never been systematically studied. Illumination from above results in the perception of the correct locomotion direction, while illumination from below can completely flip the perceived direction of locomotion. As shown by an additional control experiment, the observed illusion is not just a side-effect of classical shape-from-shading mechanisms for the perception of static shapes, and their dependence on illumination direction. Instead, it must be based on a specific previously unknown mechanism that seems to combine temporally changing intrinsic shading gradients of object surfaces (i.e. gradients that are not caused by the object boundaries) with the perceived illumination direction.

In the following, we present two experiments. Our first experiment establishes the illusion, showing that flipping the light-source position from above to below can completely change the perceived walking direction of a biological motion stimulus. In a second experiment, we isolate the visual features that critically drive this visual illusion. Our experiments motivate a computational model that accounts for the illusion, and which proposes a way how the underlying visual prior might be encoded by physiologically plausible neural mechanisms within the visual pathway.

Results

To investigate the influence of illumination direction on the perception of walking direction, we developed a novel biological motion stimulus, consisting of 11 conic volumetric elements with reflectional symmetry (Figs 1A and 2B-E). The movements of the elements were derived from motion-captured movements of a human walker (see Methods for details). It is well-known that 2D images of illuminated three-dimensional surfaces specify shading gradients that allow an estimation of the surface orientation. This estimation is also known as classical

¹Section for Computational Sensomotrics, Dept. Cognitive Neurology, CIN & HIH, UKT, University of Tübingen, Otfried-Müller Strasse 25, 72076, Tübingen, Germany. ²Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076, Tübingen, Germany. ³International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Spemannstrasse 38, 72076, Tübingen, Germany. Correspondence and requests for materials should be addressed to M.A.G. (email: martin.giese@uni-tuebingen.de)

Received: 14 June 2017

Accepted: 2 January 2018

Published online: 24 January 2018

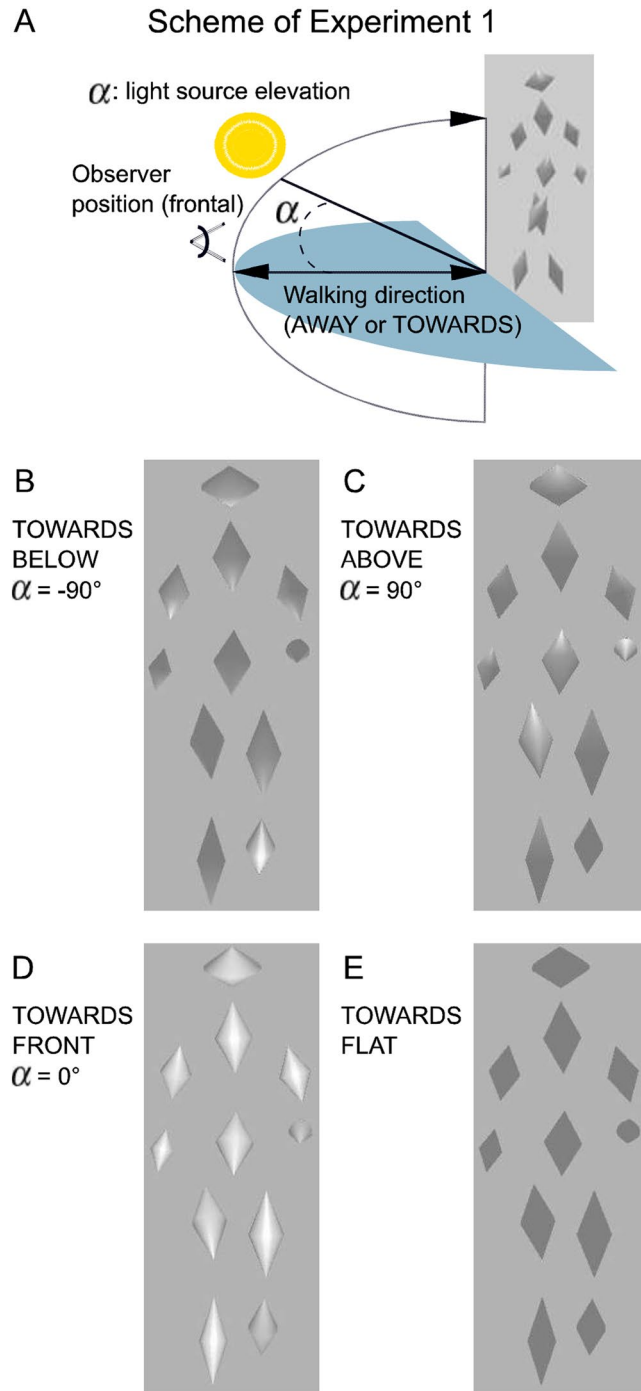


Figure 1. Experimental paradigm and stimuli snapshots. **(A)** Scheme of experimental setup. Participants were viewing a walker moving TOWARDS or AWAY from them. The walker consists of volumetric conic elements with a reflective grayscale surface. It was rendered assuming a light source position with a fixed elevation angle. The walker performed two gait cycles before participants were asked to report the perceived walking direction. **(B–D)** Characteristic snapshots of the same body configuration of the walker during the TOWARDS gait with light source positioned at different elevation angles. **(E)** Snapshot of the walker with ‘flat’ shading with uniform shading within the individual elements. Movie 1 shows these 4 walker stimuli.

‘shape-from-shading’ problem¹⁸. In this paper we investigated the influence of the such shading gradients on the perceived locomotion direction from a biological motion stimulus that consists of volumetric elements.

In Experiment 1 the elements were illuminated by a light source whose position was systematically varied (Fig. 1A). Previous work on the perception of walking from point-light stimuli has shown that direction perception can become ambiguous for particular view angles if no additional depth cues are provided^{3,6}. The view of the body was chosen to minimize occlusions between different stimulus elements (see e.g. Figure 2B–E), which

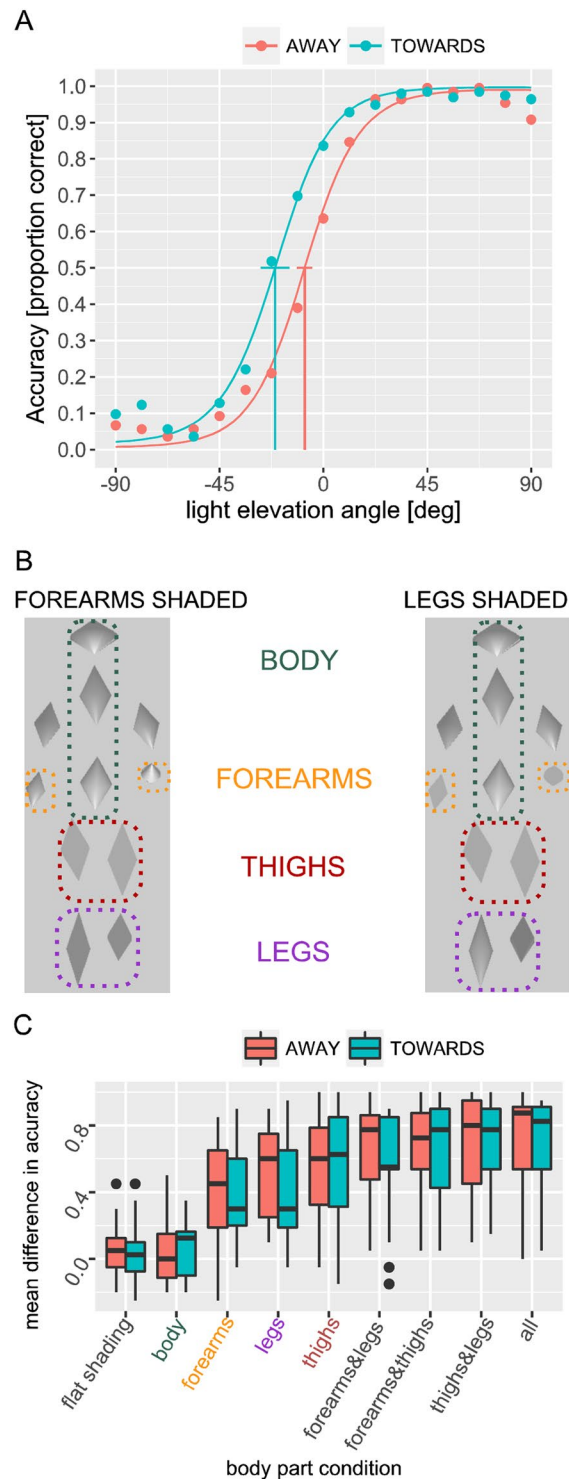


Figure 2. Experimental Results. **(A)** Results of Experiment 1. Accuracy of reporting the veridical walking direction as a function of the light source elevation angle α . Accuracy is defined as a probability of perceiving the true walking direction. Plotted points represent the means of the veridical binary responses per condition. The psychometric function was fitted with a generalized linear mixed effects model using cosine and sine of the light elevation angle, and walking direction as predictors (GLMM). **(B)** Snapshots from example stimuli lit from BELOW walking AWAY. Walkers illuminated from ABOVE and BELOW (light elevation angles ± 45 deg) were presented for which the gradual shading was removed from different combinations of stimulus elements. Left: 'forearms' condition where gradual shading was removed from the thighs and the legs. Right: 'legs' condition where gradual shading was removed from the thighs and the forearms. Except for the 'flat shading' condition the trunk and the upper arms always had gradual shading. Movie 2 shows these 2 walker stimuli. **(C)** Results of Experiment 2. Boxplot of the mean difference of the response accuracies (probabilities of correct reporting of the veridical walking direction) between stimuli illuminated from ABOVE and BELOW. This measure of

the size of the illusion is shown for different combinations of elements with gradual shading. Boxes indicate the ranges of the data (middle 50% interquartile range (IQR)). Black thick lines within the boxes indicate the medians. Whiskers mark intervals of 1.5 times the IQR ranges, and the dots indicate outliers that do not fall within these intervals.

maximizes the ambiguity in absence of shading cues because occlusions provide relative depth information¹⁹. The true walking direction of the walker was either straight out of the image plane in the direction of the observer (TOWARDS) or into the image plane away from the observer (AWAY). The light source elevation angles α (Fig. 1A) varied between 90 deg (illumination exactly from above) to -90 deg (illumination exactly from below). Elevation angle $\alpha = 0$ deg corresponds to an illumination directly from the side. Within a forced-choice task, participants responded whether the walker was perceived as walking 'towards' them or 'away' from them.

Participants always reported perceiving a walking human character. Whether this character was perceived as walking towards them or away from them depended on the light source elevation angle. Figure 2A shows the accuracy of the responses (proportion of correct responses where the reported direction matched the true walking direction of the walker) averaged over 13 observers (represented as points in Fig. 2A for illustration). Individual accuracies are reported in Fig. S1 in the Supplemental Information (SI). Illumination from above ($\alpha > 20$ deg) results typically in correct perception of the veridical walking direction, while illumination from below ($\alpha < -40$ deg) results in an illusion: the perception of walking opposite to the veridical direction. In an intermediate regime of elevation angles (about -40 to $+20$ deg) the stimulus was multi-stable and the percept alternated between the two veridical walking directions. Within individual trials, observers never reported switches during the stimulus presentation. Figure S1 (SI) shows that the illusion was present in the responses of all 13 observers.

Responses of all observers were fitted with a logistic mixed-effects model (see Methods) with the cosine and sine of light angle and veridical motion direction both as fixed and random effects. This analysis uncovers systematic effects for all observers while still allowing for individual differences. The resulting fixed effects curves are plotted in Fig. 2A, one for AWAY and one for TOWARDS and the random effect ones are plotted in Fig. S1. The fits showed a highly significant effect of light angle ($p < 10^{-16}$) on the perceived walking direction for both true walking directions (AWAY and TOWARDS). In addition, our analysis revealed a significant small effect of walking direction ($p < 0.05$). The small significant effect of walking direction is consistent with a bias that favors perception of walking 'towards' the observer, which has been also observed in previous studies with point-light-walkers²⁰. A further analysis of the condition with frontal lighting (cf. Fig. S2) reveals that the veridical walking directions can be perceived with a performance above chance level even for this condition. This might be explained by the presence of subtle temporally changing shading variations within the stimulus elements even for this illumination condition.

Summarizing, the results of Experiment 1 show that the perception of body motion is influenced by a 'lighting-from-above prior'. For illumination from above the walking direction is correctly perceived from the 2D stimulus and identical with the veridical locomotion direction of the 3D stimulus. Illumination from below, however, results in a misperception where the walker is perceived as walking in the opposite direction of the veridical locomotion.

Our stimuli were designed in a way that minimizes occlusions between different stimulus elements. Occlusion is a strong relative depth cue, which also could disambiguate the three-dimensional structure of our biological motion stimuli. Since this cue was minimized, this leaves mainly shading variations within the elements (intrinsic shading gradients) as possible depth cue. This motivated us to investigate what happens when we eliminate all intrinsic gradual shading cues from our stimuli. For this purpose, we replaced the luminance values of all pixels belonging to an element by the average luminance, averaging over all pixels that form the element and over all frames of a gait cycle. The resulting stimulus elements have a 'flat' shading profile that was constant over time within the elements (Fig. 1E). Note that this flat shading stimulus is different from the stimulus with frontal lighting, which still contains small changing shading gradients within the moving stimulus elements that change their orientation relative to the light source (Fig. 1D). We embedded trials with flat shading in the stimulus trials with gradual shading (as presented before in Fig. 1B–D). Consistent with our expectation, stimuli with flat shading were perceptually ambiguous and sometimes perceived as walking towards and sometimes as away from the observer. To test if there was any information about walker direction used by the observers, we fitted a logistic mixed-effects model using only an intercept as predictor to the data from trials with flat shading. We found a small but significant ($p < 0.05$) negative deviation of the response accuracy from chance level, i.e. observers performed worse than chance. Thus, observers made no use of the remaining information about walking direction in the flat shaded stimuli. If anything, the remaining information resulted in the perception of the wrong walking direction. Further statistical analysis results on this stimulus class is presented in the SI and Fig. S3. This result implies that the information about the walking direction is largely carried by the gradual shading within the stimulus elements, while variations of the movement kinematics and shape variations of the boundaries of the stimulus elements are apparently not exploited by the visual system even though they also contain information about the walking direction.

