Action in Mind: Neural Models for Action and Intention Perception

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt von
Mohammad Hovaidi-Ardestani
aus Teheran, Iran

Juli, 2020

Tag der mündlichen Prüfung:	01.10.2020
Dekan der Math.-Nat. Fakultät:	Prof. Dr. József Fortágh
Dekan der Medizinischen Fakultät:	Prof. Dr. Bernd Pichler
1. Berichterstatter:	Prof. Dr. Martin A. Giese.
2. Berichterstatter:	Prof. Dr. P. Thier
Prüfungskommission:	1.Prof. Dr. Martin A. Giese
	2. Prof. Dr. P. Thier
	3.Prof. Dr. Martin V. Butz
	4.Prof. Dr. Hanspeter A. 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*

Acknowledgments

In academic life, Ph.D. is the most challenging chapter. When I write these lines, I remember all the sleepless nights before the deadlines, all the challenging tasks for which I was the only responsible person to tackle and all the moments that I was disappointed to reach the finish line. Along the process of writing this chapter of my life, I have been really grateful to have a well-respected high-profile supervisor that without his precious support it would not be possible to conduct this research. My sincere thanks goes to my advisor Prof. Giese for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. I appreciate all his contributions of time, ideas, and funding to make this academic experience productive and stimulating.

I would like also to express my gratitude to all the members of computational sensomotorics team and all of my former colleagues whose assistance was a milestone in the completion of this project. I would always remember all my fellow groupmates for the fun-time we spent together, sleepless nights that gave us the courage to complete tasks before deadlines, the feedbacks and comments on rehearsing talks, and for stimulating the discussions. Among them I would like to specially thank my mentor Dr. Tjeerd Dijkstra for his friendship, empathy, and great sense of humor. He inspired me by vision, sincerity, and his scientific talks and showed me how to present the research work as clearly as possible by avoiding unnecessary technical jargon. I would be remiss if I did not thank Mirjana Angelovska, who deserves credit for providing much needed assistance with administrative tasks, reminding us of impending deadlines, and keeping our work running smoothly. During the years of my work, she was always so helpful and had a sympathetic ear for any kind of concern.

Last but not least, I would like to thank my family for supporting me spiritually throughout writing this thesis and my life in general. They gave me enough moral support, encouragement and motivation to accomplish the personal goals. At the end I would like to express my special appreciation to my beloved wife Behnaz who experienced all of the ups and downs of my research and was always my support in the moments when there was no one to answer my queries. I will keep on trusting You for the future.

Abstract

To notice, recognize, and ultimately perceive the others' actions and to discern the intention behind those observed actions is an essential skill for social communications and improves markedly the chances of survival. Encountering dangerous behavior, for instance, from a person or an animal requires an immediate and suitable reaction. In addition, as social creatures, we need to perceive, interpret, and judge correctly the other individual's actions as a fundamental skill for our social life. In other words, our survival and success in adaptive social behavior and nonverbal communication depends heavily on our ability to thrive in complex social situations. However, it has been shown that humans spontaneously can decode animacy and social interactions even from strongly impoverished stimuli and this is a fundamental part of human experience that develops early in infancy and is shared with other primates.

In addition, it is well established that perceptual and motor representations of actions are tightly coupled and both share common mechanisms. This coupling between action perception and action execution plays a critical role in action understanding as postulated in various studies and they are potentially important for our social cognition. This interaction likely is mediated by action-selective neurons in the superior temporal sulcus (STS), premotor and parietal cortex. STS and TPJ have been identified also as coarse neural substrate for the processing of social interactions stimuli. Despite this localization, the underlying exact neural circuits of this processing remain unclear. The aim of this thesis is to understand the neural mechanisms behind the action perception coupling and to investigate further how human brain perceive different classes of social interactions.

To achieve this goal, first we introduce a neural model that provides a unifying account for multiple experiments on the interaction between action execution and action perception. The model reproduces correctly the interactions between action observation and execution in several experiments and provides a link towards electrophysiological detailed models of relevant circuits. This model might thus provide a starting point for the detailed quantitative investigation how motor plans interact with perceptual action representations at the level of single-cell mechanisms. Second we present a simple neural model that reproduces some of the key observations in psychophysical experiments about the perception of animacy and social interactions from stimuli. Even in its simple form the model proves that animacy and social interaction judgments partly might be derived by very elementary operations in hierarchical neural vision systems, without a need of sophisticated or accurate probabilistic inference.

Contents

Synopsis

Action enables us to acquire perceptual information about the environment and to understand the presented information differently. When we turn around, for instance, our spatial relations to the surrounding environment will be changed. Hearing a car horn and a loud noise afterwards, might signal the occurrence of some undesirable events while we are crossing the street. Touching an object already provides useful information about its texture, shape, and temperature. Even as simple a task as eating food involves a sequence of perceptions and actions woven together by expectations and experience.

Aforementioned examples clearly suggest that perception and action are interdependent and as it was formulated by Gibson (1966), the perceptual information is used primarily in the organization of action and can subsequently facilitate the interactions of individual and the related environment.: ”We must perceive in order to move, but we must also move in order to perceive”. For example, when we move around, the pattern of optic flow in the retinal image keeps updating and therefore gives us information about our current heading direction, while the motion parallax as a monocular depth cue estimate the relative distances of objects (depth perception) by considering the relative velocities of objects moving across our field of view.

The main focus of this dissertation is directed in developing neurophysiologically plausible models which could qualitatively simulate the behavior of the brain in processing and perceiving actions. Here, it is important to emphasize that this study does not claim that the whole categories of action types and their perception mechanisms can be modeled by these novel neural models. However, the ultimate goal of designing these models would be achieved if these new models can provide a better understanding of action perception from a different perspective.

In this section, first I attempt to sketch out the action processing mechanisms as far as it has been known, and the possible underlying neural substrate by briefly reviewing and summarizing the relevant literature. Since understanding the other agents’ actions is an integral prerequisite for social interaction, the relation between action perception and action execution will be discussed adequately in the next part and then I report the research on animacy and social interaction perception as one of the most important skill granted to social creatures. Forth, in line with the focus of this thesis a short review of action recognition models and their importance will be provided. A brief overview of agent navigation models will be discussed to complete the relevant literature review section of this chapter. Finally, I will focus on the novelty of this PhD dissertation and give an overview of following papers by describing the motivation and goal of each together with emphasis on importance of their results and contribution of my work.

1. Action Perception

To notice, recognize, and ultimately perceive the others' actions and to discern the intention behind those observed actions is an essential skill for social communications and improves markedly the chances of survival. Encountering dangerous behavior, for instance, from a person or an animal requires an immediate and suitable reaction. In addition, as social creatures, we need to perceive, interpret, and judge correctly the other individual's actions as a fundamental skill for our social life.

Since the beginning of the third millennium, action observation and recognition have garnered enormous interest and have been one of the core topics in the field of neuroscience (Rizzolatti et al., 2001, Keysers and Perrett, 2004, Schütz-Bosbach and Prinz, 2007, Keysers, 2011, Rizzolatti and Fogassi, 2014, Caggiano et al., 2016, Etzel et al., 2016, Savaki and Raos, 2019). However, the fundamental neural mechanisms that provide constraints for underlying computations remains only partially explained, in contrast to the fact that a wide range of speculative theories in this direction have been proposed to shed light on understanding how the brain accomplishes this remarkable skill.

Coarsely speaking, action-selective neurons have been found in different number of brain structures, including premotor, primary motor, and parietal cortex (Rizzolatti et al., 2001, Puce and Perrett, 2003, Nelissen et al., 2011, Rizzolatti and Fogassi, 2014, Geiger et al., 2019). In addition, selective activation for biological movement stimuli has been reported in the ventral lateral occipital cortex and the lingual gyrus at the cuneus border that are sensitive to motion (Grossman et al., 2000, Servos et al., 2002, Freitag et al., 2008) as well as non-visual areas including amygdala and cerebellum (Bonda et al., 1996). These studies suggest that several parts of the brain are engaged to accomplish action recognition. The discussion here, however, is restricted to the cases relevant for the work presented in this thesis particularly experimental results from action selective regions in the superior temporal sulcus, parietal as well as premotor cortex.

A deeper understanding of the neural basis of this complex procedure will require knowledge of how single neurons encode action related stimuli. Earlier studies showed that superior temporal sulcus (STS) of monkey brain (see Figure.1) is particularly sensitive to the kinematic and dynamical signatures of biological movement and plays a crucial role to analyze visually others' actions including walking, arm movement, goal-directed hand actions and some other biological movements (Bruce et al., 1981, Perrett et al., 1985a, Oram and Perrett, 1994, Jellema et al., 2000, Vangeneugden et al., 2014). In fact, Bruce, Desimone, and Gross (1981) reported the very first study that showed response selectivity of STS to views of walking person. They recorded from single neurons in the superior temporal polysensory (STP), an area of dorsal bank, and fundus of the anterior portion of STS. A few years later, Perret and his colleagues who have extensively investigated the tuning of neurons in STS to observed body actions, also showed response selectivity of STP to the observation of walkers and concluded that motion direction is stronger cue driving neural responses than body orientation (Perrett et al., 1985a, Oram and Perrett, 1994, 1996).

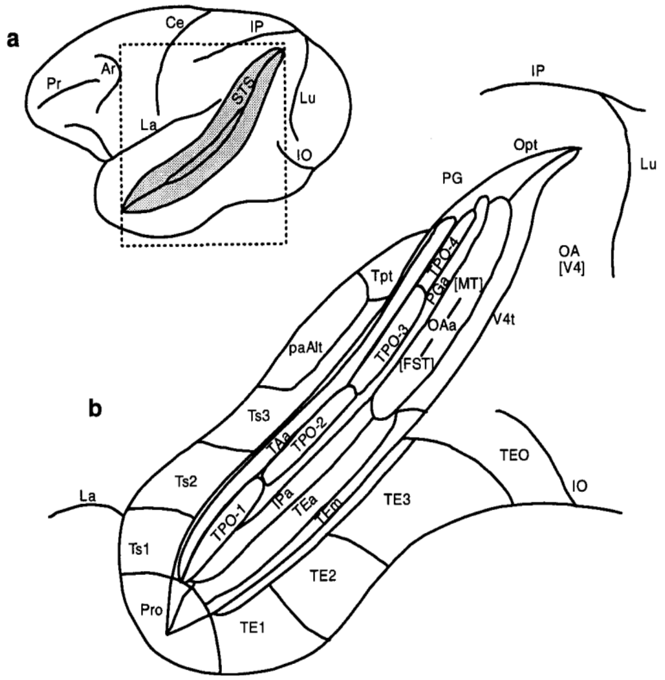


Figure 1: Subdivision of monkey inferior temporal lobe centered around the superior temporal sulcus (STS). (a) Lateral view of the cortical surface showing the upper bank, depth, and lower bank of STS with major visible sulci labeled inferior occipital (IO). lunate (Lu), intraparietal (IP). central (Ce), lateral (Sylvian) fissure (La), arcuate (Ar), and principal (Pr). (b) Enlarged drawing of the inferior temporal areas surrounding the STS. [Adapted from Seltzer and Pandya, 1994.]

The discovery of mirror neurons by neurophysiologists of university of parma (Pellegrino et al., 1992, Rizzolatti et al., 1996, Gallese et al., 1996), aroused widespread interest among researchers in the neuroscience community to study action processing and understanding more elaborately. Mirror neurons, which have been discovered in a sector of the ventral premotor region F5 (see Figure.2.a) of the macaque monkey by single cell recordings in the parieto-frontal areas, represent a distinctive class of visuomotor neurons that discharge both in monkey's brain when it executes a goal-directed action as well as when it observes the same or a similar motor act performed by another individual (Gallese and Goldman, 1998, Rizzolatti et al., 1996, Gallese et al., 1996, Rizzolatti et al., 2001)(see Figure.2.b). Findings of the seminal work (Pellegrino et al., 1992) indicated that mirror neurons can retrieve movements not only on the basis of stimulus characteristics, but also on the meaning of the observed goal-directed actions (e.g. placing or grasping an object).