The new illusion was further investigated in Experiment 2, in which we varied the amount of shading information provided by the individual stimulus elements. As illustrated in Fig. 2B, we removed the shading from combinations of elements (e.g. the ones forming the forearms or the legs). In total, we used a set of nine stimuli, ranging from the original fully shaded stimulus to a stimulus with flat shading within all stimulus elements (Fig. 1E). Specifically, we removed the gradual shading from the elements forming the head, torso, the forearms and upper arms, the thighs and the lower legs. The veridical motion of the walker was again either AWAY or TOWARDS the observer. As in the first experiment, participants responded within a forced-choice task whether they perceived

the stimulus as walking towards or away from them. For this experiment, we used only two light source directions ($\alpha = -45$ deg and $\alpha = 45$ deg), which showed large differences in accuracy between illumination from above and from below for fully shaded stimuli in Experiment 1.

The results from the second experiment are shown in Fig. 2C, separately for the stimuli with veridical motion AWAY from and TOWARDS the observer. The figure shows the mean difference in accuracy between the lighting from above and below conditions for stimuli with gradual shading in different combinations of stimulus elements (see Fig. S4 for per-observer averages). The differences were averaged over repetitions, where the colored bars indicate the ranges of the means across the different participants. The mean difference in accuracy between the two illumination conditions characterizes the size of the illusion. Consistent with our expectation, the size of the illusion increases with the fraction of elements with gradual shading. Consistent with the findings in the first experiment, for the condition with gradual shading of all elements ('all') we observe a large difference in accuracy (0.85), which is close to the one found in Experiment 1 (0.88). For all conditions with gradual shading of stimulus elements in the forearms, the legs, the thighs or combinations of them the difference in accuracy deviated significantly from zero (one-sample t test, $p < 10^{-4}$ for both AWAY and TOWARDS). Contrasting with this result, stimuli without any gradual shading within the elements (labeled "none") and the ones with gradual shading of the elements that form the head, torso and the upper arms (labeled "body") show mean differences in accuracy that do not deviate significantly from zero (one-sample t test, $p > 0.15$). This indicates the absence of the illusion for those stimuli. A more detailed statistical analysis using a linear mixed-effects regression²¹ is presented in the SI. This analysis confirms that the head, torso and upper arm elements do not contribute to the illusion, while forearm, thigh and lower leg elements induce a significant illusion. In addition, the analysis shows that of the three groups of moving elements, the forearms are least effective in inducing the illusion, followed by the (lower) legs, whereas the thighs were most effective in inducing the illusory effect. From this we conclude that the gradual shading cues from the mobile elements of the walker are critical for the illusion, since in our stimulus the elements representing the head, the torso and the upper arms do not show much motion.

To further support our conclusion that the illusion is driven by the intrinsic shading gradients in the mobile stimulus elements, we developed a computational neural model that recognizes body motion by an analysis of luminance gradients. The model is based on a hierarchical neural architecture and is compatible with facts known about the visual pathway. The model learns a perceptual prior from training data that contains only stimuli that are illuminated from above. We demonstrate that this model reproduces the illusion shown in Experiment 1 and that it also reproduces qualitatively the results about the most informative features from Experiment 2.

The model is illustrated in Fig. 3. It is formed by a hierarchy of four layers that consist of neural detectors. The first layer consists of Gabor filters, modeling V1 simple cells, where the uneven Gabor filters estimate local luminance gradients. The second layer performs nonlinear gating to suppress the strong gradients on the boundaries of the stimulus elements. Since typically the background contrast is different from the one of the stimulus element this creates strong contrast edges, which without suppression would dominate in the higher levels of the hierarchy. The gating operation suppresses the responses of the detectors to these contrast edges. The third layer pools these gated filter outputs over limited spatial regions using a maximum operation, resulting in detector responses with increased position invariance^{22,23}. To determine the connections to the fourth layer we applied a feature selection algorithm, which selects only those receptive fields in layer 3 whose responses vary significantly over the training set. A further reduction of the dimensionality of the feature space is accomplished by Principal Component Analysis (PCA). The resulting reduced feature vectors provide input to the highest layer that consists of two Gaussian radial basis function units, which model a two-component Gaussian mixture distribution mixture (one component encoding AWAY and the other TOWARDS walking). The parameters of this distribution were learned in an unsupervised manner from stimuli from both veridical walking directions that were illuminated from above ($\alpha = 78.75$ deg). Our model thus assumes that the visual system is trained with typical stimuli, which are illuminated from above, implementing a learned perceptual prior. The 'perceptual response' of the model was then determined by the radial basis function unit with the largest response, where it can be shown that this decision rule implements a Bayesian classifier. (See SI for further details).

When the model was tested with the stimuli from Experiment 1 it reproduced the experimentally observed illusion. This is illustrated in Fig. 4A that shows the probability to classify the veridical walking direction of the walking stimuli, which were illuminated from the same directions as in the experiment. Like the human participants, the model misclassifies the walking direction for stimuli that are illuminated from below. The likelihood of correct classification increases as a function of the elevation angle of the light source, consistent with the results from Experiment 1, where in our analysis we averaged the responses of the AWAY and TOWARDS conditions. (See SI for details).

We also tested the model with the stimulus variants from Experiment 2. Like in Experiment 2, the size of the illusion increases with the number of gradually shaded elements of the mobile body parts, while shading of the head and torso is not providing reliable information for the classification of the walking direction. The stimulus without gradual shading cues results in completely ambiguous responses with equal probability of both classification results. The model largely reproduces the relative importance of the individual elements for the illusion size. This is shown in Fig. 4B that shows the mean differences in accuracy (that quantifies the strength of the illusion) from the experiment and the one derived from the model. The two measures are significantly correlated ($R^2 = 0.8$, $p < 0.001$). This result further supports the hypothesis that the illusion is based on an analysis of intrinsic luminance gradients of body motion stimuli. The higher illusion size for stimuli with shaded legs might result from the fact that during the observation of body movement humans tend to attend the body center²⁴ while the model treats all body parts equally and does not account for such attentional biases.

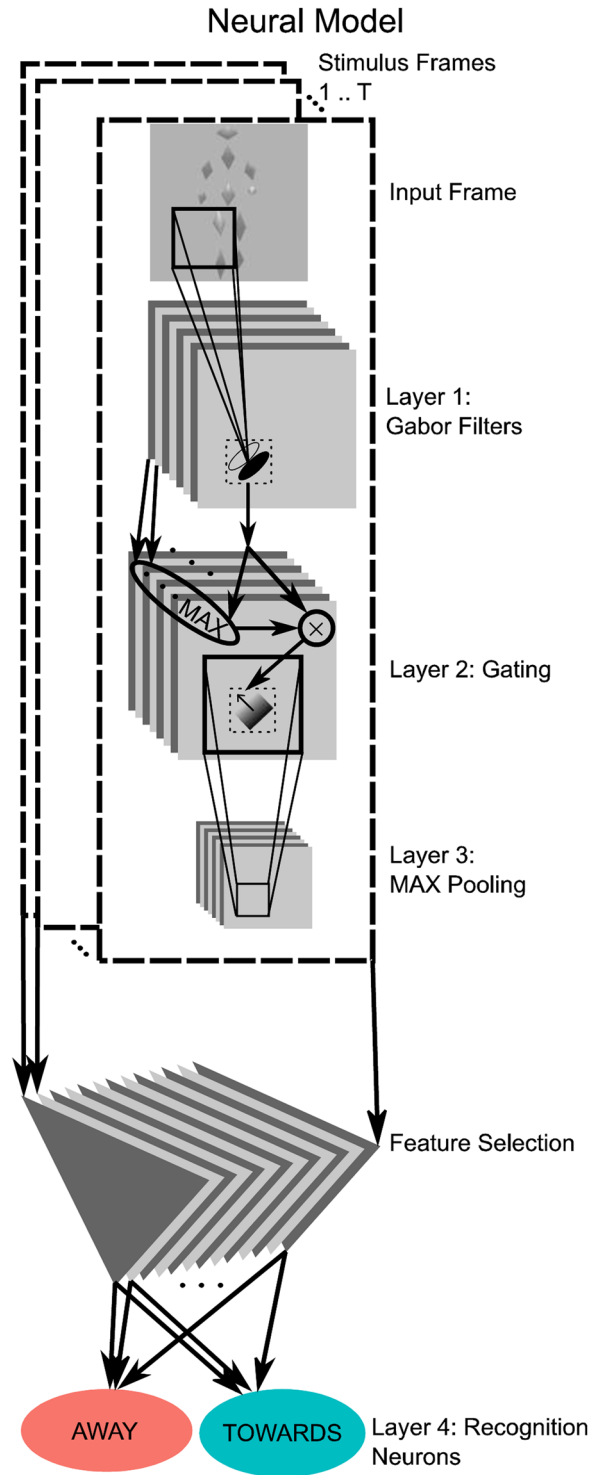


Figure 3. Neural model. The model consists of four neural layers: (1) uneven Gabor filters that are sensitive to shading gradients, (2) a gating stage that suppresses the strong contrast edges on the boundary of the silhouette of the walker; (3) partly position-invariant neurons that detect the strengths and direction of luminance gradients within the individual parts of the moving Figure; (4) a recognition level that processes selected features transmitted from the previous level. This level is composed from two Gaussian radial basis functions units that are trained to approximate the statistics of training patterns, which all have been illuminated from above. (See text and SI for further details.).

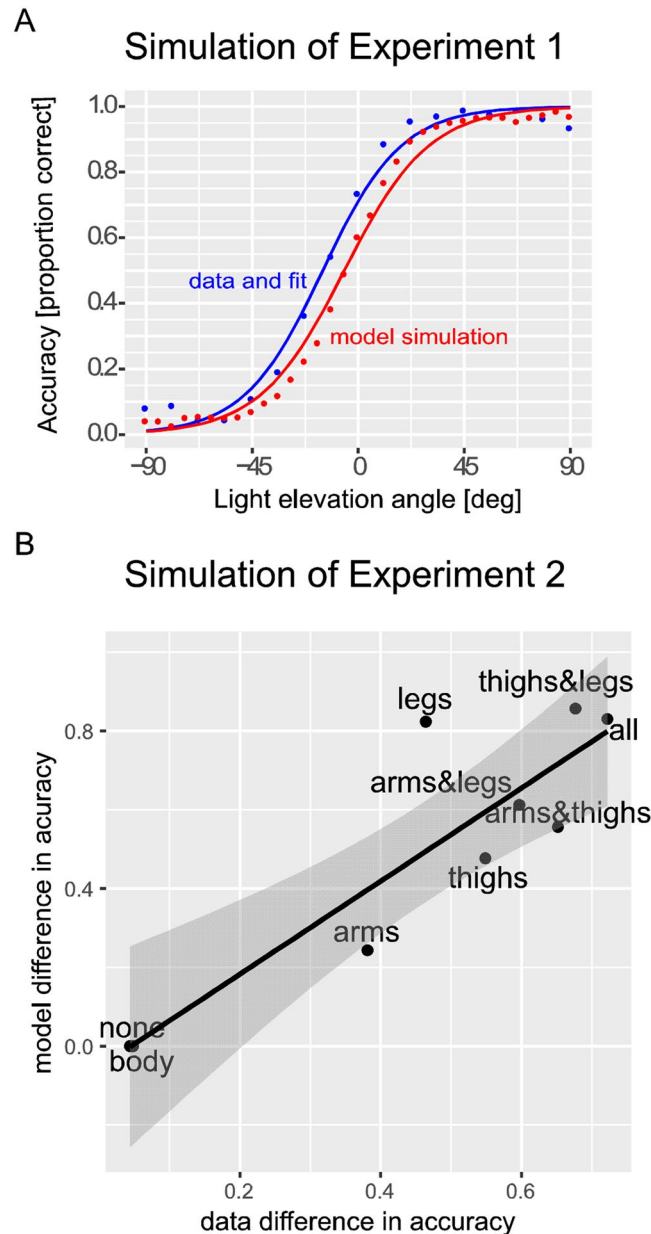


Figure 4. Simulation of experiments by the model. **(A)** Simulation of Experiment 1. The model reproduces the illusory effect, closely approximating the functional form obtained from the experimental data. Psychometric functions are averaged over patterns with the veridical walking directions AWAY and TOWARDS. **(B)** Simulation of Experiment 2. Separately for each stimulus type in Experiment 2, the correlation plot shows the mean differences of accuracies between illumination from above and below (cf. Figure 2C), as computed from the experimental data and the model predictions. The correlation between both measures is high (adjusted $R^2 = 0.7695$) and significant ($p < 0.01$).

Discussion

We presented a new psychophysical illusion that provides evidence that the perception of body motion is influenced by a lighting-from-above prior. Using a novel biological motion stimulus that consists of moving volumetric elements, we showed that the perceived walking direction matches the veridical walking direction when the stimulus was illuminated from above. If the stimulus was illuminated from below, however, the walking direction was misperceived. This implies that body motion perception integrates the stimulus information with the a-priori assumption that the light source is typically positioned above. While our first experiment established this novel psychophysical illusion, our second experiment narrowed down the relevant visual features. Critical for the illusion were the shading gradients within the most mobile stimulus elements. We could qualitatively reproduce the illusion, and its dependence on these critical features by a neural model that analyzes intrinsic shading gradients inside the moving stimulus elements. In the model the lighting-from-above prior was learned from training patterns that were illuminated from above. Such training reflects what humans might experience in the visual world

during the maturing of the visual system. Without further training the model spontaneously shows the illusion, i.e. the misperception of walking direction for stimuli that are illuminated from below. In addition, the model reproduced the dependence on critical shading features as tested in Experiment 2.

Priors for illumination direction have been reported previously for other visual functions, including the perception of static shapes^{10,13–15,25,26}, visual search, and reflection perception²⁷. One might thus argue that our illusion does not reveal a new perceptual process, because it might be explained by the well-known dependence of static shape perception on illumination direction in individual frames. We argue against this criticism, maintaining the claim that our illusion reveals a novel and fundamentally different perceptual process that directly analyses the dynamically changing intrinsic shading information. In order to provide support for this claim, we ran an additional control experiment.