In a detailed follow up study by (Gallese et al., 1996), researchers recorded 532 neurons in area F5 in the premotor cortex of the macaque monkey and detected that ninety two of these neurons have mirror properties and were consequently named mirror neurons for the first time. In this study, action-selective responses of mirror neurons for

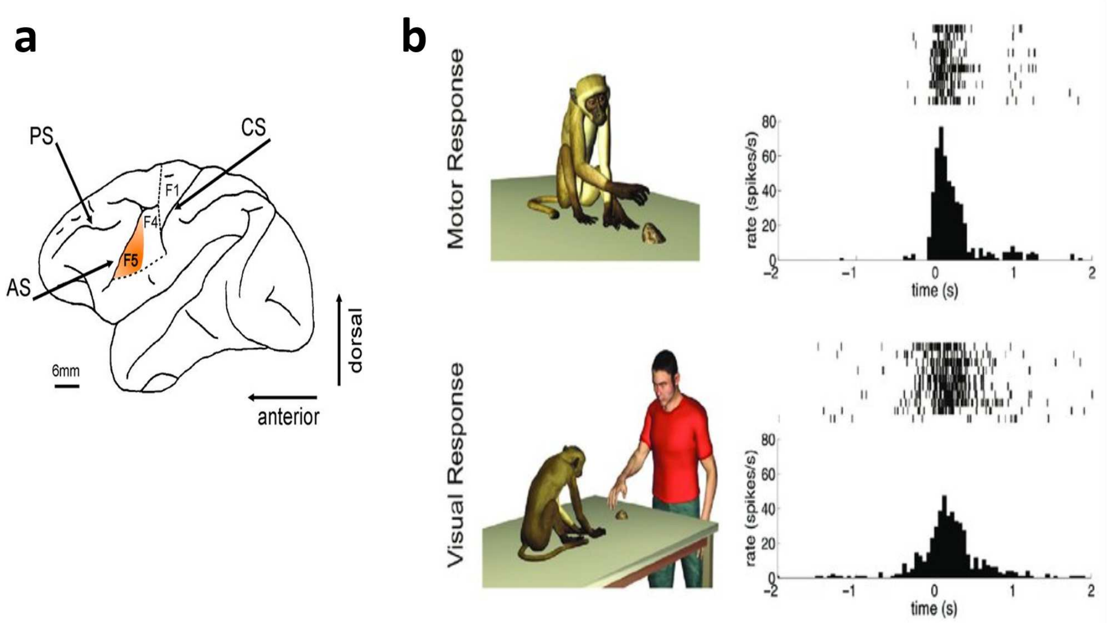


Figure 2: Examples of F5 mirror neurons. (a) Lateral view of the monkey brain showing the location of area F5 highlighted as part of the ventral premotor cortex (AS, arcuate sulcus; CS, central sulcus; PS, principal sulcus). [Adapted from Caggiano et al., 2012]. (b) Response properties of a mirror neuron. Column in the left shows the experiment conditions including action execution and action observation by the monkey. The raster plot and peri-stimulus spike density show the corresponding responses of the same neuron during active goal-directed motor acts of the monkey (e.g., grasping small objects of different shapes) and the observation of the same acts performed by the experimenter. [Adapted from Casile et al., 2011.]

different actions such as holding, placing, and grasping an object have been shown while authors noted that size of the object did not have influence on most of the neurons. This distinguishes mirror neurons from other motor or sensory neurons whose discharge is associated either with execution or observation, but not both.

Mirror neurons in human brain have also been widely investigated and challenged in a vast number of experiments from different perspectives (e.g., Rizzolatti et al., 2001, Gazzola and Keysers, 2009, Mukamel et al., 2010). However, finding individual mirror neurons in humans is nearly impossible due the fact that electrophysiology is rarely possible at only specific brain regions. Therefore, the majority of the studies that focus on the ”mirror” properties of the human brain use non-invasive methods including transcranial magnetic stimulation (TMS), electroencephalography (EEG) functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) to assess the analogous mechanisms in human (for reviews see Rizzolatti and Sinigaglia, 2010, Molenberghs et al., 2012, Kilner and Lemon, 2013, Rizzolatti and Sinigaglia, 2016).

Many studies have shown analogous functional properties of the mirror mechanism in human including understanding action, mirroring and understanding emotions, and action execution and observation and some other controversial ones (e.g., Abdollahi et al., 2013, Hickok, 2013, Keysers and Gazzola, 2014, Campbell and Cunnington, 2017). However, some differences between monkeys and humans not only challenged some of these functions but also raised questions even about the existence of a mirror neuron system in human (e.g. Heyes, 2010). One of this differences is, that brain regions in human that were implicated in mirror mechanism functioning, not only found in homologous areas, but also in regions outside those where mirror neurons have been reported in monkeys. This could be traced back both to the variety of approaches and inherent accuracy problem of indirect measures applied in mirror systems in human. Nevertheless, action perception and action execution function of the mirror system is more ubiquitously agreed upon and the understanding of their processes has profoundly been changed by the discovery of motor neurons. This ability is important for our non-verbal social communication (for reviews see Frith and Frith, 2012) and understanding and simulating this function of the human brain was one of the main motivations behind commencing this PhD work.

From a functional viewpoint, action execution and action perception are closely-related processes with a candidate neural substrate of mirror neuron system and specifically the premotor and prefrontal cortex and the STS that is especially sensitive to point-light animations of biological motion. In addition, several studies discuss also that understanding of others' actions is accomplished with the prediction of sensory consequences from ongoing movements, processed in cerebellum and the posterior parietal cortex (e.g., Miall, 2003; Sokolov et al., 2010; Sokolov et al., 2017). Regardless of the exact neural substrate for action perception, the evidences commonly show a remarkable overlap between brain regions recruited during action understanding. Some even claimed that our ability to interpret the actions of others requires the involvement of our own motor system meaning that an action could be understood only if there is a mapping between the agent's observed movements and the observer's motor repertoire (e.g., Rizzolatti et al., 2001). Obviously, if this was true we could not understand a bird's fly in the sky since we cannot map a bird's wing movements onto our own motor repertoire (Jacob, 2009). In addition , this mapping hypothesis suggests that the brain represents others' actions like one's own. This has been investigated in some studies and was concluded that motor facilitation strongly depends on the agent to whom the observed action is attributed which simply means that motor systems may be involved differently in processing self-action and actions executed by others, which signifies a role in social cognition (e.g., Schütz-Bosbach et al., 2006). This paper and some others in which researchers scrutinized also the neuroimaging data and found the replicated results led to more cautious conclusions such that activation in the motor system could reflect alternative mechanisms, such as encoding of the semantic features of actions (e.g., Press et al., 2012).

The employing of the same motor system in action perception is also of particular interest for the theoretical accounts of this close link between neural activity in the motor-system during action observation and our ability to execute the same action (e.g., Kilner et al., 2007; Friston et al., 2011; Donnarumma et al., 2017; Kahl and Kopp, 2018). The central principle in all these theoretical models is that the same motor models used during execution of action is exploited for the interpretation and inference during observation of actions.

Although there is an ever-expanding body of literature about action perception and the role of mirror neurons in brain, many questions yet need to be addressed. A comprehensive knowledge about the connectivity of mirror neurons and their comparative biology across different species might decode the the true role of mirror neurons and their putative functional roles as ”The most hyped concept in neuroscience” (Jarrett, 2012). The focus of the next section is to review briefly the main studies that illustrate more clearly how the action perception and action execution are linked in the brain.

1.1 Action Perception Cycle

As suggested in the previous section, action and perception are found to be tightly coupled as opposed to the traditional approaches to human information processing that tend to deal with perception and action planning in isolation. The roots of the concept of this close entanglement can be traced to 19th century when British and German researchers independently postulated that, an association between action and sensory consequences needed to be formed at first and then the ”soul” can trigger the action automatically by the anticipation of their intended perceptual effects (Gibson, 1852; Herbart, 1852). This, later called ideomotor principle as a combination of the idea and the motor act and it is known as the first principle that theorized the link between action and perception. The aim of this section however, is to review the most recent and relevant studies that focus on the theories that stress this close link between perception and action and show the influences of concurrent action on perception and the remarkable impact of perception on actions.

As it has been shown, an important functional aspect of mirror neurons is the relation between their visual and motor properties and can be assumed as common neural substrate for action and perception. The relevant studies commonly discuss that, almost all mirror neurons show congruence between the visual actions they respond to and the motor responses they code. These studies have also suggested that action perception and action execution are intrinsically coupled in the human brain (Rizzolatti and Craighero, 2004, Rizzolatti and Sinigaglia, 2010, Brucker et al., 2015). In addition, the common coding theory (Prinz, 1997) as well as the theory of event coding (Hommel et al., 2001), have also been introduced as a functional version of sharing mechanism of action and perception with possibility that mirror neurons could be the neural substrate of this bidirectional link. In addition to these behavioral and neurophysiological studies, computational accounts have also endorsed this tight link between action perception and action execution in the brain (e.g., Wolpert and Ghahramani, 2000; Wolpert and Kawato, 1998).

Following up on the above mentioned studies, a tight interplay between action and perception is highly expected. Especially based on the common coding theory and mirror neurons, both processes share a common neural substrate, and contemporary ideomotor approaches, the linkage between perception and action originates from shared representational resources, one can expect that action and perception may induce or interfere with each. Under the theory of event coding, Müsseler and colleagues in 1997 (Müsseler and Hommel, 1997) demonstrated, as one of the first studies discussing the effects of action on perception, an inhibiting influences of concurrent action on perception. They showed that the identification of a left- or right-pointing arrow is impaired when it was presented with an arrow as visual stimulus during the preparation and execution of corresponding key press. Since the congruent executed hand posture and shown pictures induced less deterioration of perceptual judgments and made the participants ”blind” for this similar postures, authors concluded that an action can induce blindness.

This interference effects have become the focus of a number of similar investigations that tried to obtain the effects of action on perception (e.g., Lindemann et al., 2006; Zwickel et al., 2007; Roussel et al., 2013; Christensen et al., 2011; Thomaschke et al., 2018). An example that shows how action can compromise perceptual judgments is the study conducted by Hamilton and colleagues (Hamilton et al., 2004). They examined the subjects' perceptual judgments of weight of a box from other people's action while they performed different motor tasks. They measured the weight estimation of observers while lifting actively, holding passively, or maintaining a neutral condition. Subjects underestimated the weight of the box in the video clip when they were lifting a heavy box vice versa overestimated the weight of box while physically lifting a light box. Another study that experimented how action impairs or biases perception of related actions has been shown by (Jacobs and Shiffrar, 2005). They showed that participants as walking observers demonstrated the poorest sensitivity to walking speed in comparison to concurrently performing cycling and standing. This results indicated that a clear influence of related motor activity on action perception.

Besides the discussed impairment of perceptual sensitivity by motor execution, a variety of studies indicate that motor expertise may afford better understanding of action-congruent stimuli (e.g., Fagioli et al., 2007; Wykowska and Schubö, 2012; Catmur et al., 2018). In addition, the facilitation of the body motion perception has been reported in a number of interesting studies (e.g., Casile and Giese, 2006; Calvo-Merino et al., 2006; Calvo-Merino et al., 2010; Calmels et al., 2018). For instance, Casile and Giese in 2006, for the first time, demonstrated a direct and highly selective improvement of the visual recognition performance for the novel acquired motor programs independent of visual learning. In this study researchers, using a new experimental paradigm, dissociated visual and motor learning during the acquisition of new motor patterns. They assessed the visual recognition of gait patterns from point-light stimuli before and after nonvisual motor training while subjects were blindfolded and learned based only on verbal and haptic feedback. Their results proved the direct influence of motor learning on visual recognition of action even if they have been acquired in the absence of visual learning. In a more recent study, researchers exploited a novel virtual reality paradigm for the online control of biological motion stimuli to indicate that execution of motor behavior influences concurrent visual action observation (Christensen et al., 2011). In this signal-detection task, subjects sitting in front of a projection screen had to detect a point-light arm controlled by their own movements in a scrambled mask. The results demonstrated that if the concurrent motor execution is temporally synchronous and spatially congruent with corresponding arm movement the biological motion detection is facilitated and inhibited otherwise. This study showed, for the first time, a range of influences between visual action recognition and action execution from facilitatory interactions to interference.

As discussed earlier, perception and action rely on a 'shared representational system' as coined by Prinz in 1997. This suggest that not only action execution influences on perception but also mere observation of an action activates a corresponding motor representation in the observer. Therefore, substantial efforts have been made to investigate the effect of perception on action as well. The interference effect of perception on action has been investigated in several studies (e.g., Kilner et al., 2003; Bouquet et al., 2007). In the first investigations into this interference, the variance in the executed movement was measured as an index of interference to the movement while subjects made arm movements observing another human making the same or qualitatively different arm movements (Kilner et al., 2003). This study demonstrated a significant interference effect on executed movements in case of incongruent movements.