In this control experiment we presented stimuli that prevented the reconstruction of 3D limb orientation from individual frames, while maintaining approximately the temporally varying intrinsic shading gradients of the individual stimulus elements. For this purpose, we replaced the rigid conic stimulus elements by elements with a fixed circular shape. The intrinsic luminance patterns of these elements were obtained by spatial warping of the texture of the conic stimulus elements in Experiment 1 onto these circular shapes (see Fig. S5A and the SI for details). Control subjects observing these control stimuli (see Supplementary Movies 7 and 8) perceived the elements as ‘deforming rubber sheets’, and they were not able to reconstruct reliably the 3D orientation of these elements from individual frames. However, the illusory effect was retained for these stimuli (Fig. S5B). Fitting a logistic mixed effects models to the data as for Experiment 1 we obtained a significant illusory effect of walking direction ($p < 0.001$). This result provides strong support for the claim that the illusion reported in this paper cannot be explained by classical shape-form-shading mechanisms, by an estimation of 3D segment orientations in individual keyframes. Rather, it must be based on a special potentially body motion-specific process.

Our neural model predicts the illusion by learning the relevant shading cues from example movies with illumination from above. A thorough analysis of the similarities of the intrinsic shading features for walking in opposite directions for opposite illumination directions explains why the model, if trained only with patterns illuminated from above, explains the misperception of walking direction by the model. To our knowledge, our model is the first one that accounts for the influence of shading on body motion perception. Further extensions of the model account also for dynamical aspects of the multi-stable perception of such body motion stimuli, such as switching rates and hysteresis²⁸, as well as for the integration of the intrinsic shading features with the contour cues of the body silhouette²⁹.

In order to rule out that the observed psychophysical results, and specifically the illusion, can be explained by simple low-level motion perception, instead of a more sophisticated process related to biological motion, we performed an analysis of the average optic flow generated by the stimuli in Experiment 1. We varied the light source position and walking direction and for each condition computed responses of hypothetical motion-sensitive neurons representing the total motion energy in one of eight directions. While we found that these neural responses showed reliable differences between the different conditions, the pattern of these differences was incompatible with the observed psychophysical results (smoothness of response curves and their dependence on the light source position). Moreover, even the misclassification when the light source is flipped was not reproduced by this simplified model.

Since our model is based on simple mechanisms that, in principle, can be implemented with cortical neurons (filtering, pooling, gain modulation/multiplicative gating, template matching) it makes specific predictions about cell types in the proposed visual pathway. Neurons involved in the processing of body motion stimuli have been found in macaque superior temporal cortex^{4,30}. In addition, our model postulates a suppression of the contour information on the boundary of the body silhouette. Such a suppression of information on figure boundaries has been proposed also in models for other visual functions^{31,32}. Electrophysiological studies will be required to unravel whether the postulated mechanisms for shading analysis really approximate computations in the biological visual pathway.

Methods

Apparatus. Both experiments were performed on a Dell Precision computer using the MATLAB Psychophysics Toolbox version 3. Stimuli were displayed on a 24-inch BenQ XL2420-B LCD monitor with 1920×1080 pixels resolution and a refresh rate of 120 Hz. Stimuli were viewed from a distance of 60 cm.

Stimuli. All stimuli were pre-rendered before an experimental session and were identical for all participants. The stimuli presented a movie of a walking figure with a resolution of 800 by 600 pixels. The walking figure itself always fit into a 250×600 pixel box. The walker performed two gait cycles (four steps) and then a text was displayed that asked for the perceived perceptual alternative. The response was given by pressing one of the two buttons on the keyboard. Subsequently, the movie with the next experimental condition was started. One gait cycle took about 1 second.

Procedure. Different observers participated in Experiments 1 and 2. Before both experiments, participants were presented with two movies of walkers lit from a 78.75-degree elevation angle walking away and towards. In this instructional step, the movies were viewed continuously until the participants confirmed seeing the veridical walking direction in both cases. To make sure the participants can follow the experimental procedure, they were then presented with a short experimental block consisting of a 20% random subset of all conditions in the experiment. No feedback was given.

In both experiments, all conditions were block-wise randomly permuted. They were presented subsequently without breaks between the blocks. In case participants wanted a break, they could stop the stimulus sequence and continue after the break. The whole experimental procedure including the instruction phase lasted less than 1 hour in both cases.

Experiment 1 comprised 36 conditions (17 light source positions and the condition with flat shading, each presented for two veridical walking directions). Each condition was repeated 15 times, resulting in a total of 540 trials. Experiment 2 included 36 conditions (9 different combinations of shaded elements, two walking directions, and two different light source positions). Because we expected the effects in this experiment to be subtler, we used 20 repetitions, resulting in a total of 720 trials

Participants. Thirteen volunteers (mean age 25) participated in Experiment 1 of which seven were females. In Experiment 2 sixteen volunteers participated of which 7 were females. All participants were naïve about the goals of the study and were compensated by 10 EURO per hour. After the experiments, they were debriefed and informed about the study.

Motion capture. We used the processed motion capture data from the experiments of Roether and colleagues³³. For both experiments we used a single female walker performing an emotionally neutral gait, as defined in the above reference. For the model simulation, we used motion capture data from 3 extra actors (2 male, 1 female) also performing an emotionally neutral gait.

Consent. Psychophysical experiments were performed with informed consent of participants. All experimental procedures were approved by the ethics board of the University of Tübingen (Germany) and all experiments were performed in accordance with relevant guidelines and regulations.

Rendering of the surface shading and light source position. The walker was composed of conic elements rendered as surfaces in MATLAB 2014b. The element sizes were adjusted manually to match the geometry of a walking human. We chose an infinite light source distance, resulting in parallel light rays of the illumination field (choosing white as ray color). In both experiments we varied the elevation angle of the light source. In Experiment 1 the elevation varied from -90 degrees to 90 degrees, in 17 equidistant steps, and in Experiment 2 we used the two elevation angles -45 and 45 degrees, which maximized the size of the illusory effect. We used the ZBuffer renderer, which allows to specify the parameters AmbientStrength, SpecularStrength, DiffusionStrength and SpecularExponent of the surface. AmbientStrength refers to the amount of light present at every point of a scene, while the other three parameters refer to the surface reflectance properties. The walking figure was rendered on a gray background. For both experiments, we use the settings: AmbientStrength = 0.5, SpecularStrength = 0.3, and SpecularExponent = 10. For Experiment 1 we chose DiffusionStrength = 0.5, and BackgroundColor = [0.75 0.75 0.75]. For Experiment 2 we chose DiffusionStrength = 0.4, and BackgroundColor = [0.8 0.8 0.8]. For all shaded surfaces we specified FaceColor = [0.99 0.99 0.99] and removed all surface edges. The Gouraud lighting algorithm was exploited to compute the pixel colors for the specified light source positions.

For the elements with ‘flat’ shading we set the FaceColor to a constant. To calculate its value, we individually rendered each shaded element and computed the average pixel brightness over a full gait cycle.

Neural model. Space permits only a very brief summary of the model here and we refer to our previous work²⁹ for a more complete description (parameters of the model are summarized in Supplementary Table 1). The stimulus set for training and testing of the model was generated from motion capture data from 4 actors (2 male and 2 female). From each actor, we generated 25 three-dimensional body models with randomly varying sizes of the conic elements. One of these models was identical with the one used to generate the stimuli for the psychophysical experiment. Each model was rendered for the two veridical walking directions (AWAY and TOWARDS), assuming 33 different light source positions with elevation angles that varied equidistantly between -90 deg to 90 deg.

The model was trained with the stimuli (both veridical walking directions) rendered with a single elevation angle of 78.75 degrees, simulating illumination from above and using the data from all 4 actors and 25 body shape models with varying shape parameters. This variability in the training set prevents overfitting of individual training stimuli and makes the recognition more robust. For Experiment 1 the model was tested with all generated stimuli (in total 6600) using all 33 light source positions. To simulate Experiment 2, the training set was the same, and the test stimuli were generated from, also, 4 different actors, using only two light source positions, but rendered with 9 experimental conditions with different combinations of elements with or without intrinsic shading gradients (Fig. 4A). In total, only $\sim 3\%$ of the movies were used for training the model of Experiment 1 and $\sim 10.0\%$ of the movies were used for training the model of Experiment 2.

Data availability statement. All relevant data are available as supplementary information files, with captions included in the SI.

References

- Giese, M. A. & Poggio, T. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* **4**, 179–192, <https://doi.org/10.1038/nrn1057> (2003).
- Lange, J. & Lappe, M. A model of biological motion perception from configural form cues. *J Neurosci* **26**, 2894–2906, <https://doi.org/10.1523/JNEUROSCI.4915-05.2006> (2006).
- Vanrie, J. & Verfaillie, K. Perceiving depth in point-light actions. *Percept Psychophys* **68**, 601–612 (2006).
- Vangeneugden, J. *et al.* Distinct mechanisms for coding of visual actions in macaque temporal cortex. *J Neurosci* **31**, 385–401, <https://doi.org/10.1523/JNEUROSCI.2703-10.2011> (2011).
- Bülthoff, I., Bülthoff, H. & Sinha, P. Top-down influences on stereoscopic depth-perception. *Nat Neurosci* **1**, 254–257, <https://doi.org/10.1038/699> (1998).
- Vanrie, J., Dekeyser, M. & Verfaillie, K. Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception* **33**, 547–560 (2004).
- Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* **14**, 201–211 (1973).

8. Ramachandran, V. S. Perception of shape from shading. *Nature* **331**, 163–166, <https://doi.org/10.1038/331163a0> (1988).
9. Sun, J. & Perona, P. Shading and stereo in early perception of shape and reflectance. *Perception* **26**, 519–529 (1997).
10. Adams, W. J., Graf, E. W. & Ernst, M. O. Experience can change the ‘light-from-above’ prior. *Nat Neurosci* **7**, 1057–1058, <https://doi.org/10.1038/nn1312> (2004).
11. Brewster, D. On the conversion of relief by inverted vision. *Edinburgh Philosophical Transactions* **15**, 657–662, <https://doi.org/10.1017/S0080456800030234> (1844).
12. Kleffner, D. A. & Ramachandran, V. S. On the perception of shape from shading. *Percept Psychophys* **52**, 18–36 (1992).
13. Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z. & Connor, C. E. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* **11**, 1352–1360, <https://doi.org/10.1038/nm.2202> (2008).
14. Tsutsui, K., Sakata, H., Naganuma, T. & Taira, M. Neural correlates for perception of 3D surface orientation from texture gradient. *Science* **298**, 409–412, <https://doi.org/10.1126/science.1074128> (2002).
15. Fleming, R. W., Holtmann-Rice, D. & Bulthoff, H. H. Estimation of 3D shape from image orientations. *Proc Natl Acad Sci USA* **108**, 20438–20443, <https://doi.org/10.1073/pnas.1114619109> (2011).
16. Gerardin, P., Kourtzi, Z. & Mamassian, P. Prior knowledge of illumination for 3D perception in the human brain. *Proc Natl Acad Sci USA* **107**, 16309–16314, <https://doi.org/10.1073/pnas.1006285107> (2010).
17. Mamassian, P. Bayesian inference of form and shape. *Prog Brain Res* **154**, 265–270, [https://doi.org/10.1016/S0079-6123\(06\)54014-2](https://doi.org/10.1016/S0079-6123(06)54014-2) (2006).
18. Horn, B. & Brooks, M. *Shape From Shading*. 586 (MIT Press, 1989).
19. Howard, I., Rogers, B. *Seeing in Depth*. (Oxford University Press, 2008).
20. Schouten, B., Davila, A. & Verfaillie, K. Further explorations of the facing bias in biological motion perception: perspective cues, observer sex, and response times. *PLoS One* **8**, e56978, <https://doi.org/10.1371/journal.pone.0056978> (2013).
21. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48, <https://doi.org/10.18637/jss.v067.i01> (2015).
22. Fukushima, K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* **36**, 193–202 (1980).
23. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019–1025, <https://doi.org/10.1038/14819> (1999).
24. Saunders, D. R., Williamson, D. K. & Troje, N. F. Gaze patterns during perception of direction and gender from biological motion. *J Vis* **10**, 9, <https://doi.org/10.1167/10.11.9> (2010).
25. Welchman, A. E., Deubelius, A., Conrad, V., Bulthoff, H. H. & Kourtzi, Z. 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci* **8**, 820–827, <https://doi.org/10.1038/nn1461> (2005).
26. Stone, J. V., Kerrigan, I. S. & Porrill, J. Where is the light? Bayesian perceptual priors for lighting direction. *Proc Biol Sci* **276**, 1797–1804, <https://doi.org/10.1098/rspb.2008.1635> (2009).
27. Adams, W. J. A common light-prior for visual search, shape, and reflectance judgments. *J Vis* **7**, 11 11–17, <https://doi.org/10.1167/7.11.11> (2007).
28. Sting, L., Fedorov, L., Dijkstra, T., Hock, H. & Giese, M. Dynamics of multistable biological motion perception. *Journal of Vision* **17**, 70, <https://doi.org/10.1167/17.10.70> (2017).
29. Fedorov, L., Vangeneugden, J. & Giese, M. Neural Model for the Influence of Shading on the Multistability of the Perception of Body Motion. *Proceedings of the 8th International Joint Conference on Computational Intelligence* **2**, 69–76, <https://doi.org/10.5220/0006054000690076> (2016).
30. Singer, J. M. & Sheinberg, D. L. Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *J Neurosci* **30**, 3133–3145, <https://doi.org/10.1523/JNEUROSCI.3211-09.2010> (2010).
31. Koch, C., Marroquin, J. & Yuille, A. Analog “neuronal” networks in early vision. *Proc Natl Acad Sci USA* **83**, 4263–4267 (1986).
32. Lee, T. S., & Yuille, A. Efficient Coding of Visual Scenes by Grouping and Segmentation. *Bayesian Brain: Probabilistic Approaches to Neural Coding*, 145–188, <https://doi.org/10.7551/mitpress/9780262042383.001.0001> (2006).
33. Roether, C. L., Omlor, L., Christensen, A. & Giese, M. A. Critical features for the perception of emotion from gait. *J Vis* **9**, 15 11–32, <https://doi.org/10.1167/9.6.15> (2009).