The coupling of action perception and action execution is also important for our interactions with others. Our survival and success in adaptive social behavior and nonverbal communication depends heavily on our ability to thrive in complex social situations (Kennedy and Adolphs, 2012). In the next section we show that our mirror system namely inferior frontal gyrus (IFG), the inferior parietal lobule (IPL), the medial temporal gyrus (MTG), and the superior temporal sulcus (STS), play a critical role in action understanding as postulated in various studies and they are potentially important for our social cognition (Gallese et al., 2004).

1.2 Social Interaction Perception

Seeing and understanding the movements of others lie at the basis of social interaction perception and it is critical for social life. This requires the ability to perceive not just individuals and their actions but to accurately interpret conspecifics' actions and the interactions between them. This crucial cognitive function of human is shared with other primates as inherently social beings, and have made them very good at interpreting intent and social behavior from others, i.e., determining the social relationship between agents, and making predictions about their intentions or mental states.

As it has been mentioned earlier the fundamental computations and neural mechanisms involved in the perception and processing of social intent have remained relatively unexplored. Studies that are reviewed in this section are among the most appreciated ones in this emerging field of study that attempt to show brain regions that are recruited during the observation of social interactions. More importantly in this section, we review quickly how a well-known social interaction judgment experiment demonstrated that displaying moving simple 2D shapes can give rise to precepts of animacy and to impute human characteristics even to geometric figures in nonrandom motion. This seminal work and its subsequent studies that investigated the cues that can directly invoke the perception of interaction were the main motivations behind the second part of this PhD work.

Heider and Simmel in 1944 demonstrated that humans can perceive intent or social interaction from strongly impoverished stimuli. In this seminal work the human participants were shown a video with simple 2D figures in motion. The participants not only perceived the objects as alive, but also they attributed different features of social interactions including fighting, following, attacking and etc. (Fig.3). These anthropomorphic interpretations were spontaneous and highly consistent from one subject to another. The striking ability of humans that can appreciate the cause-effect relationship of interacting simple geometrical shapes (social interaction perception), and also spontaneously recognize and appropriately perceive that another entity is an alive agent (animacy perception), have been demonstrated widely in several other seminal works (e.g., Michotte, 1963; Leslie, 1995; Scholl and Tremoulet, 2000; Schlottmann et al., 2006).

It is worth mentioning here that the interesting literature, developed specifically after initial demonstrations of Michotte and Heider, have referred to this phenomena variously. Social causality (e.g., Rochat et al., 1997), social meaning (e.g., Tavares et al., 2008), action understanding (e.g., Baker et al., 2009), soical interaction perception (e.g., Walbrin et al., 2018, Isik et al., 2019), goal-directedness (e.g., Csibra, 2008), intentionality (e.g., Dasser et al., 1989), or perception of animacy (e.g., Leslie, 1995) are the terms that have been coined more frequently by researchers. Although there are some differences among these terms and their applications, for the purpose of the discussion in this thesis we refer to the animacy perception as the ability of recognizing a single living agent and social interaction perception for the ability to detect and understand other agents' social interactions. Before we dive into the discussion of crucial features for both animacy and social interaction perception, in the following we will see shortly what is known already as the neural mechanism underlying this precept.

While much is now known about the different brain mechanisms involved in the recognition of objects (e.g., Riesenhuber and Poggio, 2002), scenes (e.g., Nakamura et al., 2000), bodies (e.g., Aviezer et al., 2012), faces (e.g., Todorov et al., 2008), and emotion (e.g., Frühholz et al., 2015; Adolphs et al., 2003), the neural circuitry that underlies interaction processing and the fundamental computations and mechanisms involved in the perception and processing of social intent are less explored. Although it has been frequently shown that a region in the pSTS is sensitive to the presence of social interactions, it is not the only area in the brain that shows activation in presence of social relevant stimuli. Several studies have demonstrated that its the neighboring temporo-parietal junction (TPJ) as well as bilateral inferior frontal gyrus (IFG), and angular gyrus, particularly in the right hemisphere can be activated when viewing social interaction stimuli (e.g., Iacoboni et al., 2004; Laurie et al., 2011; Dolcos et al., 2012; Molenberghs et al., 2012). However, converging fMRI evidence suggests that the posterior superior temporal sul-

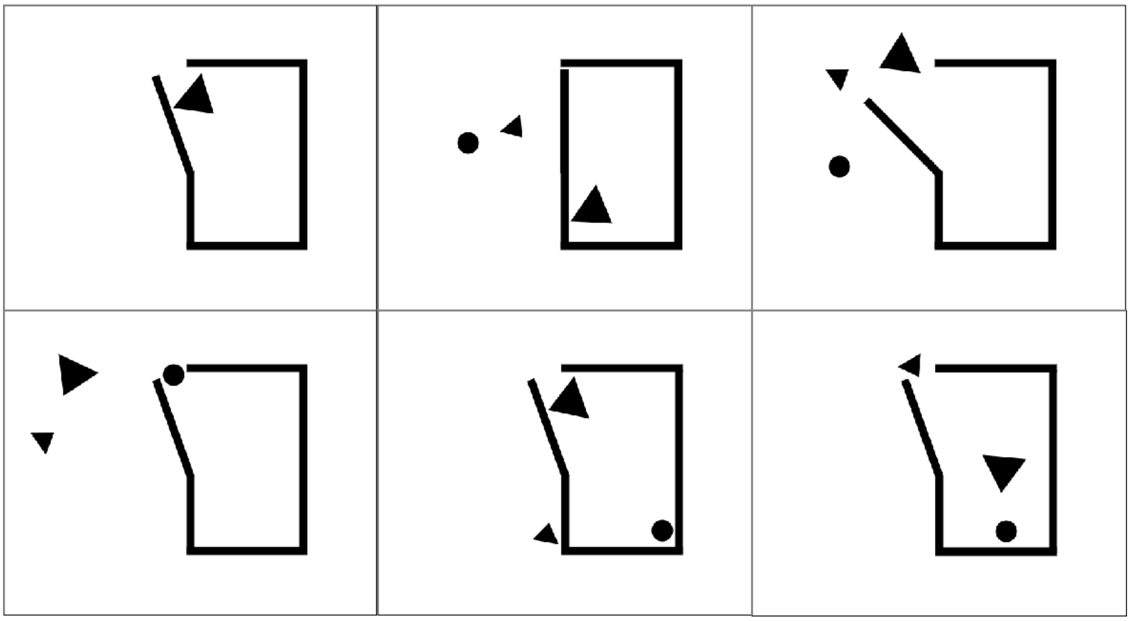


Figure 3: Snapshots from the chase scene in the Heider-Simmel video [Heider and Simmel, 1944]. The video shows simple 2D shapes (a large triangle, a small triangle and a small circle) moving around and having interactions. Observers attributed personality traits (e.g. shyness, being a bully) and emotions (e.g. frustration, anger) to these geometric figures

cus (pSTS) in humans is likely candidate and has been described in the literature as the ”hub” of the social brain (e.g., Lahnakoski et al., 2012; Yang et al., 2015; Deen et al., 2015; Walbrin et al., 2018; Walbrin et al., 2020). Moreover, a recent study in macaques also demonstrates that STS is a central region in the visual analysis of conspecific social interactions (Sliwa and Freiwald, 2017). This study discovered a network centered in the medial and ventrolateral prefrontal cortex that is selective for social interactions and did not respond to any other stimulus. More importantly, the results of this interesting study, using whole-brain functional magnetic resonance imaging in macaque monkeys, revealed that the large parts of shape-selective STS are interaction-selective (i.e. competition and cooperation) and introduced this as a new dimension of tuning and functional organization of region of the STS.

One of the most intriguing aspect of Heider and Simmel’s work was the visual display itself that contains only a few simple moving 2D geometric shapes that do not have features of any living agents (e.g., faces, hands, biological motions). This suggested that motion features together with a few simple visual features of objects (e.g., anteroposterior axis) could be sufficient not only to induce a strong animacy perception but also to attribute intention characteristics to such moving agents. For the rest of this discussion and in alignment with the purpose of this thesis, we focus only on researches that attempted to reveal features that characterize social interaction and animacy perception of agents using such simple displays and do not discuss other experiments that have used more specialized and rich stimuli. This means that high level visual attributes including facial features (e.g., Johnson, 2000), surface information, and body configuration (e.g., Eimas and Quinn, 1994) that are considered relevant for identifying animacy and intention perception are out of the scope of this work.

The ability to perceive social information given visual motion appears to develop early. Exploiting such simple stimuli and based upon only the movement patterns of simple geometrical shapes, a series of developmental studies have revealed that children have the ability to infer goal-directedness action and further interpret such moving 2D objects as animate agents (Premack, 1990; Gergely et al., 1995; Csibra et al., 1999; Csibra, 2008; Csibra and Southgate, 2009). Some studies even demonstrated that children and adults with Autism Spectrum Disorder (ASD) show no deficit in the ability of animacy perception after they have reached criterion in the training phase (Rutherford et al., 2006; Vanmarcke et al., 2017). Although studies that have examined goal-directed action perception in children have suggested that this ability is constrained by other factors such as cultural context (e.g., Green et al., 2016), emotional context (e.g.,Trautmann et al., 2009), and facial expressions (e.g., Rennels et al., 2017), we can still can claim that motion features of the stimuli play an important role for perceptual animacy and social inferences. Having demonstrated that, researchers attempted to discover further the specific motion cues that arise such precepts.

Since 1944 and after Heider and Simmel published their seminal work, several researchers have attempted to reveal which visual cues of motion promote the perception of animacy. Violations of the conservation energy principle, such as heading and acceleration (e.g., Scholl and Tremoulet, 2000), and Newtonian laws of motion (e.g., Kaduk et al., 2013), together with speed and trajectory direction changes (e.g., Scholl and Tremoulet, 2000; Szego and Rutherford, 2008; Träuble et al., 2014) are among the minimal kinematic cues for the emergence of animacy in moving shapes. In addition, some behavioral properties such as goal-directedness (e.g, Schlottmann and Ray, 2010), being reactive to social contingencies (e.g., Dittrich and Lea, 1994), and self-propelledness (e.g., Csibra, 2008) also seem crucial for identifying animate beings. Among all mentioned features, self-propelled motion (i.e. without the application of an external force), seems to be the most powerful cues to animcy perception implying the perceptual evidence of a hidden energy source (e.g., Premack, 1990; Hauser, 1998; Csibra, 2008). It means that if an object seems to possess a hidden energy source, it is perceived as intentional which is known in the literature as ”Energy Violation Hypothesis” (e.g., Scholl and Tremoulet, 2000). This hypothesis implies that the perception of animacy highly depends on motion cues. However, some studies (e.g., Tremoulet and Feldman, 2006) argue that the motion information can be ambiguous and insufficient for such percept and humans exploit additional information for correct identification of animate or inanimate object. Therefore, the more precise conclusion could be that the contextual information along with the motion information guides the animacy perception.

Social interaction perception or the ability to infer the intentions of others could be achieved solely based on motion patterns which have been regarded as sufficient to trigger a strong impression of intentionality (e.g., Gergely et al., 1995; Csibra et al., 1999; Scholl and Tremoulet, 2000; Barrett et al., 2005; McAleer and Pollick, 2008). Barrett et al. in 2005, introducing a novel method for generating six basic categories of intentional motion (see Blythe et al., 1999), analyzed the specific motion cues that allow these intention-from-motion judgments and examined the accuracy of these judgments across cultures. To gather a set of naturalistic whole-body motion trajectories reflected the six categories of social interactions namely chasing, fighting, courting, following, guarding, and playing, they developed a two-person computer game. Since the instructions on how to generate each types of intention could bias the players, the experimenters just gave the label of the intentional motion and asked participants to produce trajectories that closely matched their intuitive motion schemes. Researchers noticed that judging the intentions of others based only on motion cues. They also elected several simple cues that could be computed readily from a motion trajectory and showed how diagnostic they are for categorizing two-agent social interactions.

The results of this work were replicated by McAleer and Pollick in 2008 and additionally revealed the advantage for viewing intentional motion from an overhead viewpoint. In this study, using a novel approach to create animate shapes whose motions are directly derived from human actions, people were surprisingly better at attributing intentions to displays shown from an overhead (i.e., 2D stimuli) rather than side view. The results also confirmed the important predictors namely relative distance, relative and absolute heading, speed and acceleration of agents as the most informative motion cues in the understanding and attribution of social intention.

2. Action Recognition Models

Visual recognition of action has been an active research area over the last two decades in the field of cognitive neuroscience (Rizzolatti et al., 2001; Keysers and Perrett, 2004; Keysers, 2011; Rizzolatti and Fogassi, 2014; Thompson et al., 2019) as well as other disciplines namely, computer vision (e.g., Zhu et al., 2017), robotics (for review see Vrigkas et al., 2015), and even philosophy (e.g., Tsakiris and Haggard, 2005; Jacob, 2009; Sinigaglia, 2013). A wide range of applications, including, video surveillance (e.g., Han et al., 2015), human machine interaction (e.g., Lee and Lee, 2011), video retrieval (e.g., Chaquet et al., 2013), and intelligent driving (Meiring and Myburgh, 2015) has attracted an outstanding interest for this topic. Therefore, a vast number of speculative explanations together with conceptual and computational models for action recognition have already been discussed and developed.

In the previous sections mirror neurons as the possible neural substrate of action processing and understanding together with action perception cycle and its role in social interaction perception were discussed. Here, in this section we only focus on models that narrow down the underlying computational mechanisms with explicit mathematical

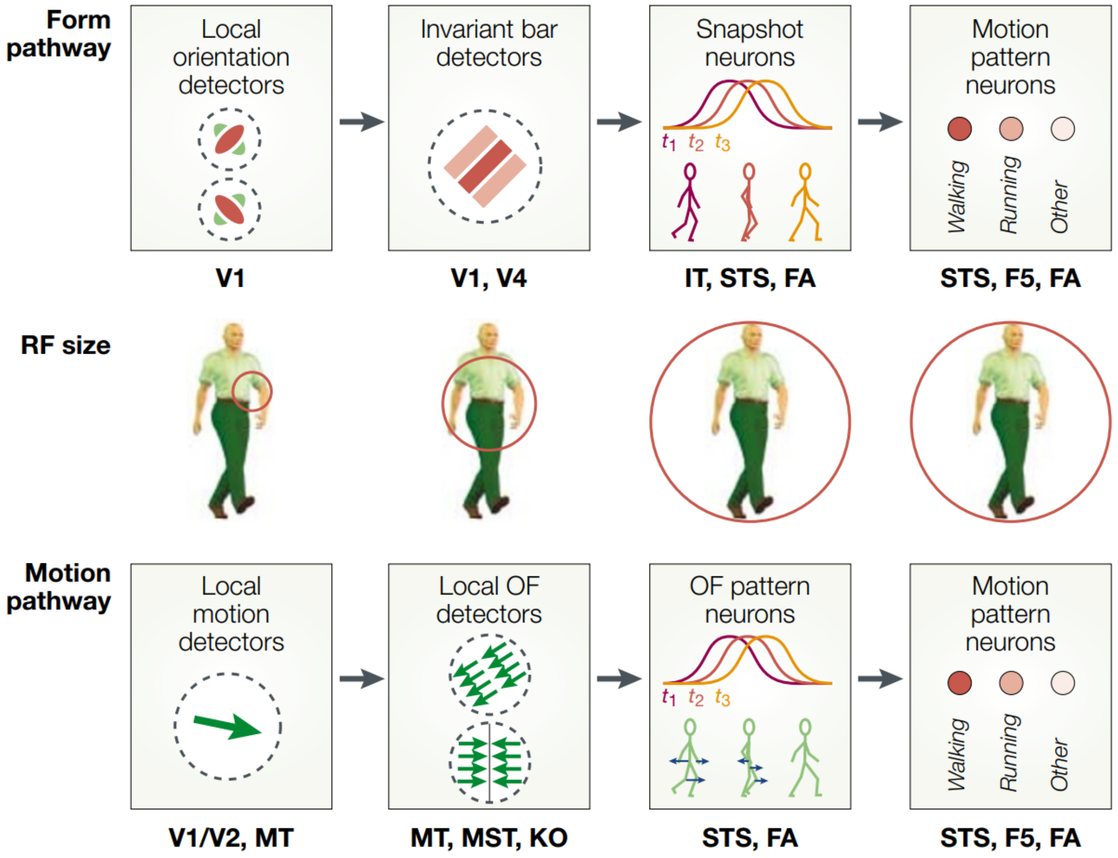


Figure 4: Overview of the model (Adopted from Giese and Poggio, 2003) for the recognition of biological body movements with two pathways for the processing of form and optic flow information. (V: visual cortex, F5: premotor cortex, M(S)T: medial (superior) temporal cortex, KO: kinetic-occipital area, IT: inferior temporal cortex, EBA: extrastriate body area, FBA: fusiform body area, IPL: inferior parietal lobule).

implementations. Thus, in alignment with the focus of this thesis I review briefly the dynamic neural field models as well as artificial neural networks from classics to the recent deep learning models (for review of other important models see this comprehensive review by Giese and Rizzolatti, 2015).

A class of biologically motivated models for action recognition is based on recurrent neural networks and a theoretical framework of dynamic neural fields (DNFs). DNFs are tissue level models that describe the spatiotemporal evolution of distributed activation patterns as the average activity of cortical neurons (Wilson and Cowan, 1973; Amari, 1977; Meijer and Coombes, 2014; Schöner and Spencer, 1966). Since this model of the recurrently connected space-continuous ensembles of neurons is sometimes mathematically traceable, its emerging neural activity patterns can easily be understood and for some cases predicted. For instance, the existence of self-sustained neural population behavior that has been shown in many areas of higher association cortices (e.g., Miller, 2000) can be explained by the positive reciprocal feedback among neighboring neurons (Amari, 1977). This persistent behavior also is widely assumed to support a multitude of related cognitive functions such as working memory, decision making and the learning of connections between events that are time-separated (Curtis and Lee, 2010). DNFs have also been employed as building blocks in architectures of models for the distributed representation of motor programs as well as for the encoding of perceived visual pattern sequences (e.g., Zhang, 1996; Erlhagen and Schöner, 2002; Giese and Poggio, 2003; Cisek and Kalaska, 2010). In robotics also researchers modeled the STS, F5 and prefrontal cortex exploiting DNFs to account for the interaction between action planing and movement recognition (e.g., Sousa et al., 2015).

One of the most influential neurophysiologically plausible model of biological movement recognition designed based on DNFs was introduced in 2003 by, Giese and Poggio. In this model the recurrent neural networks with asymmetric lateral connections were deployed to encode the perceived visual pattern sequences of point light walker stimuli. The resulting network dynamics suppresses responses to the randomization of the temporal order of the frames of a movie which leads to the destruction of a biological movement's perception. Since the computational mechanisms of this model formed the basis of the models developed in this thesis, a very short description of this model will be provided subsequently.

In accordance with the neural tuning properties in the ventral and dorsal stream of visual processing that are specialized for the analysis of form and motion (optic flow) information (Ungerleider and Mishkin, 1982) , the model is divided into two parallel processing streams consisting of hierarchies of neural feature detectors that mimic properties of cortical neurons (illustrated in Fig.4). Consistent with the known properties of cortical neurons in the visual pathway, the complexity of these feature detectors become more complex and the receptive fields of neurons tend to get larger along with the complexity of their optimal stimuli. In this hierarchical scheme, translation invariance is achieved by pooling over afferents tuned to different transformed versions of the same stimulus as it was sketched originally by (Perrett and Oram, 1993). The maximum pooling technique is further used to achieve the scale invariance similar to related shape recognition models (e.g., Riesenhuber and Poggio, 1999 ; Serre et al., 2005).

The form pathway achieves recognition of actions by recognizing sequences of 'snapshots' of body shapes. Neurons on the first level of the form pathway model the simple cells (Hubel and Wiesel, 1962) in primary visual cortex (V1) responding selectively for local oriented contours using Gabor filters. Local orientation information is extracted by neurons on the second hierarchy level that are selective for position- and scale invariant bar detectors corresponding to complex cells in areas V2 and V4. The third layer contains snapshot neurons that are selective for the particular configurations of the human body that are characteristic for actions and biological movements. These view-tuned neurons that have been found in inferotemporal cortex (area IT) of monkeys (Logothetis and Sheinberg, 1996), and STS of monkeys and humans (e.g., Vaina et al., 2001) are modeled by Gaussian Radial Basis Functions (RBF). The highest level of the form pathway is formed by motion pattern detectors that temporally integrate and smooth the activity of all snapshot neurons that represent the same movement pattern. The motion pathway of the model has the same hierarchical architecture which contributes to achieve the recognition of biological movements, by analyzing the optic flow pattern. The local motion detectors corresponding to direction-selective neurons in V1 and motion-selective neurons in area MT comprise the first level of this hierarchy. The local structure of the optic-flow fields induced by movement stimuli is evaluated by neurons with larger receptive fields in the second level. Equivalent to the form pathway snapshot neurons that are trained by RBFs, the optic-flow pattern neurons on the next level are selective for more complex optic flow patterns that appear at each individual pattern of biological movement. Like the form pathway, the highest level of that hierarchy is comprised of motion pattern neurons the integrate and smooth temporally the output signals of the optic-flow pattern neurons. This hierarchical (deep) visual recognition model which was originally developed as model for V1-MT cortical processing, later extended extended to applications in computer vision (e.g., Jhuang et al., 2007; Schindler et al., 2008; Abdul-Kreem, 2019).

Artificial systems for action processing have received considerable attention over the past few decades. The first implemented models for action processing and mirror systems were indeed based on artificial neural networks. The seminal models (Oztop and Arbib, 2002; Bonaiuto et al., 2007), represent the circuitry of the action processing network (e.g., F5 mirror system, the STS, and parietal areas such as AIP or IPS [intraparietal sulcus], or LIP [lateral intraparietal cortex]). This model architecture that was primary introduced to model visual feedback for grasping of objects (Fagg et al., 1998) and used a classical training scheme similar to backpropagation, accomplished the recognition of grips and trajectory prediction. This model that was fully implemented, introduced the concept of hand state which consists of sequence of locations and grasp shapes of hand in correspondence to the target object is calculated in STS to provide feedback for visually directed grasping. Performing a reach action, F5 mirror neurons learn an association of the sequence of motor signals and the evolving hand state that makes action recognition possible by evoking F5 activity during action observation. This detailed computational model that demonstrated the temporal sequence of hand states can then be exploited to recognize the action inspired many other schemes for this purpose and extended further the study of the mirror system by introducing and examining the novel hypotheses in the context of artificial neural networks (e.g., Oztop et al., 2006; Bonaiuto and Arbib, 2010; Schrodt et al., 2014).

The recent surge of interest in deep learning methods is due to the fact that in recent years, ”deep learning architectures” or ”convolutional neural networks” have been shown to outperform previous state-of-the-art techniques in several computer vision problems, first in object detection (e.g., LeCun et al., 2015; Ouyang et al., 2017), but later also in motion tracking (e.g., Doulamis and Voulodimos, 2016; Doulamis, 2017), action recognition (e.g., Karpathy et al., 2014; Lin et al., 2016; Tacchetti et al., 2017; Yang et al.,

2019), and human pose estimation (e.g., Brau and Jiang, 2016; Zhu et al., 2018). In spite of this great success, there are many nonbiological features of deep (multilayer) learning such as applied filter kernels, regularization by ”drop out”, or backpropagation algorithm in current supervised training process that requires massive amounts of labeled data, and non-local learning for changing the weights. These, and many other details that are not neurobiologically plausible in architecture of deep learning networks, have made machine learning researchers and neuroscientists to examine the similarities in the computational properties between deep neural networks and human brain (e.g., Illing et al., 2016).
Notwithstanding that the increasing interest in neuron-like architectures with local learning rules motivated by the current advances in neuromorphic hardware, has led to new elaborate models for biologically plausible variants of deep learning (e.g., Lake et al., 2017; Tavanaei et al., 2019; for review see also Nawrocki et al., 2016), the degree of the correlation between the computational properties of the deep neural networks and those of the human brain remains unclear. Therefore, further review of deep learning methods for action recognition will not necessarily provide more insights to our understating of action processing in real cortical neurons. However, some recent comprehensive reviews of this topic are highly encouraged to be read (e.g., Yao et al., 2019; Serre, 2019).