Acknowledgements

The first author thanks Albert Mukovskiy for help with rendering the stimuli. The research leading to these results has received funding from DFG GI 305/4-1, DFG GZ: KA 1258/15-1; FP7-PEOPLE-2011-ITN (Marie Curie): ABC PITN-GA-011-290011, CogIMon H2020 ICT-23-2014 /644727, HFSP RGP0036/2016. We also acknowledge support by Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of University of Tübingen.

Author Contributions

L.F. performed the experiments, analyzed the data, implemented the model and wrote the paper. T.D. analyzed the data and wrote the manuscript. M.G. designed the research, supported the data analysis and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-19851-8>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Supplementary Information

Title: "Lighting-from-above prior in biological motion perception"

Authors: Leonid A Fedorov^{1,3}, Tjeerd MH Dijkstra^{1,2} and Martin A Giese^{1,3}

Author Affiliation:

¹Section for Computational Sensomotorics, Dept. Cognitive Neurology, CIN & HIH, UKT, University of Tübingen, Otfried-Müller Strasse 25, 72076 Tübingen, Germany

²Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

³International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Spemannstrasse 38, 72076 Tübingen, Germany

Neural model

Our hierarchical neural model consists of four layers that implement a position-invariant recognition of walking direction based on the intrinsic gradual shading variations of the individual stimulus elements. The model reproduces qualitatively the illusion, as well as the dependence of the illusion size on the available shading information.

The first layer of the model is composed from uneven Gabor filters G_u that are ordered within a rectangular spatial grid. Such filters are sensitive to oriented local luminance gradients. We assume that the receptive field center of the filter is specified by the vector (x_c, y_c) , and that its preferred gradient direction is given by the angle α . We assume further that the receptive field size is specified by the parameter σ , and the preferred spatial frequency by the constant k_0 . With these parameters, the filter functions are defined as:

$$G_u(x, y; x_c, y_c, \alpha, \sigma) = \exp\left(-\frac{\tilde{x}^2 + \tilde{y}^2}{2\sigma^2}\right) \sin(2\pi k_0 \tilde{x}) \quad (1)$$

where $\tilde{x} = (x - x_c) \cos \alpha + (y - y_c) \sin \alpha$ and $\tilde{y} = -(x - x_c) \sin \alpha + (y - y_c) \cos \alpha$. We used 810,000 spatial grid points for the receptive field centers and eight different angles α (cf. Tab. 1). Assuming $I(x, y)$ as the gray-level input pixel image, the output signal of the filter with these parameters is given by the sum:

$$R^1(x_c, y_c, \alpha) = \sum_x \sum_y I(x, y) G_u(x, y; x_c, y_c, \alpha, \sigma) \quad (2)$$

In our experiments the walking direction could not be reliably extracted from the stimuli with ‘flat’ shading. These stimuli specify strong luminance gradients on the boundaries of the stimulus elements, but no shading gradients inside of them. The strong gradients on the element boundaries dominate the responses of the filters on the first hierarchy layer. In a set of additional simulations, we found that a more robust processing of the weak luminance gradients inside the elements can be accomplished by suppressing the gradient responses on the silhouette boundaries. This suppression is accomplished by the second layer of our model, exploiting a multiplicative gating mechanism. The underlying operation can be implemented by a simple feed-forward network that multiplies a gating signal with the filter responses from the previous layer. The gating signal is computed from all filter responses of the first layer with the same receptive field center, taking the maximum over the responses with selectivity for

different direction angles: $C(x, y) = \max_{\alpha} R^1(x, y, \alpha)$. The output signals of the second layer are then given by thresholded products of the form:

$$R^2(x, y, \alpha) = \left[H(\lambda_1 - C(x, y)) R^1(x, y, \alpha) \right]_+, \quad (3)$$

with the Heaviside function $H(x) = 1$ for $x > 0$ and $H(x) = 0$ otherwise, the positive constant λ_1 and the linear rectifier function $[x]_+ = \max(x, 0)$. The resulting output signals form a population code of the luminance gradients inside the individual stimulus elements. This layer contains a total of 6,480,000 model neurons.

The third layer of the model pools the responses with same direction selectivity from the second level over limited spatial regions $U(x, y)$ using a maximum operation, proving a (spatially subsampled) set of detector responses with partial position invariance. Mathematically these responses were given by:

$$R^3(x, y, \alpha) = \max_{(x', y') \in U(x, y)} R^1(x', y', \alpha), \quad (4)$$

The responses of the 648 neural detectors on this level can be interpreted as ‘mid-level features’ within the visual hierarchy. The recognition of walking direction was based on the classification of the activation patterns, exploiting a subset of these mid-level feature-detector responses, including only those detectors whose response varied significantly over the training set. In order to determine this set of detectors automatically, we computed, separately for each position (x, y) , a population vector from the detector responses with different direction specificity, which was given by the complex number:

$$p(x, y) = \sum_{\alpha} R^3(x, y, \alpha) \exp(i\alpha). \quad (5)$$

Exploiting circular statistics, we computed the variance of these population vectors (see (29) for details) and retained only the responses of those detectors whose circular variance over the whole training set exceeded a fixed threshold. This operation corresponds to a feature selection stage that identifies features with robust variation within the set of training patterns. Interestingly, the selected feature detectors correspond nicely to the body parts whose shading, according to Experiment 2, is most critical for the size of the illusion.

The highest level of the model (layer 4) is given by a classifier stage that is formed by two radial basis function (RBF) units that have been trained with example stimuli walking TOWARDS and AWAY, and which were illuminated from above. The underlying training set contained 100 movies of walkers that were lit with an elevation angle of 78.75 deg, where the veridical motion direction of 100 stimuli was AWAY and the one of the other 100 movies TOWARDS. The different movies were rendered using 4 different actors (2 male, 2 female) and randomly varying sizes of the body elements. This training

encodes a perceptual prior, which reflects statistical properties of stimulus patterns that are illuminated from above, and forms thus the implementation of the ‘lighting-from-above prior’ in our model.

In the model implementation, the basis function units were trained using a two-step algorithm, where this training was completely unsupervised without knowledge of the true class memberships (TOWARDS vs. AWAY) of the stimuli. First, the selected features were subject to a principle component analysis (PCA), over all training examples and stimulus frames, to further reduce the dimensionality of the feature vector. We retained only the first three principal components, explaining about 70% of the variance of the original feature vectors. This percentage was sufficient since the two walking directions were linearly separable in the corresponding reduced subspace. In order to test whether the chosen number of principle components influences the results critically, we also tested a model with 20 principal components, capturing about 90% of the variance, and obtained essentially the same model predictions (results not shown). Denoting by \mathbf{x} the output vector of the PCA, we fitted the distribution of these vectors over the whole training data set by a mixture of two Gaussians, which correspond in the neural implementation to the RBFs. The mixture distribution was fitted using Expectation Maximization (EM). The underlying mixture density is given by:

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{n=1}^2 \pi_n g(\mathbf{x}; \boldsymbol{\mu}_n, \sigma_n), \quad (6)$$

The EM procedure fits the mixture weights π_n (both very close to 0.5), and the means $\boldsymbol{\mu}_n$ and the covariance parameters σ_n of the Gaussian distributions, where we assume the Gaussian density functions:

$$g(\mathbf{x}; \boldsymbol{\mu}_n, \sigma_n) = (2\pi\sigma_n^2)^{-3/2} \exp\left(-\frac{|\mathbf{x} - \boldsymbol{\mu}_n|^2}{2\sigma_n^2}\right) \quad (7)$$

Interpreting the parameters π_n as the prior probabilities of the two classes ‘walking TOWARDS and ‘walking AWAY’ and the functions $g(\mathbf{x}; \boldsymbol{\mu}_n, \sigma_n)$ as the likelihoods of the vectors \mathbf{x} conditioned on the class C_n , according to the Bayes formula the posterior probability of the two classes C_n is given by the ratio:

$$p(C_n | \mathbf{x}) = \frac{\pi_n g(\mathbf{x}; \boldsymbol{\mu}_n, \sigma_n)}{\sum_{m=1}^2 \pi_m g(\mathbf{x}; \boldsymbol{\mu}_m, \sigma_m)} = \frac{g(\mathbf{x}; \boldsymbol{\mu}_n, \sigma_n)}{\sum_{m=1}^2 g(\mathbf{x}; \boldsymbol{\mu}_m, \sigma_m)}, \quad (8)$$

given that approximately $\pi_1 = \pi_2$. The posterior class probability can thus be computed by a simple normalization operation from the RBF neuron activities, and the most likely class can be found by simple winner-takes-all competition.

In order to compare the simulation results for Experiment 1 with the experimental data, we fitted the classification responses of the model with a logistic regression, and the experimental data with a logistic mixed-effects model (see section on statistical analysis of data from Experiment 1), using only the elevation angle as fixed and random effect predictor (smooth curves in Fig. 4A). Data was collapsed across the veridical walking directions since the model treats AWAY and TOWARDS walking equally and does not contain a special mechanism that can account for the walking towards bias that has been observed in biological motion vision (18). For Experiment 2, for each test stimulus, we computed the differences of the accuracies between the two light source positions (above and below) in order to quantify the size of the illusion. The similarity of the illusion sizes derived from the model with the real data for the different shading conditions was quantified by a linear regression analysis, predicting the illusion size obtained from the model with the ones in the experimental data (Fig. 4B).

Statistical analysis of data from Experiment 1

Statistical analysis was realized using R version 3.4, RStudio version 1.1.4, and lme4 version 1.1-12. We use ggplot2 version 2.2 to generate the statistical plots and multcomp version 1.4-6 for the multiple comparisons.

For the analysis of the conditions from Experiment 1 with graded shading of all elements we combined the responses of all 13 observers and fitted them with a logistic mixed-effects model (function `glmer` from package `lme4`) with cosine and sine of the light elevation angle and veridical walking direction both as fixed and random effects. In total, we fitted data points from 13 (observers) * 2 (veridical walking directions) * 17 (light angle directions) * 15 (repetitions) = 6630 data points. We did not include the interaction between the light elevation angle and veridical walking direction in the regression because this interaction was not significant. Fig. S1 shows the probability of perceiving the veridical walking direction as a function of light elevation angle in a separate panel for each observer (the panels are ordered by increasing difference in participant's points of subjective equality). The data points in each panel are jittered to minimize overlap (hence some probabilities seem smaller than zero or larger than one). Different colors indicate the two veridical walking directions, and the curves from the random effects fit are shown in the same color. The illusion is present for all observers and both veridical locomotion directions, while the fitted curves show different offsets (thresholds) that vary between participants, and partly between the two veridical locomotion directions.

A particularly interesting question is whether the condition with frontal illumination ($\alpha = 0$), which is the condition with the smallest luminance gradients within the different stimulus elements, conveys information about then veridical walking direction. The accuracy for this class of stimuli for the different observers is shown in Fig. S2. All participants show accuracies above 0.5, indicating that there is some remaining information about walking directions even for the stimuli with illumination from the side, potentially mediated by the subtle time-varying shading gradients in the stimulus elements that are present for this stimulus class. This stands in contrast with the stimuli with flat shading, that showed accuracies even below 0.5, indicating on average a perception of the wrong walking direction. A further set of statistical analyses investigated potential biases in the perception of walking direction in favor of walking TOWARDS or AWAY. Such biases were studied for the class of stimuli of all shaded walker elements as well as for the stimuli with flat shading.

To quantify potential biases for the stimuli with gradual shading of all elements we analyzed the points of subjective equality (PSEs) of the random effects from the model fitted above (as plotted in Fig. S1). As shown in Fig. S3A, only observer 10 has a bias to perceive 'away', while eight participants have a bias to perceive walking 'towards' them. Further four observers have little to no bias (observers 9, 5, 13 and 12). Thus, the bias in perceived walking direction is different for different observers, but most have a 'towards' bias, consistent with results on biological motion stimuli in the literature (18).

We investigated also potential biases for the perception of walking direction for the stimuli with flat shading. Response accuracies are shown in Fig. S3B, separately for the AWAY and TOWARDS conditions. Observers are ordered by the sizes of the differences between the PSEs between AWAY and TOWARDS walking. Observers 10, 9, 5, 13 and 12, have similar accuracy for the AWAY and TOWARDS walkers, whereas all others are more accurate for the TOWARDS condition.

There is a striking analogy between the two panels, in that observers with a larger difference in PSE's while viewing the fully shaded stimuli also tend to have a larger difference in accuracy between AWAY and TOWARDS walkers while viewing stimuli with flat shading. To make this relationship explicit, we plotted the difference in accuracy for the stimuli with flat shading against the difference in PSE derived from the fully shaded stimuli in Fig. S3C. Testing the correlation between both variables, by an ordinary linear regression analysis, we found the slope of the regression line to be significantly different from zero ($p < 0.01$). This shows that the observer-specific bias in perceiving an ambiguous walker as walking towards the observer is similar for the two different stimulus classes: the fully shaded stimuli and the ones with flat shading.

Statistical analysis of data from Experiment 2

We used the same software for statistical analysis as for Experiment 1. The combined responses of all 16 observers were fitted with a linear mixed-effects model (function `lmer` from package `lme4` (19)) with body part and veridical walking direction both as fixed and random effects. In total, we fitted 16 (observers) * 2 (veridical walking directions) * 9 (body part conditions) = 288 data points, each of which was defined by the difference in accuracy between 15 repetitions with lighting from above and 15 repetitions with lighting from below. The data was collected testing 9 levels of the factor body part, while only 8 levels were used in the main analysis, since the results for the levels “none” (flat shading) and “body” (gradual shading of head, torso and upper arms only) were not statistically different and hence combined. Fig. S4 shows all of these 288 means, with the panels ordered by increasing average accuracy difference. The trend of increasing mean accuracy with body part condition holds for most observers, although there are also clear inter-individual differences. Interestingly, there are only minor differences between the stimuli with different veridical walking directions.