3. Agent Navigation Models

Although path planing and navigation (see Figure.5) can be sometimes as survival as chasing prey, humans and other animals select paths leading to their targets effortlessly through a complex environment. To understand this fundamental ability, a good deal of research has been carried out to model the behavioral dynamics of locomotion and to design of artificial behaving autonomous agents. However, aligned with the focus of this thesis, this section gives only a simplified review of the dynamical systems models of visually-guided locomotor behavior.
Towards an understanding of the visual control system of the fly, in 1976 Reichardt and Poggio (Reichardt and Poggio, 1976), provided a quantitative analysis of navigation model that describes mathematically how a fly steers toward moving targets which they chase as part of their mating behavior. This was the first study that described the orientation behavior of an autonomous agent using a dynamical system with an attractor at the direction in which targets lie. However, a detailed navigation behavior cannot be described based only on target acquisition. To address this problem, Schöner and Dose (Schöner and Dose, 1992, Schöner et al., 1995) provided a dynamical system framework that integrates the target acquisition and obstacle avoidance for navigation and exploration of an autonomous agent. This dynamical model consists of a system of differential equations with attractors and repellers that correspond to goals and obstacles. In this seminal paper, authors have developed a model that seeks to understand the behavioral

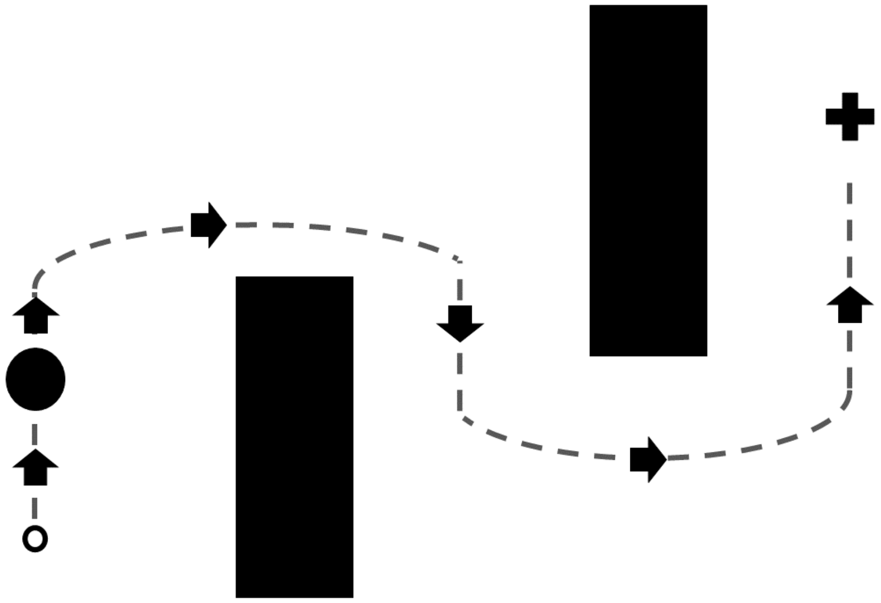


Figure 5: Path planing and navigation: from the starting point (small circle), the autonomous agent has to navigate through walls (indicated in black) by avoiding obstacles towards goal (marked as cross). Dash line indicates the planned path.

dynamics of locomotion and shows the time evolution of behavior as a robot interacts with its environment. In this dynamical system based approach, the current state of the system, as well as intended and avoided states, can be expressed as (sets of) points in the space of behavioral variables, while trajectories of the agent through this space define the behavior.

Inspired by the approach of this paper (Schöner et al., 1995), a new investigation over visually-guided locomotion in such a dynamical framework was proposed to compute a potential field over the robot heading (Fajen et al., 2003). Researchers in this study, developed a biologically-inspired model deriving from experiments on human walking, to identify a set of behavioral variables for steering and obstacle avoidance. They ultimately demonstrated that a successful route selection can be accounted for by the on-line steering dynamics, without explicit path planning. The main difference between this method with its preceding ones was to exploit angular acceleration ($\ddot{\phi}$) rather than angular velocity ($\dot{\phi}$) motivated from measurements of human walking. The resulting model showed smoother and more efficient route selection with continuous curvatures through a cluttered environment. However, this model later was extended (Huang et al., 2006) to use the angular width of an obstacle instead of its distance, and to accommodate obstacles of finite width. More importantly the new extended version of the model was able to use speed control to guarantee that the robot will not collide with an obstacle. This method was introduced, however, for a single agent to track a moving target while avoiding collision with moving obstacle, and was not suitable for applications that required multiple agents. In a new scheme presented by Yan et al. in 2010, an interpotential field, determined by the relative positions among agents, was introduced to address the problems of target tracking and obstacle avoidance for multi-agent systems.

4. The Focus of This Thesis

The aim of this thesis is to study, model and hopefully provide an account that explains how humans understand the actions of other people. First we focused on modeling dynamic neural processes underlying the bidirectional link between action perception and action execution. As it was shown previously action perception and the control of action execution are intrinsically linked in the human brain. Experiments show that concurrent motor execution influences the visual perception of actions and biological motion. This interaction likely is mediated by action-selective neurons in the STS, premotor and parietal cortex. We aimed to answer the question how action execution and action perception might be coupled in the brain and how they influence on each other. Therefore, we developed a model that is based on electrophysiologically plausible mechanisms. It combines mechanisms from previous models that accounted separately for electrophysiological results from action recognition and the neural encoding of motor programs. We demonstrated that our model provides a unifying account for multiple experiments on the interaction between action execution and action perception.

The second part of this PhD work studies and models perception of animacy and social interaction. As it was shown previously, humans derive spontaneously judgments about agency and social interactions from strongly impoverished stimuli, as impressively demonstrated by the seminal work by Heider and Simmel (1944). The neural circuits that derive such judgments from image sequences are entirely unknown. It has been hypothesized that this visual function is based on high-level cognitive processes, such as probabilistic reasoning. Taking an alternative approach, we show that such functions can be accomplished by relatively elementary neural networks that can be implemented by simple physiologically plausible neural mechanisms, exploiting an appropriately structure hierarchical (deep) neural model of the visual pathway.

In order to investigate further the fundamental basis of the neural encoding of social intent and semantics the creation of appropriate stimulus sets for humans and monkeys is unavoidable. Synthesis of such stimuli used in the animacy and social interaction perception studies is a challenging task. The handmade displays like that of the Heider and Simmels has rich motion features but are not quantizable. Also, the experiments generally need many displays with varying cues, and it is difficult to mass produce the handmade displays. The goal of last part of this work is to present a dynamical model that can generate different classes of social interactions controlling the dynamics of the most important factors of social interaction perception namely speed and motion direction. Although still recent studies argue that the usage of artificial displays for the animacy perception lack the rich motion features and hence cannot capture the natural looking motions and interactions, we will show that our generated videos have been validated by human subjects in a psychophysical experiment. These displays have the advantage that the motion cues are quantizable and can be controlled precisely with tweaking parameters of the model and more importantly one can generate arbitrary number of videos per social interaction type.

5. Overview of The Thesis

This thesis is concerned with introducing new physiologically plausible neural models for the coupling of action perception and execution as well as social interaction perception as one of the main application of action perception. Following lines provide a brief overview of the major studies in this PhD work and a summary of presented papers in the subsequent chapters.

In the first study (Chapter 2: ”Neurodynamical Model for the Coupling of Action Perception and Execution”). In this work we developed a model that is based on neural representations of different motor actions by mutually coupled neural fields. One field model represents the perceived action (vision field), and the other one the associated motor program (motor field). Input stimulus pattern for the vision field is a traveling input peak that is derived from 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. For the implementation, we used only the form-pathway which analyzes biological movements by recognizing sequences of ’snapshots’ of body shapes. Two pairs of neural fields of this type were then integrated within an agent that realizes coupled distributed representations of visual and motor patterns. Both fields are reciprocally coupled by interaction kernels that result in a mutual excitation of the fields if the traveling solutions are at the same position along the field, and which induce inhibition if the peak positions strongly differ. As consequence, the motor representation enhances the activity in the visual field when the motor peak propagates with the same speed and phase as the observed visual input. Finally, visual stimuli are reconstructed through a pathway that provides learning of linear neural net-works that map the peak neural activity of the motor field at position onto the joint angles (key poses) within a body-centered frame of reference.

We used the model to reproduce the results of a several experiments that focus on the action perception cycle and mirror neurons. Since the model parameters were identical for all simulations, it thus provides a unifying quantitative account for the experimental. The model reproduces correctly the interactions between action observation and execution in several experiments and provides a link towards electrophysiological detailed models of relevant circuits.

In the second study (Chapter 3: ”Neural model for the visual recognition of animacy and social interaction”) we extended the focus of the study to understand and show that how animacy and social interaction can be perceived in human brain by introducing and developing a neural model that can classifies different social interaction types based mainly on motion cues.

Extending classical biologically-inspired models for object and action perception (Riesenhuber and Poggio, 1999; Giese and Poggio, 2003) by a front-end that exploits deep learning for the construction of low and mid-level feature detectors, we built a hierarchical neural model that reproduces elementary psychophysical results on animacy and social perception from abstract stimuli. The lower hierarchy levels of the model consist of position-variant neural feature detectors that extract orientation and intermediately complex shape features. The next-higher level is formed by shape-selective neurons that are not completely position-invariant, which extract the 2D positions and orientation of moving agents. A second pathway extracts the 2D motion of the moving agents. Exploiting a gain-field network, we compute the relative positions of the moving agents. The top layers of the model combine the mentioned features into more complex high-level features that represent the speed, smoothness of motion and spatial relationships of the moving agents. The highest level of the model consists of neurons that have learned to classify the agency of the motions, and different categories of social interactions.

Based on input video sequences, the model successfully reproduces results of Scholl and Tremoulet, 2000 on the dependence of perceived animacy on motion parameters, and its dependence on the alignment of motion and body axis. The model reproduces the fact that a moving figure that has a body axis, like a rectangle, results in stronger perceived animacy than a circle if the movement, and that the rating is highest if the body axis is aligned with the motion than if it is not aligned. In addition, the model correctly classifies six categories of social interactions that have been frequently tested in the psychophysical literature (following, fighting, chasing, playing, guarding, and flirting). Using simple physiologically plausible neural circuits, the model accounts simultaneously for a variety of effects related to animacy and social interaction perception. Even in its simple form the model proves that animacy and social interaction judgments partly can be derived by very elementary operations within a hierarchical neural vision system, without a need of sophisticated probabilistic inference mechanisms. The model makes precise predictions about the tuning properties of different types of neurons that should be involved in the visual processing of such stimuli. Such predictions might serve as starting point for physiological experiments that investigate the correlate of the perceptual processing of animacy and interaction at the single-cell level.

In the third study (Chapter 4: ”A Generative Model for the Interaction of Two Moving Agents”), we modeled, using previous human navigation models, the interaction of two moving agents. Running a psychophysics study, we showed that our model can generate an arbitrary number of videos for at least twelve distinctive classes of social interaction. In order to model the interaction of two moving agents we exploited a dynamical systems approach, which before was used very successfully for the modelling of human navigation (Fajen et al., 2003). The original approach focuses on mathematical formalization of reactive control for autonomous robots using differential equations that specify attractors and repellors for behavioral variables that control the agent's heading direction and speed. These displays have the advantage that the motion cues are quantizable and can be controlled precisely with tweaking parameters of the model and more importantly one can generate arbitrary number of videos per social interaction type. This study together with the previous one provide both a descent training dataset and a recognition model that might explain the social interaction perception in human brain.

Bibliography

ABDOLLAHI, R. O., J. JASTORFF, AND G. A. ORBAN (2013): “Common and Segregated Processing of Observed Actions in Human SPL,” *Cerebral Cortex (New York, N.Y.: 1991)*, 2734–53.

ABDUL-KREEM, L. I. (2019): “Computational Architecture of a Visual Model for Biological Motions Segregation,” *Network: Computation in Neural Systems*, 40(1-4), 58–78.

ADOLPHS, R., D. TRANEL, AND A. R. DAMASIO (2003): “Dissociable Neural Systems for Recognizing Emotions,” *Brain and Cognition*, 52(1), 61–69.

AMARI, S. (1977): “Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields,” *Biological Cybernetics*, 27(2), 77–87.

AVIEZER, H., Y. TROPE, AND A. TODOROV (2012): “Body Cues, Not Facial Expressions, Discriminate between Intense Positive and Negative Emotions,” *Science (New York, N.Y.*, 338(6111), 1225–29.

BAKER, C. L., R. SAXE, AND J. B. TENENBAUM (2009): “Action Understanding as Inverse Planning,” *Cognition, Reinforcement learning and higher cognition*, 113(3), 329–49.

BARRETT, H. C., P. M. TODD, G. F. MILLER, AND P. W. BLYTHE (2005): “Accurate Judgments of Intention from Motion Cues Alone: A Cross-Cultural Study,” *Evolution and Human Behavior*, 26(4), 313–31.

BLYTHE, P. W., P. M. TODD, AND G. F. MILLER (1999): “How Motion Reveals Intention: Categorizing Social Interactions,” *Simple Heuristics That Make Us Smart*, Evolution and Cognition, 275–85.

BONAIUTO, J. AND M. A. ARBIB (2010): “Extending the Mirror Neuron System Model, II: What Did I Just Do? A New Role for Mirror Neurons,” *Biological Cybernetics*, 102(4), 341–59.

BONAIUTO, J., E. ROSTA, AND M. ARBIB (2007): “Extending the mirror neuron system model, I. Audible actions and invisible grasps,” *Biological Cybernetics*, 96, 9–38.

BONDA, E., M. PETRIDES, D. OSTRY, AND A. EVANS (1996): "Specific Involvement of Human Parietal Systems and the Amygdala in the Perception of Biological Motion," *Journal of Neuroscience*, 3737–44.

BOUQUET, C. A., V. GAURIER, T. SHIPLEY, L. TOUSSAINT, AND Y. BLANDIN (2007): "Influence of the Perception of Biological or Non-Biological Motion on Movement Execution," *Journal of Sports Sciences*, 51930.

BRAU, E. AND H. JIANG (2016): "3D Human Pose Estimation via Deep Learning from 2D Annotations," *Fourth International Conference on 3D Vision*, 582–91.

BRUCE, C., R. DESIMONE, AND C. G. GROSS (1981): "Visual Properties of Neurons in a Polysensory Area in Superior Temporal Sulcus of the Macaque," *Journal of Neurophysiology*, 369–84.

BRUCKER, B., A.-C. EHLIS, F. B. HÄUSSINGER, A. J. FALLGATTER, AND P. GERJETS (2015): "Watching Corresponding Gestures Facilitates Learning with Animations by Activating Human Mirror-Neurons: An FNIRS Study," *Learning and Instruction*, 24–37.

CAGGIANO, V., F. FLEISCHER, J. K. POMPER, M. A. GIESE, AND P. THIER (2016): "Mirror Neurons in Monkey Premotor Area F5 Show Tuning for Critical Features of Visual Causality Perception," *Current Biology*, 3077–82.

CAGGIANO, V., L. FOGASSI, G. RIZZOLATTI, A. CASILE, M. A. GIESE, AND P. THIER (2012): "Mirror Neurons Encode the Subjective Value of an Observed Action," *Proceedings of the National Academy of Sciences*, 1184853.

CALMELS, C., M. ELIPOT, AND L. NACCACHE (2018): "Probing Representations of Gymnastics Movements: A Visual Priming Study," *Cognitive Science*, 152951.

CALVO-MERINO, B., S. EHRENBERG, D. LEUNG, AND P. HAGGARD (2010): "Experts See It All: Configural Effects in Action Observation," *Psychological Research PRPF*, 400406.

CALVO-MERINO, B., J. GRZES, D. E. GLASER, R. E. PASSINGHAM, AND P. HAGGARD (2006): "Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation," *Current Biology*, 190510.

CAMPBELL, M. E. J. AND R. CUNNINGTON (2017): "More than an Imitation Game: Top-down Modulation of the Human Mirror System," *Neuroscience and Biobehavioral Reviews*, 195–202.

CASILE, A., V. CAGGIANO, AND P. F. FERRARI (2011): "The Mirror Neuron System: A Fresh View," *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 52438.

CASILE, A. AND M. A. GIESE (2006): "Nonvisual Motor Training Influences Biological Motion Perception," *Current Biology*, 6974.

CATMUR, C., E. L. THOMPSON, O. BAIRAKTARI, F. LIND, AND G. BIRD (2018): "Sensorimotor Training Alters Action Understanding," *Cognition*, 1014.

CHAQUET, J. M., E. J. CARMONA, AND A. FERNNDEZ-CABALLERO (2013): "A Survey of Video Datasets for Human Action and Activity Recognition," *Computer Vision and Image Understanding*, 117(6), 633–59.

CHRISTENSEN, A., W. ILG, AND M. A. GIESE (2011): "Spatiotemporal Tuning of the Facilitation of Biological Motion Perception by Concurrent Motor Execution." *Journal of Neuroscience*, 31, 3493–3499.

CISEK, P. AND J. F. KALASKA (2010): "Neural Mechanisms for Interacting with a World Full of Action Choices," *Annual Review of Neuroscience*, 33(1), 269–98.

CSIBRA, G. (2008): "Goal Attribution to Inanimate Agents by 6.5-Month-Old Infants," *Cognition*, 107(2), 705–17.

CSIBRA, G., G. GERGELY, S. BÍRÓ, O. KOÓS, AND M. BROCKBANK (1999): "Goal Attribution without Agency Cues: The Perception of Pure Reason in Infancy," *Cognition*, 72(3), 237–67.

CSIBRA, G. AND V. SOUTHGATE (2009): "Inferring the outcome of an ongoing novel action at 13 months," *Developmental Psychology*, 45, 1794–1798.

CURTIS, C. E. AND D. LEE (2010): "Beyond Working Memory: The Role of Persistent Activity in Decision Making," *Trends in Cognitive Sciences*, 14(5), 216–22.

DASSER, V., I. ULBAEK, AND D. PREMACK (1989): "The Perception of Intention," *Science*, 243(4889), 365–67.

DEEN, B., K. KOLDEWYN, N. KANWISHER, AND R. SAXE (2015): "Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus," *Cerebral Cortex*, 25(11), 4596–4609.

DITTRICH, W. H. AND S. E. G. LEA (1994): "Visual Perception of Intentional Motion," *Perception*, 23(3), 253–268.

DOLCOS, S., K. SUNG, J. J. ARGO, S. FLOR-HENRY, AND F. DOLCOS (2012): "The Power of a Handshake: Neural Correlates of Evaluative Judgments in Observed Social Interactions," *Journal of Cognitive Neuroscience*, 24(12), 2292–2305.

DONNARUMMA, F., M. COSTANTINI, E. AMBROSINI, K. FRISTON, AND G. PEZZULO (2017): "Action Perception as Hypothesis Testing," *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 89, 46–50.

DOULAMIS, N. (2017): "Adaptable deep learning structures for object labeling/tracking under dynamic visual environments," *Multimedia Tools and Applications*, 1–39.

DOULAMIS, N. AND A. VOULODIMOS (2016): "FAST-MDL: Fast Adaptive Supervised Training of multi-layered deep learning models for consistent object tracking and classification," *Proceedings of the 2016 IEEE International Conference on Imaging Systems and Techniques, IST*, October, 318–323.

EIMAS, P. D. AND P. C. QUINN (1994): "Studies on the Formation of Perceptually Based Basic-Level Categories in Young Infants," *Child Development*, 65(3), 903–17.

ERLHAGEN, W. AND G. SCHÖNER (2002): "Dynamic Field Theory of Movement Preparation," *Psychological Review*, 109(3), 545–72.

ETZEL, J. A., N. VALCHEV, V. GAZZOLA, AND C. KEYSERS (2016): "Is Brain Activity during Action Observation Modulated by the Perceived Fairness of the Actor?" *PLOS ONE*, 107–25.

FAGG, A. H., , AND M. A. ARBIB (1998): "Modeling Parietal-Premotor Interactions in Primate Control of Grasping," *The Official Journal of the International Neural Network Society*, 11(7-8), 1277–1303.

FAGIOLI, S., B. HOMMEL, AND R. I. SCHUBOTZ (2007): "Intentional Control of Attention: Action Planning Primes Action-Related Stimulus Dimensions," *Psychological Research*, 2229.

FAJEN, B. R., W. H. WARREN, S. TEMIZER, AND L. P. KAELBLING (2003): "A Dynamical Model of Visually-Guided Steering, Obstacle Avoidance, and Route Selection," *International Journal of Computer Vision*, 54(1), 13–34.

FREITAG, C. M., C. KONRAD, M. HÄBERLEN, C. KLESER, A. VON GONTARD, W. REITH, N. F. TROJE, AND C. KRICK (2008): "Perception of Biological Motion in Autism Spectrum Disorders," *Neuropsychologia*, 1480–94.

FRISTON, K., J. MATTOUT, AND J. KILNER (2011): "Action Understanding and Active Inference," *Biological Cybernetics*, 137–60.

FRITH, C. D. AND U. FRITH (2012): "Mechanisms of Social Cognition," *Annual Review of Psychology*, 287–313.

FRÜHHOLZ, S., C. HOFSTETTER, C. CRISTINZIO, A. SAJ, M. SEECK, P. VUILLEUMIER, AND D. GRANDJEAN (2015): "Asymmetrical Effects of Unilateral Right or Left Amygdala Damage on Auditory Cortical Processing of Vocal Emotions," *Proceedings of the National Academy of Sciences*, 112(5), 1583–88.

GALLESE, V., L. FADIGA, L. FOGASSI, AND G. RIZZOLATTI (1996): "Action recognition in the premotor cortex," *Brain*, 593–609.

GALLESE, V. AND A. GOLDMAN (1998): "Mirror Neurons and the Simulation Theory of Mind-Reading," *Trends in Cognitive Sciences*, 493–501.

GALLESE, V., C. KEYSERS, AND G. RIZZOLATTI (2004): "A Unifying View of the Basis of Social Cognition," *Trends in Cognitive Sciences*, 8, 396–403.

GAZZOLA, V. AND C. KEYSERS (2009): "The Observation and Execution of Actions Share Motor and Somatosensory Voxels in All Tested Subjects: Single-Subject Analyses of Unsmoothed FMRI Data," *Cerebral Cortex (New York, N.Y.: 1991)*, 1239–55.

GEIGER, A., G. BENTE, S. LAMMERS, D. ROTH, D. BZDOK, AND K. VOGELEY (2019): "Distinct Functional Roles of the Mirror Neuron System and the Mentalizing System," *NeuroImage*.

GERGELY, G., Z. NADASDY, G. CSIBRA, AND S. BÍRÓ (1995): "Taking the Intentional Stance at 12 Months of Age," *Cognition*, 56(2), 165–93.

GIBSON, J. J. (1852): *Medicinische Psychologie oder Physiologie der Seele Weidmann*, Leipzig: Weidmann.

——— (1966): *The Senses Considered as Perceptual Systems*, London: Allen and Unwin.

GIESE, M. A. AND T. POGGIO (2003): "Neural Mechanisms for the Recognition of Biological Movements," *Nature Reviews. Neuroscience*, 4(3), 179–92.

GIESE, M. A. AND G. RIZZOLATTI (2015): "Neural and Computational Mechanisms of Action Processing: Interaction between Visual and Motor Representations," *Neuron*, 167–80.

GREEN, D., Q. LI, J. J. LOCKMAN, AND G. GREDEBÄCK (2016): "Culture influences action understanding in infancy: prediction of actions performed with chopsticks and spoons in Chinese and Swedish infants," *Child development*, 87(3), 736–746.

GROSSMAN, E., M. DONNELLY, R. PRICE, D. PICKENS, V. MORGAN, G. NEIGHBOR, AND R. BLAKE (2000): "Brain areas involved in perception of biological motion," *Journal of Cognitive Neuroscience*, 711–720.

HAMILTON, A., D. WOLPERT, AND U. FRITH (2004): "Your Own Action Influences How You Perceive Another Persons Action," *Current Biology*, 49398.

HAN, Y., P. ZHANG, W. HUANG, AND Z. Y (2015): "Going Deeper with Two-Stream ConvNets for Action Recognition in Video Surveillance," *Pattern Recognition Letters, Video Surveillance-oriented Biometrics*, 107(5), 83–90.

HAUSER, M. D. (1998): “A Nonhuman Primate’s Expectations about Object Motion and Destination: The Importance of Self-Propelled Movement and Animacy,” *Developmental Science*, 1(1), 31–37.

HEIDER, F. AND M. L. SIMMEL (1944): “An Experimental Study of Apparent Behavior,” *The American Journal of Psychology*, 10, 243–259.

HERBART, J. (1852): *Psychologie als Wissenschaft neu gegrndet auf Erfahrung*, Unzer, Knigsberg: Metaphysik und Mathematik.

HEYES, C. (2010): “Where Do Mirror Neurons Come From?” *Neuroscience and Biobehavioral Reviews*, 575–83.