In a linear mixed-effects regression analysis we found a significant effect of factor body part ($p < 10^{-15}$) and an insignificant effect of veridical walking direction ($p > 0.05$). Subsequently, we realized also selected pairwise comparisons with corrections for multiple testing (R package `multcomp` (20)). First, we first asked whether adding the forearms, the thighs or the (lower) legs increased the mean difference in accuracy. We found that adding each of these elements significantly increased (all $p < 0.001$) the illusion relative to the baseline condition with gradual shading of the head, torso and upper arms. Second, we compared which of the three additions, forearms, thighs or (lower) legs lead to a larger illusion and found no significant differences between these three conditions (all $p > 0.5$). Third, we asked if the addition of a second moving limb region to the first increased size of the illusion. In detail, we asked whether adding (lower) legs or thighs to the forearms increased the illusion and found that it did (all $p < 0.01$). Adding the forearms to the (lower) legs or thighs did not increase the illusion (all $p > 0.1$). Adding the (lower) legs to the thighs or the thighs to the (lower) legs did increase the illusion by a modest amount ($p < 0.05$). Fourth, we asked if adding a third moving limb region to any other two moving limb regions resulted in a significant increase of performance. Only adding of the thighs to the forearms and (lower) legs resulted in a significant difference ($p < 0.05$), while this was not the case for the addition of the other two moving regions ($p > 0.2$).

Rendering of stimuli for Control Experiment

In a control experiment we used a stimulus that does not provide classical shape-from-shading cues that allow the reconstruction of the three-dimensional orientation of the body segments. However, the elements of this control stimulus approximate the internal luminance profiles of the elements of the original stimuli used in Experiment 1. The purpose of this experiment was to rule out the possibility that the observed

illusion just is a consequence of a classical shape-from-shading mechanism that estimates the stimulus element orientations in space.

The walker for this control experiment was composed of elements with fixed circular shape in order to eliminate silhouette-based orientation cues. The internal luminance profiles of these circular elements were determined from the luminance patterns of the original stimuli in Experiment 1 by warping of the shading profiles of the conic elements onto these circular patches. For this purpose, each conic element was rendered separately on a black background. Using a Sobel edge detection algorithm, we extracted the coordinates of the boundary points of the elements. For each element we computed the center of mass (X_c, Y_c) and represented the boundary points (X_b, Y_b) in terms of polar coordinates relative to this center, i.e. $(X_b - X_c, Y_b - Y_c) = R(\phi) \cdot (\cos \phi, \sin \phi)$. To each boundary point on the conic elements we assigned a corresponding boundary point (X_b^*, Y_b^*) on the circular element according to the equation $(X_b^* - X_c, Y_b^* - Y_c) = R^* \cdot (\cos \phi, \sin \phi)$, where the corresponding radius R^* was given by the maximal radius of all boundary points of the conic element. This construction assigns to each point inside the conic element with polar coordinates $(\lambda R, \phi)$ a corresponding point $(\lambda R^*, \phi)$ in the circular element with $0 \leq \lambda \leq 1$. This implies for the corresponding points within the circular element the coordinates $X_b^* = X_c + R^* \cdot \cos \phi$ and $Y_b^* = Y_c + R^* \cdot \sin \phi$. Exploiting these corresponding coordinate systems, we warped the luminance profiles of the original conic elements on the ones with circular shape using nearest neighbor interpolation. Fig. S5A illustrates example frames from the generated control stimuli. The circular elements look like deforming rubber sheets and do not provide a clear impression about orientation in space.

Control Experiment

In order to assess the illusion size for the control stimulus that prevents the use of shape-from-shading for the estimation of element orientations in space, we used only three light source positions with illumination from ABOVE and three from BELOW. Like in Experiments 1 and 2, we asked the participants to report the perceived walking direction. A new group of 12 observers participated in this control experiment. The experimental protocol was identical to experiment 1 and 2, except for the fact that performed 8 instead of 4 steps (4 gait cycles instead of 2) with 20 repetitions per condition.

The results were analyzed using the same procedure as in experiments 1 and 2. We combined the responses of all 12 observers and fitted them with a logistic mixed-effects model with light elevation angle and veridical walking direction both as fixed and random effects. In total, we fitted 12 (observers) * 2 (veridical walking directions) * 6 (light angle directions) * 20 (repetitions) = 2880 data points.

The data (Fig. S5B) shows a weaker, but significant effect of the light source position on the accuracy of perceived walking direction ($p=0.02$). This shows that the illusion was present even for stimuli which make a reconstruction of the three-dimensional orientation of individual limb segments impossible. However, taking the holistic body configuration and the structure of the inner luminance gradients into account, the visual system was able to extract the walking direction from such stimuli and shows the same illusion as for the stimuli with conic elements in Experiment 1. The observed illusion is thus not just a consequence of a classical shape-from-shading process that estimates the orientation of individual stimulus elements. Instead, the observed phenomenon must be based on a mechanism that integrates body shape and information about internal shading gradients, as for example the mechanism realized by the proposed model.

Supplementary Figure Captions

Fig. S1. Per-participant data from Experiment 1. Accuracy of reporting the veridical walking direction as a function of the light source elevation angle α . The participants are ordered by increasing difference between their AWAY and TOWARDS points of subjective equality.

Fig. S2. Accuracy of reporting veridical walking direction of the frontal lighting condition ($\alpha = 0$) for the fully shaded walkers. The participants are ordered by increasing difference between their AWAY and TOWARDS points of subjective equality (same as Fig. S1).

Fig. S3. Towards walking bias derived from conditions with and without gradual shading. (A) Towards walking bias derived from the conditions with full gradual shading. The figure shows the points of subjective equality (PSE) per participant, estimated from the random effects component of the GLMM fitted to the data of Experiment 1, using the conditions with gradual shading of all elements. (B) Accuracy of reporting the veridical walking direction for the stimuli with flat shading. (Participants are ordered by increasing differences of the PSEs between their AWAY and TOWARDS walking.) (C) Correlation between participant accuracies for stimuli with flat shading (panel (B)) and the PSEs derived from stimuli with full graded shading (panel (A)) (adjusted $R^2 = 0.4831$; $p < 0.01$).

Fig. S4. Per-participant accuracies of reporting the veridical walking direction as a function of the body part condition of Experiment 2. Participants are ordered by increasing difference in their accuracies, averaged over body part conditions and veridical walking directions.

Fig. S5. Control Experiment. (A) Snapshots from the control experiment with left: one of the ABOVE light elevation angles and right: one of the BELOW light elevation angles, both walking AWAY. Shading of each individual element of the walker figure was stretched from original conic shape (used in experiments 1 and 2) to a circular boundary via interpolation. Supporting Information (SI) Movie 3 shows these 2 walker stimuli. (B) Results. Walkers illuminated from 3 ABOVE and 3 BELOW light angles were presented. Plotted points represent the means of the veridical binary responses per condition. The data were fitted using a generalized linear mixed effects model (GLMM).

Supplementary Movie Captions

Movie 1. One of the stimuli used for Experiment 1. The walker is walking TOWARDS and is lit below with - 90 deg elevation angle.

Movie 2. One of the stimuli used for Experiment 1. The walker is walking TOWARDS and is lit above with 90 deg elevation angle.

Movie 3. One of the stimuli used for Experiment 1. The walker is walking TOWARDS and is lit in front with 0 deg elevation angle.

Movie 4. One of the stimuli used for Experiment 1. The walker is walking TOWARDS and has flat shading.

Movie 5. One of the stimuli used for Experiment 2. The walker is walking AWAY, lit BELOW and has it's thighs and legs flattened (FOREARMS condition).

Movie 6. One of the stimuli used for Experiment 2. The walker is walking AWAY, lit BELOW and has it's forearms and thighs flattened (LEGS condition).

Movie 7. One of the stimuli used for Control Experiment. The walker is walking AWAY and is above with 67.5 deg elevation angle.

Movie 8. One of the stimuli used for Control Experiment. The walker is walking AWAY and is below with - 22.5 deg elevation angle.

Movie 9. Demo video to show two walking figures next to each other. Both walkers are rendered to walk AWAY (the camera is positioned looking at the back of the walking figure). The left walker has light source positioned ABOVE, and the right walker has the light source positioned BELOW. Both walkers are looped for multiple gait cycles. Try judging the walking direction of each of them separately. Most observers would judge the left one as walking AWAY (correctly), but the right one as walking TOWARDS (opposite).

Supplementary Table 1

Table 1. Parameter values used to simulate the model with experimental stimuli.		
Parameter	Value	Meaning
α	{0°,45°,90°,135°,180°,225°,270°,315°}	Orientation of maximal direction sensitivity of neurons in Layer 1 modeled as an angle value of the 2D uneven Gabor filter.
σ	60	Gabor filter scale in Layer 1.
k_0	40^{-1}	Spatial filter frequency in Layer 1.
λ_1	25	Excitation threshold in Layer 2 exceeding which activates the gating feedback inhibitory mechanism for boundary suppression.
m	81	Number of partially-position invariant detectors at Layer 3, arranged in overlapping square grid with large spatial receptive field sizes.
T	30	Number of keyframes taken in every individual walker movie.
$X \times Y$	900×900	Resolution of a single keyframe.
$U(x, y)$	300×300	Size of spatial max-pooling region in Layer 3.

Supplementary Dataset Captions

Dataset 1. Collected participant data for Experiment 1, conditions with shaded walker.

Columns indicate, in order: participant ID, light elevation angle in degrees, walking direction, participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not), participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not) in the previous trial.

Dataset 2. Collected participant data for Experiment 1, single condition with flat walker.

Columns indicate, in order: participant ID, light elevation angle set to 'Inf' indicating no shading present, walking direction, participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not), participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not) in the previous trial.

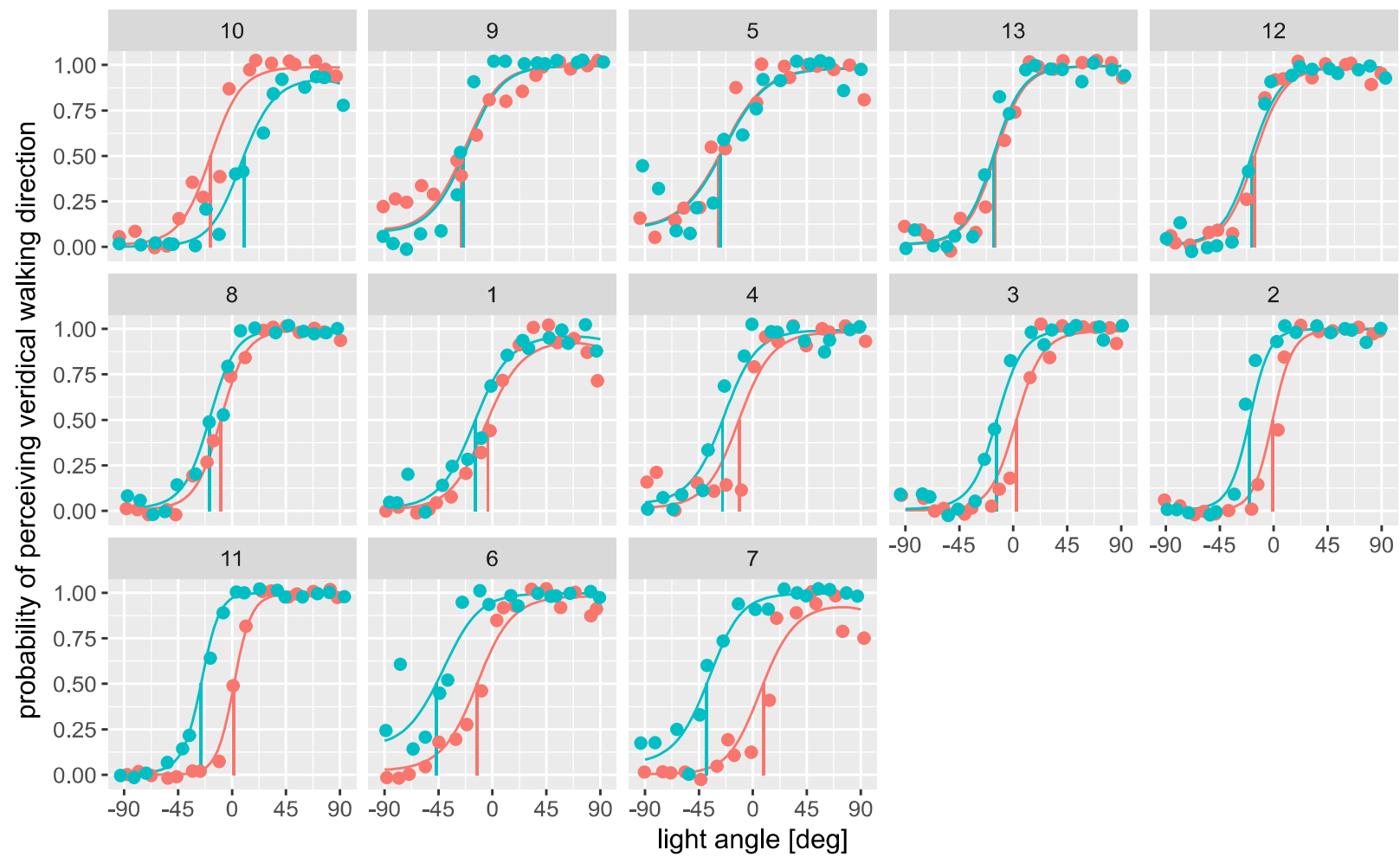
Dataset 3. Collected participant data for Experiment 2.

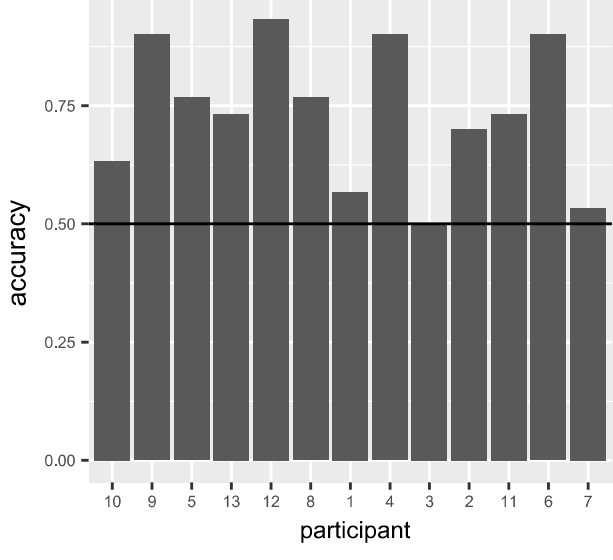
Columns indicate, in order: participant ID, light elevation angle in degrees, walking direction, participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not), body part condition name, recoded light source position (same as light elevation angle, coded ABOVE if 45, BELOW if -45), participant's perceived walking direction.