HICKOK, G. (2013): “Do Mirror Neurons Subserve Action Understanding?” *Neuroscience Letters*, 56 –58.

HOMMEL, B., J. MÜSSELER, G. ASCHERSLEBEN, AND W. PRINZ (2001): “The Theory of Event Coding (TEC): A Framework for Perception and Action Planning,” *Behavioral and Brain Sciences*, 84978.

HUANG, W. H., B. R. FAJEN, J. R. FINK, AND W. H. WARREN (2006): “Visual Navigation and Obstacle Avoidance Using a Steering Potential Function,” *Robotics and Autonomous Systems*, 54(4), 288–99.

HUBEL, D. H. AND T. N. WIESEL (1962): “Receptive Fields, Binocular Interaction and Functional Architecture in the Cats Visual Cortex,” *The Journal of Physiology*, 2, 106–154.

IACOBONI, M., M. D LIEBERMAN, B. J KNOWLTON, I. MOLNAR-SZAKACS, M. MORITZ, C. J. THROOP, AND A. P. FISKE (2004): “Watching Social Interactions Produces Dorsomedial Prefrontal and Medial Parietal BOLD FMRI Signal Increases Compared to a Resting Baseline,” *NeuroImage*, 21(3), 1167–73.

ILLING, B., W. GERSTNER, AND J. BREA (2016): “Biologically Plausible Deep Learning But How Far Can We Go with Shallow Networks?” *Neural Networks*, 118(1), 90–101.

ISIK, L., A. MYNICK, D. PANTAZIS, AND N. KANWISHER (2019): “The Speed of Human Social Interaction Perception,” *BioRxiv*, March, 579375.

JACOB, P. (2009): “A Philosophers Reflections on the Discovery of Mirror Neurons,” *Topics in Cognitive Science*, 570–95.

JACOBS, A. AND M. SHIFFRAR (2005): “Walking Perception by Walking Observers,” *Journal of Experimental Psychology: Human Perception and Performance*, 15769.

JARRETT, C. (2012): "Mirror Neurons: The Most Hyped Concept in Neuroscience?" *Psychology Today*.

JELLEMA, T., C., B. WICKER, AND D. I. PERRETT (2000): "Neural Representation for the Perception of the Intentionality of Actions," *Brain and Cognition*, 280–32.

JHUANG, H., T. SERRE, L. WOLF, T. POGGIO, AND IEEE (2007): "A biologically inspired system for action recognition," *Ieee 11th International Conference on Computer Vision*, 4(3), 1253–1260.

JOHNSON, S. (2000): "The Recognition of Mentalistic Agents in Infancy," *Trends in Cognitive Sciences*, 4(1), 22–28.

KADUK, K., B. ELSNER, AND V. M. REID (2013): "Discrimination of Animate and Inanimate Motion in 9-Month-Old Infants: An ERP Study," *Developmental Cognitive Neuroscience*, 6(10), 14–22.

KAHL, S. AND S. KOPP (2018): "A Predictive Processing Model of Perception and Action for Self-Other Distinction," *Frontiers in Psychology*, 137–60.

KARPATHY, A., G. TODERICI, S. SHETTY, T. LEUNG, R. SUKTHANKAR, AND L. FEI-FEI (2014): "Large-Scale Video Classification with Convolutional Neural Networks," *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society*, 1725–1732.

KENNEDY, D. P. AND R. ADOLPHS (2012): "The Social Brain in Psychiatric and Neurological Disorders," *Trends in Cognitive Sciences*, 16, 559–72.

KEYSERS, C. (2011): *The Empathic Brain: How the Discovery of Mirror Neurons Changes Our Understanding of Human Nature*, CreateSpace Independent Publishing Platform.

KEYSERS, C. AND V. GAZZOLA (2014): "Hebbian Learning and Predictive Mirror Neurons for Actions, Sensations and Emotions," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 20130175.

KEYSERS, C. AND D. I. PERRETT (2004): "Demystifying Social Cognition: A Hebbian Perspective," *Trends in Cognitive Sciences*, 201–7.

KILNER, J. AND R. LEMON (2013): "What We Know Currently about Mirror Neurons," *Current Biology*, R1057–62.

KILNER, J. M., K. J. FRISTON, AND C. D. FRITH (2007): "The Mirror-Neuron System: A Bayesian Perspective," *Neuroreport*, 18, 619–23.

KILNER, J. M., Y. PAULIGNAN, AND S. J. BLAKEMORE (2003): "An Interference Effect of Observed Biological Movement on Action," *Current Biology*, 52225.

LAHNAKOSKI, J. M., E. GLEREAN, J. SALMI, I. P. JÄÄSKELÄINEN, R. SAMS, AND L. NUMMENMAA (2012): "Naturalistic FMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception," *Frontiers in Human Neuroscience*, 6(223), 263–75.

LAKE, B. M., T. D. ULLMAN, J. B. TENENBAUM, AND S. J. GERSHMAN (2017): "Building Machines That Learn and Think like People," *Behavioral and Brain Sciences*, 40.

LAURIE, C., C. ASSAIANTE, B. NAZARIAN, J.-L. ANTON, AND C. SCHMITZ (2011): "Recruitment of Both the Mirror and the Mentalizing Networks When Observing Social Interactions Depicted by Point-Lights: A Neuroimaging Study," *Journal of Cognitive Neuroscience*, 6(1), e15749.

LECUN, Y., Y. BENGIO, AND G. HINTON (2015): "Deep Learning," *Nature*, 521(7553), 436–44.

LEE, D. AND S. LEE (2011): "Vision-Based Finger Action Recognition by Angle Detection and Contour Analysis," *ETRI Journal*, 33(3), 415–22.

LESLIE, A. M. (1995): "A Theory of Agency. In Causal Cognition: A Multidisciplinary Debate," *Symposia of the Fyssen Foundation. New York, NY, US: Clarendon Press/Oxford University Press*, 121–49.

LIN, L., K. WANG, W. ZUO, M. WANG, J. LUO, AND L. ZHANG (2016): "A deep structured model with radius-margin bound for 3D human activity recognition," *International Journal of Computer Vision*, 118(2), 256–273.

LINDEMANN, O., P. STENNEKEN, H. T. V. SCHIE, AND H. BEKKERING (2006): "Semantic Activation in Action Planning," *Journal of Experimental Psychology: Human Perception and Performance*, 63343.

LOGOTHETIS, N. K. AND D. L. SHEINBERG (1996): "Visual Object Recognition," *Annual Review of Neuroscience*, 19, 577–621.

MCALEER, P. AND F. E. POLLICK (2008): "Understanding Intention from Minimal Displays of Human Activity," *Behavior Research Methods*, 40(3), 830–39.

MEIJER, H. G. AND S. COOMBES (2014): "Travelling Waves in Models of Neural Tissue: From Localised Structures to Periodic Waves," *EPJ Nonlinear Biomedical Physics*, 2(1): 3.

MEIRING, G. A. AND H. C. MYBURGH (2015): "A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms," *Sensors*, 15(12), 30653–30682.

MIALL, R. C. (2003): "Connecting mirror neurons and forward models," *NeuroReport*, 2135–2137.

MICHOTTE, A. (1963): *The Perception of Causality*, England: Oxford, Basic Books.

MILLER, E. K. (2000): "Beyond Working Memory: The Role of Persistent Activity in Decision Making," *Nature Reviews. Neuroscience*, 1(1), 59–65.

MOLENBERGHS, P., R. CUNNINGTON, AND J. B. MATTINGLEY (2012): "Brain Regions with Mirror Properties: A Meta-Analysis of 125 Human FMRI Studies," *Neuroscience and Biobehavioral Reviews*, 341–49.

MUKAMEL, R., A. D. EKSTROM, J. KAPLAN, M. IACOBONI, AND I. FRIED (2010): "Single-Neuron Responses in Humans during Execution and Observation of Actions," *Current Biology*, 750–56.

MÜSSELER, J. AND B. HOMMEL (1997): "Blindness to Response-Compatible Stimuli," *Journal of Experimental Psychology: Human Perception and Performance*, 86172.

NAKAMURA, K., R. KAWASHIMA, N. SATO, A. NAKAMURA, M. SUGIURA, T. KATO, AND K. HATANO (2000): "Functional Delineation of the Human Occipito-Temporal Areas Related to Face and Scene ProcessingA PET Study," *Brain*, 123(9), 1903–12.

NAWROCKI, R. A., R. M. VOYLES, AND S. E. SHAHEEN (2016): "A Mini Review of Neuromorphic Architectures and Implementations," *IEEE Transactions on Electron Devices*, 63(10), 3819–29.

NELISSEN, K., E. BORRA, M. GERBELLA, S. ROZZI, G. LUPPINO, W. VANDUFFEL, G. RIZZOLATTI, AND G. A. ORBAN (2011): "Action Observation Circuits in the Macaque Monkey Cortex," *The Journal of Neuroscience*, 3743–56.

ORAM, M. W. AND D. I. PERRETT (1994): "Responses of Anterior Superior Temporal Polysensory (STPa) Neurons to Biological Motion Stimuli," *Journal of Cognitive Neuroscience*, 99–116.

——— (1996): "Integration of Form and Motion in the Anterior Superior Temporal Polysensory Area (STPa) of the Macaque Monkey," *Journal of Neurophysiology*, 109–29.

OUYANG, W., X. ZENG, AND X. WANG (2017): "DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7), 1320–1334.

OZTOP, E. AND M. ARBIB (2002): "Schema design and implementation of the grasp-related mirror neuron system," *Biological Cybernetics*, 15(12), 116–140.

OZTOP, E., M. KAWATO, AND M. ARBIB (2006): "Mirror Neurons and Imitation: A Computationally Guided Review," *Neural Networks, The Brain Mechanisms of Imitation Learning*, 19(3), 254–71.

PELLEGRINO, G. D., L. FADIGA, L. FOGASSI, V. GALLESE, AND G. RIZZOLATTI (1992): "Understanding Motor Events: A Neurophysiological Study," *Experimental Brain Research*, 176–80.

PERRETT, D. I. AND M. W. ORAM (1993): "Neurophysiology of Shape Processing," *Image and Vision Computing*, 11(6), 317–33.

PERRETT, D. I., P. A. SMITH, A. J. MISTLIN, A. J. CHITTY, A. S. HEAD, D. D. POTTER, R. BROENNIMANN, A. D. MILNER, AND M. A. JEEVES (1985a): "Visual Analysis of Body Movements by Neurones in the Temporal Cortex of the Macaque Monkey: A Preliminary Report," *Behavioural Brain Research*, 153–70.

PREMACK, D. (1990): "The Infants Theory of Self-Propelled Objectsy," *Cognition*, 36(1), 1–16.

PRESS, C., N. WEISKOPF, AND J. KILNER (2012): "Dissociable Roles of Human Inferior Frontal Gyrus during Action Execution and Observation," *NeuroImage*, –671–77.

PRINZ, W. (1997): "Perception and Action Planning," *European Journal of Cognitive Psychology*, 12954.

PUCE, A. AND D. PERRETT (2003): "Electrophysiology and Brain Imaging of Biological Motion," *Philosophical Transactions of the Royal Society of London*, 435–45.

REICHARDT, W. AND T. POGGIO (1976): "Visual Control of Orientation Behaviour in the Fly. Part I. A Quantitative Analysis," *Quarterly Reviews of Biophysics*, 9(3), 311–375.

RENNELS, J. L., J. JUVRUD, A. J. KAYL, M. ASPERHOLM, G. GREDEBCK, AND A. HERLITZ (2017): "Caregiving Experience and Its Relation to Perceptual Narrowing of Face Gender," *Developmental Psychology*, 53(8), 1437–46.

RIESENHUBER, M. AND T. POGGIO (1999): "Hierarchical Models of Object Recognition in Cortex," *Nature Neuroscience*, 2(2), 1019–25.

——— (2002): "Neural Mechanisms of Object Recognition," *Current Opinion in Neurobiology*, 12(2), 162–68.

RIZZOLATTI, G. AND L. CRAIGHERO (2004): "The Mirror-Neuron System," *Annual Review of Neuroscience*, 169–92.

RIZZOLATTI, G., L. FADIGA, V. GALLESE, AND L. FOGASSI (1996): "Premotor Cortex and the Recognition of Motor Actions," *Cognitive Brain Research, Mental representations of motor acts*, 131–41.

RIZZOLATTI, G. AND L. FOGASSI (2014): "The Mirror Mechanism: Recent Findings and Perspectives," *Biological Sciences*.