Dataset 4. Collected participant data for Control Experiment.

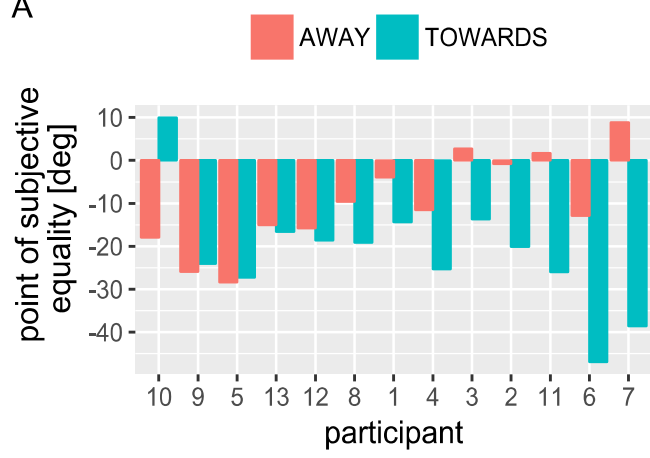
Columns indicate, in order: participant ID, light elevation angle in degrees, walking direction, participant's binary accuracy in indicating the walking direction (1 if correct, 0 if not).

AWAY TOWARDS

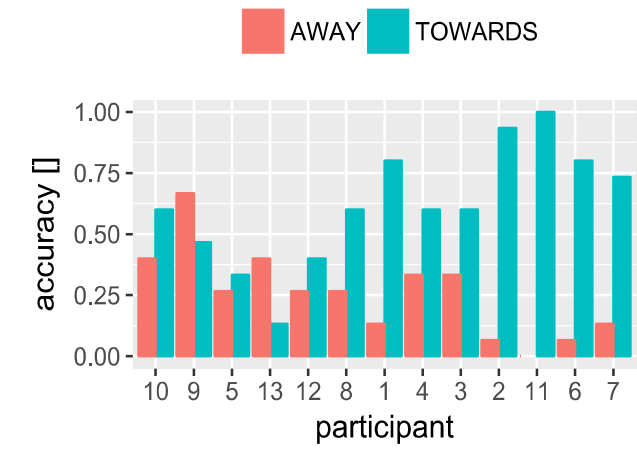




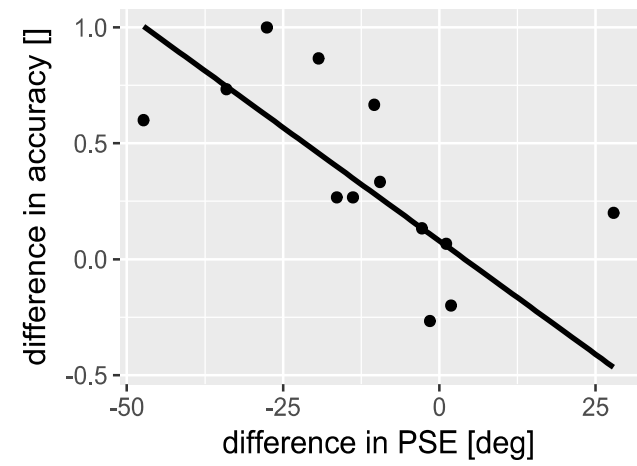
A



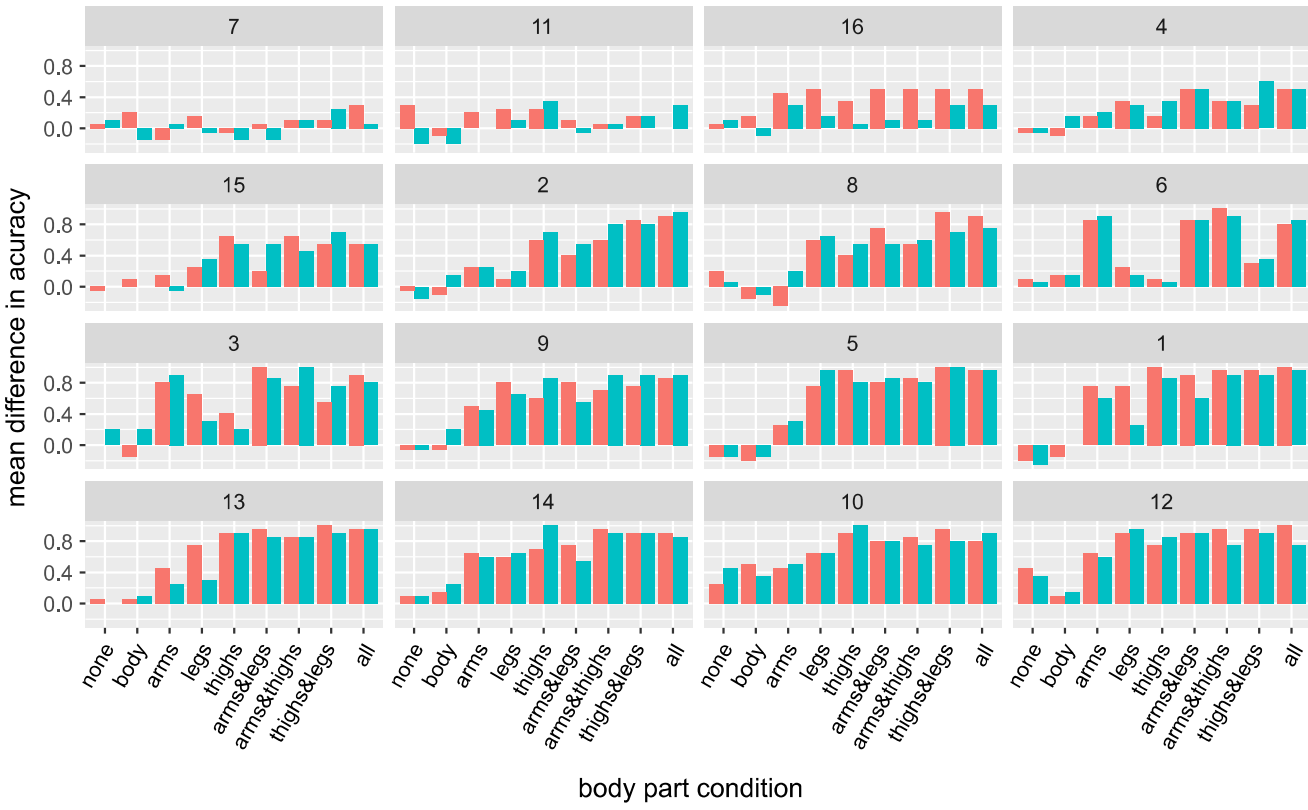
B

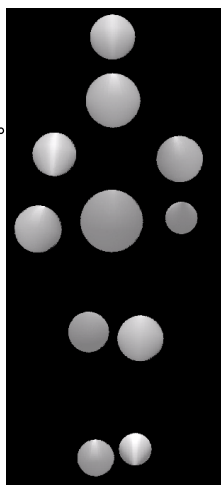
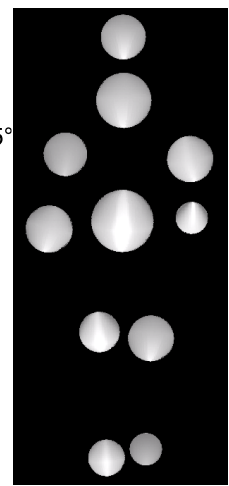
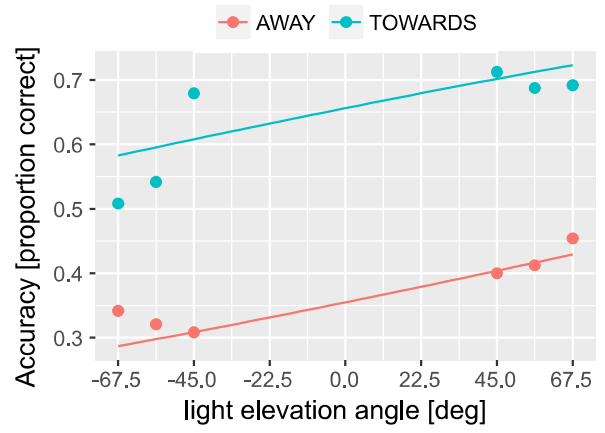


C



AWAY TOWARDS



AAWAY
ABOVE
 $\alpha = 67.5^\circ$ AWAY
BELOW
 $\alpha = -22.5^\circ$ **B**

Chapter 3. Neural model for the influence of shading on the multistability of the perception of body motion.

(published as

L Fedorov, J Vangeneugden, M Giese (2016). Neural model for the influence of shading on the multistability of the perception of body motion. NCTA Neural Computation Theory and Applications. Conf. Proceedings: 1-8.;

received Best Student Paper award)

Neural Model for the Influence of Shading on the Multistability of the Perception of Body Motion

Leonid Fedorov¹², Joris Vangeneugden³ and Martin Giese¹²

¹*Dept. of Cognitive Neurology, CIN, HIH, University Clinic Tuebingen, Tuebingen, Germany*

²*IMPRS for Cognitive and Systems Neuroscience, University of Tuebingen, Tuebingen, Germany*

³*School of Mental Health and Neuroscience, Maastricht, The Netherlands*

{leonid.fedorov, martin.giese}@uni-tuebingen.de; joris.vangeneugden@gmail.com

Keywords: Action Recognition, Multistable Perception, Biological Motion, Neural Fields, Shading

Abstract: Body motion perception from impoverished stimuli shows interesting dynamic properties, such as multistability and spontaneous perceptual switching. Psychophysical experiments show that such multistability disappears when the stimulus includes also shading cues along the body surface. Classical neural models for body motion perception have not addressed perceptual multistability. We present an extension of a classical neurodynamic model for biological and body motion perception that accounts for perceptual switching, and its dependence on shading cues on the body surface. We demonstrate that a set of psychophysical observations can be accounted for in a unifying manner by a hierarchical neural model for body motion processing that includes an additional shading pathway, which processes luminance gradients within the individual body segments. The goal of our model is to explain psychophysics and neural mechanism in the brain.

1 INTRODUCTION

The perception of body motion from image sequences requires the dynamic integration of complex spatio-temporal visual patterns. This important visual function is accomplished by processing within a hierarchy of cortical areas along the visual pathway. Psychophysical studies suggest depth cues are important for biological motion perception (Jackson and Blake, 2010). In absence of such depth information, e.g. in point-light walkers, body motion perception can become multistable (Vanrie and Verfaillie, 2004). Then the same stimulus can be perceived as alternating randomly between two interpretations that correspond to two different walking directions (Vanrie and Verfaillie, 2006). Multistable phenomena has been also investigated in the context of static ambiguous figures and binocular rivalry (Leopold and Logothetis, 1999), (Blake and Logothetis, 2001), as well as in structure from motion (Andersen and Bradley, 1998). An example of the body motion stimulus that produces such multistability is shown in Fig. 1A (panel SILHOUETTE). For this stimulus, an articulating silhouette without intrinsic shading cues, observers perceive the walker alternately walking obliquely into or out of the image plane. The two reported percepts correspond to the unambiguous walking directions indicated in pan-

els TOWARDS and AWAY. The figure illustrates also that this perceptual ambiguity disappears when shading gradients are added to the surface of the walker, which provide information about the surface orientation of the body segments and occlusions.

Existing physiologically-inspired neural models for the processing of body motion and goal-directed actions (e.g. (Giese and Poggio, 2003), (Lange and Lappe, 2006), (Escobar and Kornprobst, 2008), (Jhuang et al., 2007), (Fleischer et al., 2013) and (Layher et al., 2014)) do not reproduce such multistability, or at least never have investigated this phenomenon. Computer vision and deep learning architectures for body motion recognition do not address perceptual multistability. Thus, the study of such phenomena is important for neuroscience, even if such multistability is often unwanted in technical action recognition systems.

In the context of low-level vision, perceptual multi-stability and the underlying neural dynamics have been extensively studied e.g. in the context of binocular rivalry (see e.g. (Wilson, 2003)), visual motion integration (Rankin et al., 2014), or as general property of attractor neural networks (Pastukhov et al., 2013).

The goal of this paper is to extend existing physiologically-inspired neural models (not computer

vision algorithms) in a way that accounts for multistability in action perception, where we use as example an established model that has been shown to account jointly for many experimental results in this area (Giese and Poggio, 2003). We extend it in two ways: 1) by introduction of a multi-dimensional neural field that accounts for multi-stable behavior by lateral interactions between shape-selective neurons; 2) by addition of a new pathway that realizes robust processing of intrinsic luminance gradients along the surface of the body segments.

The paper is structured as follows: after discussing related work in the following section, we describe the developed architecture in section 3. In section 4 we show simulation results, illustrating that the model provides a unifying account for several key psychophysical results, followed by a brief discussion in section 5.

2 RELATED THEORETICAL WORK

Body motion recognition has been a core topic in computer vision and many technical neural architectures for this purpose have been proposed (Edwards et al., 2016), (Nguyen et al., 2016), (Ziaeeafard and Bergevin, 2015), (Lee et al., 2014). The goal of that work is typically a maximization of recognition performance, not a reproduction of perceptual dynamics of humans. This paper does not contribute to computer vision or machine learning and is entirely focused on modeling of the brain.

We follow the approach in physiologically-plausible models of body motion perception, such as (Giese and Poggio, 2003), (Lange and Lappe, 2006), (Escobar and Kornprobst, 2008), (Fleischer et al., 2013), (Layher et al., 2014), while other biological models in this area (e.g. (Thurman and Lu, 2014) (Thurman and Lu, 2016)) account for experimental data without direct relationship to neural mechanisms.

Diverse approaches (see (Tyler, 2011)) have been proposed for the analysis of shape from shading, but typically not related to the processing of body motion. Perceptual dynamics and perceptual switching have been extensively studied in the context of low-level vision (reviews see e.g. (Leopold and Logothetis, 1999), (Sterzer et al., 2009), (Pastukhov et al., 2013)). Multistability in the processing of non-rigid motion has been rarely studied in neural modeling.

While hierarchical technical algorithms in computer vision typically focus on the problem how the body motion patterns (e.g. the direction of body movement) might be distinguished, our model tries

to unify this account with a reproduction of the dynamics of perceptual organization in humans which emerges specifically for the SILHOUETTE stimulus, where for the same stimulus two alternating percepts emerge. This problem is typically not addressed in technical recognition systems, and to our knowledge no account for this phenomenon has been given in biologically-inspired neural models for motion recognition.

3 MODEL ARCHITECTURE

Our model builds on a previous neural model (Giese and Poggio, 2003), which has been shown to provide a unifying account for a variety of experimentally observed phenomena in body motion perception including physiological, psychophysical and fMRI data. The original model included a motion and a form pathway, processing shape and optic flow features. The pathways consist of a hierarchy of feature detectors that mimic properties of real cortical neurons. For the implementation in this paper we used only the form-pathway and extended it by a multi-dimensional neural field, and a new pathway for the processing of intrinsic luminance gradients. An extension by inclusion of an additional motion pathway is straight-forward, and will be part of future work.

3.1 Silhouette Pathway

The backbone of our model is a 'silhouette pathway' (Fig. 1B) that is identical to the the form pathway of the classical model (Giese and Poggio, 2003). Due to space limitations, we sketch here only some basics about this pathway and refer to the original publication (Giese and Poggio, 2003) with respect to details. In brief, the form pathway consists of a hierarchy of layers that process form features of increasing complexity along the hierarchy. More complex features are formed by combination of the features from previous layers. Levels that increase feature complexity are interleaved by layers that increase position and scale invariance by MAX pooling. The highest level of this shape processing hierarchy is formed by radial basis function units (called 'snapshot neurons') that have been trained with the feature vectors that correspond to keyframes from training movies showing the recognized action. Each snapshot neuron responds selectively to the body posture that corresponds to time instance θ (within the gait cycle). In addition, consistent with physiological data (Vangeneugden et al., 2011), we assume these neurons are view-specific, where the variable ϕ specifies the preferred view an-

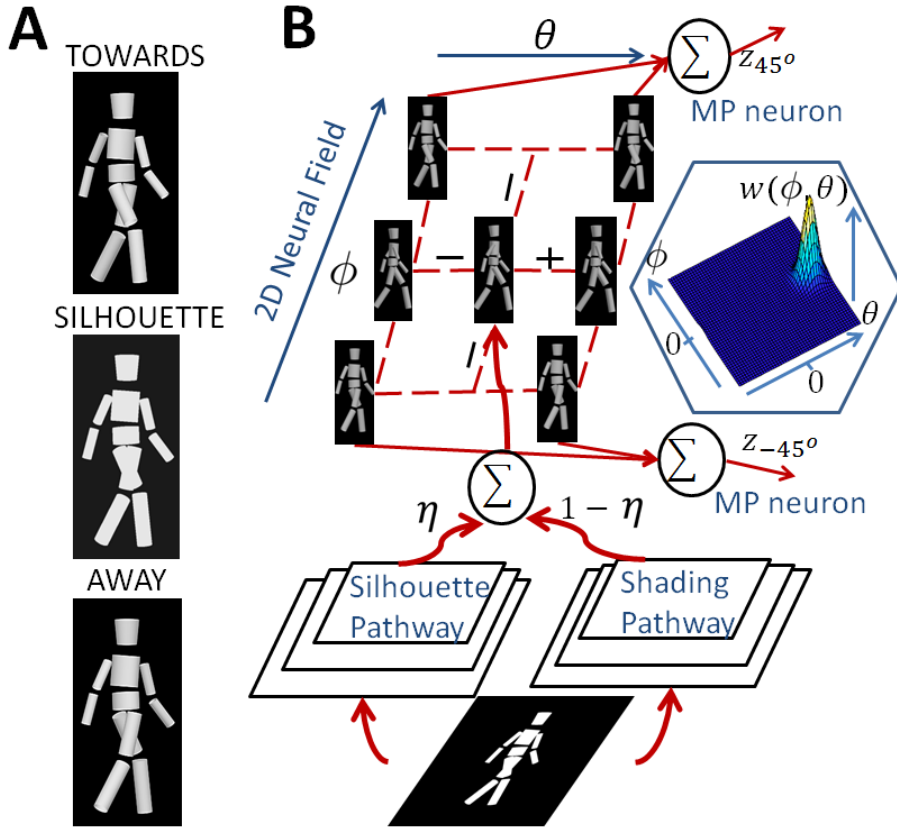


Figure 1: A. Snapshots from movies showing dynamic walker: TOWARDS shaded walker, walking direction 45 deg; SILHOUETTE bistable silhouette walker and AWAY shaded walker, walking direction -45 deg. B. Model architecture. Stimulus is analyzed by Silhouette and Shading pathways. Their outputs are linearly combined and mapped linearly onto the input of a 2D dynamic neural field that consists of laterally coupled snapshot neurons. Inset shows the lateral interaction kernel of the field. The field activity is read out by Motion Pattern (MP) neurons that encode the perceived walking directions ± 45 deg.

gle of the neuron. (We assume that the side view of a walker walking to the right in the image plane defines the view direction $\phi = 0$). Very similar architectures underlie many other classical and modern neural and deep models for object recognition, where the popular deep architectures are typically trained with much more data and often include many more layers. Since the goal of this paper is to model the perceptual dynamics, and not to maximize recognition rate, we used this simple hierarchical model, where extension with modern deep architectures as front-end seem straight-forward.

3.2 Shading Pathway

The described simple form pathway recognizes body shape on backgrounds with sufficient contrast. However, it turned out that with small amounts of training data it is difficult to accomplish with this architecture a robust recognition of the silhouette shape together

with a high sensitivity for the luminance shading gradients that disambiguate the depth structure. As one possible solution to this problem we implemented a second pathway that is specialized for the processing of intrinsic shading gradients using physiologically-plausible operations (Fig. 1B). We do not claim this is the only possible solution, but it is one that works with small amounts of training data.

The first level of this new pathway overlaps with the first hierarchy level of the silhouette pathway, described above. It consists of Gabor filters that are selective for local orientation features at different positions, and for different spatial scales. Let $G_{e,u}(x, y, \alpha, \sigma)$ signify the output signal of the even (e) or uneven (u) Gabor filter with preferred position (x, y) , preferred orientation α (we used 8 orientations), and scale σ (we used 1 scale for the given small stimuli set). The activations of the uneven Gabor filters provide a population code for the local luminance gradients.

By pooling of the responses of the Gabor filters with the same preferred position over all orientations we obtain position-specific detectors for contours with the output signals:

$$C(x, y) = \max_{\{e, u\}, \alpha, \sigma} |G_{e, u}(x, y, \alpha, \sigma)|. \quad (1)$$

This output signal was used to suppress the responses of the uneven Gabor filters along the external contour of the body, exploiting multiplicative gating. The outer contour of the body typically creates strong local contrast that dominates the detector responses, so that the weak intrinsic gradients that signal the 3D structure cannot be reliably estimated from the neural responses. A population vector signaling the intrinsic luminescence gradients is given by the gated signal:

$$L(x, y, \alpha, \sigma) = [G_u(x, y, \alpha, \sigma) \cdot H(\lambda_1 - C(x, y))]_+. \quad (2)$$

Here λ_1 is a positive constant, and the function $H(x)$ is the Heaviside function, thus $H(x) = 1$ for $x > 0$ and $H(x) = 0$ otherwise.

The next level of the shading pathway consists of (partially) position-invariant detectors for local luminance gradients. Their responses are computed by pooling of the gated responses of gradient detectors for the same preferred gradient direction α over all positions and scales in a quadratic neighborhood $\mathcal{U}(x', y')$ of the point (x', y') using a maximum operation, providing the output signals:

$$D(x, y, \alpha) = \max_{(x', y') \in \mathcal{U}(x', y'), \sigma} L(x', y', \alpha, \sigma). \quad (3)$$

These position-invariant detectors were defined for substantially less spatial positions, resulting in a strong spatial down-sampling (6,480,000 position- and scale-specific detectors vs. 648 position-invariant detector units).

In order to make recognition robust against fluctuating weak features, we selected the strongest features that provide input to the radial basis function units. We selected those features that showed the maximum variance over the training data (where clearly much more sophisticated feature selections are available that might lead to better results). We computed the circular variance of the detectors at position (x, y) , exploiting the (complex) circular mean:

The (complex) circular mean of these responses is given by:

$$m(x, y) = (1/K) \sum_{k=1}^K \sum_{\alpha} D^{(k)}(x, y, \alpha) \exp(i\alpha), \quad (4)$$

where K is the number of training patterns. A circular variance measure is then given by the formula:

$$V(x, y) = \sum_{k=1}^K \left| \sum_{\alpha} D^{(k)}(x, y, \alpha) \exp(i\alpha) - m(x, y) \right|. \quad (5)$$

We selected the direction-specific responses $D(x, y, \alpha)$ that fulfilled the relationship:

$$V(x, y) > \lambda_2, \quad (6)$$

where $\lambda_2 > 0$ is a threshold parameter. In total 9 out of 81 feature vectors were selected according to this criterion.

The next level of the shading pathway is formed by Gaussian radial basis functions, whose centers were trained with the feature vectors \mathbf{p}^l (including only the selected features) that were generated by individual keyframes from the training movies. For the results shown here, the shading pathway was trained with movies of fully shaded walkers, shown with view directions -45 deg and 45 deg. In other implementations, we have realized such models with a continuum of different views (Fleischer et al., 2013).

The RBF network returns an 50-dimensional output vector $\mathbf{R}_{SH}(t)$ for each keyframe at time t , where the components of this vector are given by:

$$R_{SH}^l(t) = \exp(-\lambda_3 \|\mathbf{p}(t) - \mathbf{p}^l\|^2), \quad (7)$$

where $\mathbf{p}(t)$ is the feature vector for the actual input frame, and where the components correspond to the different keyframes and associated training views.

In order to link the shape recognition pathway to dynamic neurons that reproduce the perceptual dynamics, the outputs of the RBF units were mapped linearly onto a discretely sampled two-dimensional input activity distribution $s_{SH}(\theta, \phi; t)$ that provides input to the neural field that is described below. Signifying by $s_{SH}(t)$ the appropriately reordered sampling points, the linear mapping was given by the equation:

$$\mathbf{s}_{SH}(t) = \mathbf{W}(t) \mathbf{R}_{SH}(t). \quad (8)$$

The weight matrices $\mathbf{W}(s)$ were learned by ridge regression from a training set that consisted of pairs of vectors $\mathbf{R}_{SH}(t)$ for each training keyframe, and a corresponding vector $\mathbf{s}_{SH}(t)$ that was computed from an idealized two-dimensional input activity distribution $s_{SH}(\theta, \phi; t)$. The idealized activity distribution was given by a Gaussian peak that was centered at the keyframe number θ and the corresponding view ϕ of the walker (s.b.). A similar input distribution $s_{SL}(\theta, \phi; t)$ was computed by a corresponding linear mapping in the silhouette pathway. The total input distribution of the neural field was then computed by 'cue fusion', modeled by a convex combination

of two input distribution functions according to the equation:

$$s(\theta, \phi; t) = \eta s_{\text{SL}}(\theta, \phi; t) + (1 - \eta) s_{\text{SH}}(\theta, \phi; t), \quad (9)$$

with $0 \leq \eta \leq 1$. Choosing $\eta = 1$ one can eliminate the influence of the shading pathway.

3.3 Dynamic Neural Field of Snapshot Neurons

The core of our model is a dynamic recognition layer that is implemented as a two-dimensional neural field of Amari type (Amari, 1977), which consists of body shape-selective neurons that are laterally connected (Fig. 1B). Consistent with physiological data (Vangeneugden et al., 2011), we assume that such neurons encode body shapes that emerge during actions in a view-specific manner. In the spatial continuum limit, we can describe the activity of neurons encoding the body shape that corresponds to the normalized time θ ($0 \leq \theta \leq 2\pi$) during the gait cycle and the view angle ϕ by the function $u(\phi, \theta, t)$. The network dynamics is given by the equation (\star signifying a spatial convolution):

$$\begin{aligned} \tau_u \frac{d}{dt} u(\phi, \theta, t) = & -u(\phi, \theta, t) + w(\phi, \theta) \star H(u(\phi, \theta, t)) \\ & + s(\phi, \theta, t) - h + \xi(\phi, \theta, t) - c_a a(\phi, \theta, t). \end{aligned} \quad (10)$$

The input signal s was described above. For the trained stimulus movies it corresponds to an activity maximum that moves in θ -direction along the field. The lateral connectivity is specified by the interaction kernel $w(\phi, \theta)$ (whose shape is indicated by the inset in Fig. 1B). It stabilizes a traveling pulse solution in θ -direction and realizes a winner-takes-all competition in the ϕ -direction. As consequence, if multiple views are consistent with the stimulus, one view is selected by competition. The positive parameters τ_u and h define the time scale and the resting potential of the field. The variable $\xi(\phi, \theta, t)$ defines a Gaussian noise process whose statistics was coarsely adapted to the noise correlations from cortical data (Giese, 2014). These fluctuations essentially drive the perceptual switching in the model. Since action perception shows adaptive properties, such as high-level after-effects and fMRI adaptation, we also included a neural adaptation process in the model, which reduces the activity of snapshot neurons after extended firing. The corresponding adaptation variable follows

the dynamical equation:

$$\tau_a \frac{d}{dt} a(\phi, \theta, t) = -a(\phi, \theta, t) + H(u(\phi, \theta, t)). \quad (11)$$

The positive constant c_a determines the strength of adaptation (τ_a is the time constant). The parameters of this adaptation dynamics were fitted to experimental data (Giese, 2014).

The activity of the neurons in the neural field was read out by motion pattern (MP) neurons, which signal the walking directions perceived in this case as AWAY from and TOWARDS the observer. These neurons compute the maximum of the neural field activity function $u(\phi, \theta, t)$ over the domains $\phi > 0$ and $\phi < 0$ in the (ϕ, θ) space, producing the output signals z_{45} and z_{-45} .

4 SIMULATION RESULTS

Testing the model after training with a non-shaded walker as illustrated in Fig. 1A, 1B and 1C, the output of the shading pathway remained silent because of the absence of intrinsic luminance gradients in this stimulus. The silhouette pathway was activated in an ambiguous way by this stimulus because the stimulus is consistent with walking in the directions ± 45 deg relative to the image plane. Consistent with simulations described in (Giese, 2014), this stimulus leads to a bistable solution of the neural field that alternates between two traveling pulse solutions that encode the spontaneous perceptual switching of a traveling pulse between the view angles $\phi = \pm 45$ deg (perception of TOWARDS or AWAY from the observer). In this case, the probabilities of the two percepts are almost identical (Fig. 2B). More detailed simulations show that the model coarsely reproduces also the switching time statistics of human perception, comparing it with experimental data (not yet published (Vangeneugden et al., 2012)). Fig. 2G shows a histogram of the percept times for the model, and Fig. 2H the percept times estimated in the psychophysical experiment.

For shaded stimuli (see Figs. 1A TOWARDS and AWAY), when both pathways are included ($\eta = 0.5$), the model successfully disambiguates the walking direction: For the AWAY stimulus (direction -45 deg) the output neuron for AWAY remains always activated while the output neuron for TOWARDS remains silent. If an TOWARDS stimulus is shown (direction 45 deg) the situation is reverse and the TOWARDS output neuron is always active (Fig. 2C and D).

If however the shading pathway is deactivated ($\eta = 0$) again perceptual switching occurs, since the

output of the silhouette pathway is ambiguous, resulting in equal percept probabilities for either direction. The silhouette pathway is not sufficiently sensitive to disambiguate the stimulus robustly based on the available luminance gradients intrinsic to the body segments (Fig.2 E-F). This demonstrates the necessity of the shading pathway in the chosen architecture for the disambiguation of the percept.

The model makes several verifiable experimental predictions in relation to the time course of the adaptation process. An example is illustrated in Fig. 1I that shows a diagram of a typical adaptation experiment to demonstrate after-effects in action perception. First, an unambiguous adaptation stimulus (TOWARDS or AWAY) is presented to participants, where the duration of the adaptor (2, 6, 10, 14, 18 or 22 gait cycles) was varied over different blocks of the experiment. After this stimulus (and a fixed Inter-stimulus Interval of 2.8 s) an ambiguous test stimulus (SILHOUETTE) is presented for 3 gait cycles, asking for the perceived walking direction.

The predicted results for such an experiment (from 20 repeated simulations) are presented in Fig. 1J, which shows the probabilities of the percept for the ambiguous test stimulus (which was identical in all cases). With increasing the duration of the adaptor stimulus the probability that participants perceive the test stimulus as walking in the same direction as the adaptor decreases. A significant decrease of the percept probability (from 0.5 without adaptor presentation) is already perceived for the shortest adaptor duration of 2 gait cycles, and we observed a further decrease with longer adaptor durations (where 1 gait cycle corresponds to 1.4 seconds of stimulus duration).

This behavior is consistent with after-effects, as investigated previously for many modalities (motion, lightness, etc) in low-level vision. Such after-effects for action perception with a similar time course have been shown for other types of action stimuli in the literature (see (Barraclough and Jellema, 2011), (de la Rosa et al., 2014)), and we are presently running psychophysical experiments to verify this prediction of the model in detail.

A further set of experiments that we are presently running, and for which the model provides quantitative predictions, investigates the interdependence of the stability of action percepts and the switching times between the different percepts (which depend on the mean-first passage times of the corresponding attractors). This extends studies that have been made for multi-stability of low-level motion perception (Hock et al., 1993) to the domain of action perception.

5 CONCLUSIONS

To our knowledge, we have described the first biologically-inspired neural model that accounts simultaneously for the following properties of body motion perception: (i) perceptual multi-stability and switching, (ii) switching time statistics and (iii) the influence of shading information on the perceptual dynamics. We showed that the model reproduces the psychophysically observed phenomenology and distributions of the percept times. Since the model is based on learned templates, these results would transfer trivially to other action patterns with the similar form of bistability in the view domain.

It is important to stress that the goal of this paper was the modeling of the perceptual dynamics, and neither the proposal of novel deep shape or action recognition architecture, nor the claim that the proposed two-pathway architecture is significantly better for shape recognition. Testing this claim would require additional experiments with larger data sets, and was not the focus of this paper. Also it remains to be shown whether any of the popular recurrent deep architectures reproduce the details of the human perceptual dynamics.

Future work will have to extend the model for more stimuli and include more accurate fits of experimental data.

ACKNOWLEDGEMENTS

The first author thanks Tjeerd Dijkstra for his insightful commentary on the analysis of the Amari field behavior. Funded by: BMBF, FKZ: 01GQ1002A, ABC PITN-GA-011-290011, CogIMon H2020 ICT-644727; HBP FP7-604102; Koroibot FP7-611909, DFG GZ: KA 1258/15-1; HFSP RGP0036/2016.

REFERENCES

- Amari, S. (1977). Dynamics of pattern formation in lateral inhibition type neural fields. *Biological Cybernetics*.
- Andersen, R. and Bradley, D. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Sciences*.
- Barraclough, N. and Jellema, T. (2011). Visual aftereffects for walking actions reveal underlying neural mechanisms for action recognition. *Psychological Science*.
- Blake, R. and Logothetis, N. (2001). Visual competition. *Nature Review Neuroscience*.
- de la Rosa, S., Streuber, S., Giese, M., Buelthoff, H., and Curio, C. (2014). Putting actions in context: Visual

- action adaptation aftereffects are modulated by social contexts. *PLOS One*.
- Edwards, M., Deng, J., and Xie, X. (2016). From pose to activity. In *Computer Vision and Image Understanding*. Elsevier Science Inc.
- Escobar, M. and Kornprobst, P. (2008). Action recognition with a bioinspired feedforward motion processing model: the richness of center-surround interactions. In *ECCV'08. 10th European Conference on Computer Vision*. Springer Berlin Heidelberg.
- Fleischer, F., Caggiano, V., Thier, P., and Giese, M. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience*.
- Giese, M. (2014). Skeleton model for the neurodynamics of visual action representations. In *Artificial Neural Networks and Machine Learning ICANN 2014, Lecture Notes in Computer Science*. Springer International Publishing.
- Giese, M. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements and action. *Nature Reviews Neuroscience*.
- Hock, H., Kelso, J., and Schoener, G. (1993). Bistability and hysteresis in the perceptual organization of apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*.
- Jackson, S. and Blake, R. (2010). Neural integration of information specifying human structure from form, motion, and depth. *Journal of Neuroscience*.
- Jhuang, H., Serre, T., Wolf, L., and Poggio, T. (2007). A biologically inspired system for action recognition. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE.
- Lange, J. and Lappe, M. (2006). A model for biological motion perception from configural form cues. *Journal of Neuroscience*.
- Layher, G., Giese, M., and Neumann, H. (2014). Learning representations of animated motion sequences a neural model. In *Topics in Cognitive Science*. Topics in Cognitive Science.
- Lee, T., Belkhatir, M., and Sanei, S. (2014). A comprehensive review of past and present vision-based techniques for gait recognition. In *Multimedia Tools and Applications*. Kluwer Academic Publishers.
- Leopold, D. and Logothetis, N. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Science*.
- Nguyen, D., Li, W., and Ogunbona, P. (2016). Human detection from images and videos. In *Pattern Recognition*. Elsevier Science Inc.
- Pastukhov, A., Garca-Rodriguez, P., Haenicke, J., Guillamon, A., Deco, G., and Braun, J. (2013). Multi-stable perception balances stability and sensitivity. *Frontiers in Computational Neuroscience*.
- Rankin, J., Meso, A., Masson, G. S., Faugeras, O., and Kornprobst, P. (2014). Bifurcation study of a neural field competition model with an application to perceptual switching in motion integration. *Journal of Computational Neuroscience*.
- Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Science*.
- Thurman, S. and Lu, H. (2014). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience*.
- Thurman, S. and Lu, H. (2016). A comparison of form processing involved in the perception of biological and nonbiological movements. *Journal of Vision*.
- Tyler, C. (2011). *Computer Vision: From Surfaces to 3D Objects*. Chapman & Hall/CRC, London, 1st edition.
- Vangeneugden, J., de Maziere, P., van Hulle, M., Jaeggli, T., van Gool, L., and Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *Journal of Neuroscience*.
- Vangeneugden, J., van Ee, R., Verfaillie, K., Wagemans, J., and de Beeck, H. (2012). Activity in areas mt+ and eba, but not psts, allow prediction of perceptual states during ambiguous biological motion. In *Society for Neuroscience Meeting*. Society for Neuroscience.
- Vanrie, J. and Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, and Computers*.
- Vanrie, J. and Verfaillie, K. (2006). Perceiving depth in point-light actions. *Perception and Psychophysics*.
- Wilson, H. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences*.
- Ziaeeafard, M. and Bergevin, R. (2015). Semantic human activity recognition: A literature review. In *Pattern Recognition*. Elsevier Science Inc.

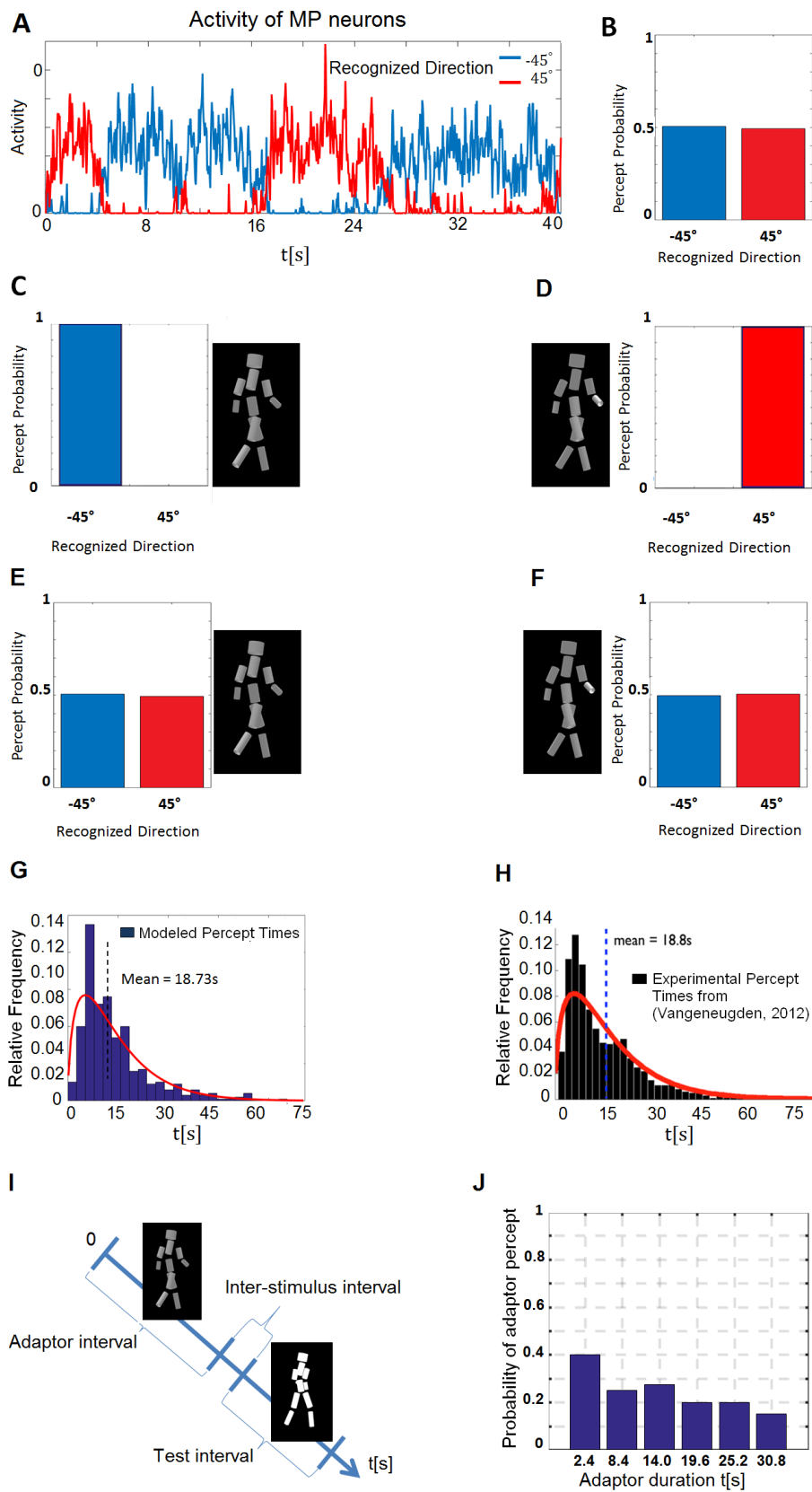


Figure 2: A. Time courses of the activity of motion pattern neurons for depth-ambiguous walker stimulus. B-F. Percept probability of the motion pattern neurons for the percepts TOWARDS and AWAY for (B) depth-ambiguous walker for model with both pathways; (C) shaded -45° (AWAY) walker for model with both pathways; (D) same for shaded 45° (TOWARDS) walker; (E) shaded 45° (TOWARDS) walker for model without shading pathway; (F) same for shaded 45° (TOWARDS) walker; G-H. Histogram of percept times (PT) from experimental data (Vangeneugden et al., 2012) and from the model. I. Paradigm for testing after-effects in action perception which is compatible with our model. After presentation of an unambiguous adaptor stimulus (AWAY or TOWARDS), and a fixed Inter-stimulus Interval, an ambiguous test stimulus (SILHOUETTE) is presented. J. Probability that test stimulus is perceived as walking in the adaptor direction as a function of the duration of the adaptor.