RIZZOLATTI, G., L. FOGASSI, AND V. GALLESE (2001): "Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action," *Nature Reviews Neuroscience*, 661–70.

RIZZOLATTI, G. AND C. SINIGAGLIA (2010): "The Functional Role of the Parieto-Frontal Mirror Circuit: Interpretations and Misinterpretations," *Nature Reviews Neuroscience*, 264–74.

——— (2016): "The Mirror Mechanism: A Basic Principle of Brain Function," *Nature Reviews Neuroscience*, 757–65.

ROCHAT, P., R. MORGAN, AND M. CARPENTER (1997): "Young Infants Sensitivity to Movement Information Specifying Social Causality," *Cognitive Development*, 12(4), 537–61.

ROUSSEL, C., G. HUGHES, AND F. WASZAK (2013): "A Preactivation Account of Sensory Attenuation," *Neuropsychologia*, 92229.

RUTHERFORD, M. D., B. F. PENNINGTON, AND S. J. ROGERS (2006): "The Perception of Animacy in Young Children with Autism," *Journal of Autism and Developmental Disorders*, 36(8), 983–92.

SAVAKI, H. E. AND V. RAOS (2019): "Action Perception and Motor Imagery: Mental Practice of Action," *Progress in Neurobiology*, 107–25.

SCHINDLER, K., L. V. GOOL, AND B. DE GELDER (2008): "Recognizing emotions expressed by body pose: a biologically inspired neural model," *Neural nwtworks*, 21, 1238–1246.

SCHLOTTMANN, A., E. D. RAY, A. MITCHELL, AND N. DEMETRIOU (2006): "Perceived Physical and Social Causality in Animated Motions: Spontaneous Reports and Ratings," *Acta Psychologica*, 123(1-2), 112–43.

SCHLOTTMANN, A. AND E. RAY (2010): "Goal Attribution to Schematic Animals: Do 6-Month-Olds Perceive Biological Motion as Animate?" *Developmental Science*, 13(1), 1–10.

SCHOLL, B. J. AND P. D. TREMOULET (2000): "Perceptual Causality and Animacy," *Trends in Cognitive Sciences*, 4(8), 299–309.

SCHÖNER, G. AND M. DOSE (1992): "A dynamical systems approach to task-level system integration used to planand control autonomous vehicle motion," *Quarterly Reviews of Biophysics*, 10(4), 253–267.

SCHÖNER, G., M. DOSE, AND C. ENGELS (1995): "Dynamics of behavior: Theory and applications for autonomous robot architectures," *Robotics and Autonomous Systems*, 16, 213–245.

SCHÖNER, G. AND J. SPENCER (1966): *Dynamic Thinking: A Primer on Dynamic Field Theory*, London: Oxford University Press.

SCHRODT, F., G. LAYHER, H. NEUMANN, AND M. BUTZ (2014): "Modeling Perspective-Taking upon Observation of 3D Biological Motion," *Proceedings of the Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics, IEEE*, 102(4), 305–310.

SCHÜTZ-BOSBACH, S., B. MANCINI, S. M. AGLIOTI, AND P. HAGGARD (2006): "Self and Other in the Human Motor System," *Current Biology*, 17, 1830–34.

SCHÜTZ-BOSBACH, S. AND W. PRINZ (2007): "Perceptual Resonance: Action-Induced Modulation of Perception," *Trends in Cognitive Sciences*, 349–55.

SELTZER, B. AND D. N. PANDYA (1994): "Parietal, Temporal, and Occipital Projections to Cortex of the Superior Temporal Sulcus in the Rhesus Monkey: A Retrograde Tracer Study," *Journal of Comparative Neurology*, 445–63.

SERRE, T. (2019): "Deep Learning: The Good, the Bad, and the Ugly," *Annual Review of Vision Science*, 5(1), 399–426.

SERRE, T., L. WOLF, AND T. A. POGGIO (2005): "Object Recognition with Features Inspired by Visual Cortex," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2(2), 994–1000.

SERVOS, P., R. OSU, A. SANTI, AND M. KAWATO (2002): "The Neural Substrates of Biological Motion Perception: An FMRI Study," *Cerebral Cortex*, 772–82.

SINIGAGLIA, C. (2013): "What Type of Action Understanding Is Subserved by Mirror Neurons," *Neuroscience Letters, The Mirror Neuron System*, 540(4), 59–61.

SLIWA, J. AND W. A. FREIWALD (2017): "A Dedicated Network for Social Interaction Processing in the Primate Brain," *Science*, 356(6339), 745–49.

SOKOLOV, A. A., A. GHARABAGHI, M. S. TATAGIBA, AND M. PAVLOVA (2010): "Cerebellar Engagement in an Action Observation Network," *Cerebral Cortex*, 486–91.

SOKOLOV, A. A., R. C. MIALL, AND R. B. IVRY (2017): "The Cerebellum: Adaptive Prediction for Movement and Cognition," *Trends in Cognitive Sciences*, 313–32.

SOUSA, E., W. ERLHAGEN, F. FERREIRA, AND E. BICHO (2015): "Off-Line Simulation Inspires Insight: A Neurodynamics Approach to Efficient Robot Task Learning," *Neural Networks, Neurobiologically Inspired Robotics: Enhanced Autonomy through Neuromorphic Cognition*, 72(12), 123–39.

SZEGO, P. A. AND M. D. RUTHERFORD (2008): "Dissociating the Perception of Speed and the Perception of Animacy: A Functional Approach," *Evolution and Human Behaviors*, 29(5), 335–42.

TACCHETTI, A., L. ISIK, AND T. POGGIO (2017): "Invariant Recognition Drives Neural Representations of Action Sequences," *PLOS Computational Biology*, 13(2), e1005859.

TAVANAEI, A., M. GHODRATI, S. R. KHERADPISHEH, T. MASQUELIER, AND A. MAIDA (2019): "Deep Learning in Spiking Neural Network," *Neural Networks*, 111(3), 47–63.

TAVARES, P., A. D. LAWRENCE, AND P. J. BARNARD (2008): "Paying Attention to Social Meaning: An FMRI Study," *Cerebral Cortex*, 18(8), 1876–85.

THOMASCHKE, R., R. C. MIALL, M. RUESS, P. R. MEHTA, AND B. HOPKINS (2018): "Visuomotor and Motorvisual Priming with Different Types of Set-Level Congruency: Evidence in Support of Ideomotor Theory, and the Planning and Control Model (PCM)," *Psychological Research*, 107390.

THOMPSON, E. L., G. BIRD, AND C. CATMUR (2019): "Conceptualizing and Testing Action Understanding," *Neuroscience and Biobehavioral Reviews*, 105(10), 106–14.

TODOROV, A., S. G. BARON, AND N. N. OOSTERHOF (2008): "Evaluating Face Trustworthiness: A Model Based Approach," *ocial Cognitive and Affective Neuroscience*, 3(2), 119–27.

TRÄUBLE, B., S. PAUEN, AND D. POULIN-DUBOIS (2014): "Speed and Direction Changes Induce the Perception of Animacy in 7-Month-Old Infants," *Frontiers in Psychology*, 5(10), 213–245.

TRAUTMANN, S., T. F. ALEXA, AND M. HERRMANN (2009): “Emotions in Motion: Dynamic Compared to Static Facial Expressions of Disgust and Happiness Reveal More Widespread Emotion-Specific Activations,” *Brain Research*, 1284(11), 100–115.

TREMOULET, P. D. AND J. FELDMAN (2006): “The Influence of Spatial Context and the Role of Intentionality in the Interpretation of Animacy from Motion,” *Perception and Psychophysics*, 68(6), 1047–58.

TSAKIRIS, M. AND P. HAGGARD (2005): “Experimenting with the Acting Self,” *Cognitive Neuropsychology*, 22(3-4), 387–407.

UNGERLEIDER, L. G. AND M. MISHKIN (1982): “In Analysis of Visual Behavior,” *MIT Press, Cambridge, Massachusetts*, 549–586.

VAINA, L. M., J. SOLOMON, S. CHOWDHURY, P. SINHA, AND J. W. BELLIVEAU (2001): “Functional Neuroanatomy of Biological Motion Perception in Humans,” *Proceedings of the National Academy of Sciences*, 98(20), 11656–61.

VANGENEUGDEN, J., M. V. PEELEN, D. TADIN, AND L. BATTELLI (2014): “Distinct Neural Mechanisms for Body Form and Body Motion Discriminations,” *Journal of Neuroscience*, 574–85.

VANMARCKE, S., S. VAN DE CRUYS, P. MOORS, AND J. WAGEMANSS (2017): “Intact Animacy Perception during Chase Detection in ASD,” *Scientific Reports*, 7(9).

VRIGKAS, M., C. NIKOU, AND I. A. KAKADIARIS (2015): “A Review of Human Activity Recognition Methods,” *Frontiers in Robotics and AI*, 2, 28.

WALBRIN, J., P. DOWNING, AND K. KOLDEWYN (2018): “Neural Responses to Visually Observed Social Interactions,” *Neuropsychologia*, 112(4), 31–39.

WALBRIN, J., I. MIHAI, J. LANDSIEDEL, AND K. KOLDEWYN (2020): “Developmental Changes in Visual Responses to Social Interactions,” *Developmental Cognitive Neuroscience*, 42(4), 100774.

WILSON, H. R. AND J. D. COWAN (1973): “A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue,” *Kybernetik*, 13(2), 55–80.

WOLPERT, D. M. AND Z. GHAHRAMANI (2000): “Computational Principles of Movement Neuroscience,” *Nature Neuroscience*, 121217.

WOLPERT, D. M. AND M. KAWATO (1998): “Multiple Paired Forward and Inverse Models for Motor Control,” *The Official Journal of the International Neural Network Society*, 131729.

WYKOWSKA, A. AND A. SCHUBÖ (2012): "Action Intentions Modulate Allocation of Visual Attention: Electrophysiological Evidence," *Frontiers in Psychology*.

YAN, J., X.-P. GUAN, AND F.-X. TAN (2010): "Target Tracking and Obstacle Avoidance for Multi-Agent Systems," *International Journal of Automation and Computing*, 7(4), 550–56.

YANG, D. Y. J., G. ROSENBLAU, C. KEIFER, AND K. A. PELPHREY (2015): "An Integrative Neural Model of Social Perception, Action Observation, and Theory of Mind," *Neuroscience and Biobehavioral Reviews*, 51(4), 263–75.

YANG, H., C. YUAN, B. LI, Y. DU, J. XING, W. HU, AND S. J. MAYBANK (2019): "Asymmetric 3D Convolutional Neural Networks for Action Recognition," *Pattern Recognition*, 85(1), 1–12.

YAO, G., T. LEI, AND J. ZHONG (2019): "A Review of Convolutional-Neural-Network-Based Action Recognition," *Pattern Recognition Letters, Cooperative and Social Robots: Understanding Human Activities and Intentions*, 118(2), 14–22.

ZHANG, K. (1996): "Representation of Spatial Orientation by the Intrinsic Dynamics of the Head-Direction Cell Ensemble: A Theory," *Journal of Neuroscience*, 16(6), 2112–26.

ZHU, A., T. WANG, AND H. SNOUSSI (2018): "Hierarchical Graphical-Based Human Pose Estimation via Local Multi-Resolution Convolutional Neural Network," *AIP Advances*, 8(3), 035215.

ZHU, H., R. VIAL, AND S. LU (2017): "A spatio-temporal convolutional regression network for video action proposal," *In Proceedings of the CVPR*, 43(3)16, 5813–5821.

ZWICKEL, J., M. GROSJEAN, AND W. PRINZ (2007): "Seeing While Moving: Measuring the Online Influence of Action on Perception," *The Quarterly Journal of Experimental Psychology*, 106371.

Statement of Contributions

This thesis comprises three manuscripts two of which are published and one is being submitted. The author of this dissertation 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.

Chapter I. Hovaidi-Ardestani M., Caggiano V., Giese M. (2017). **Neurodynamical Model for the Coupling of Action Perception and Execution**. In: Lecture Notes in Computer Science Artificial Neural Networks and Machine Learning volume 10613. Hovaidi-Ardestani designed, implemented and performed all experiments, analyzed data and wrote the submitted manuscript. Caggaino supported with providing data. Giese conceived the initial study, supported design of simulations, data analysis and writing of submitted manuscript.

Chapter II. Hovaidi-Ardestani M., Saini N., Martinez A.M., Giese M.A. (2018). **Neural Model for the Visual Recognition of Animacy and Social Interaction**. In: Lecture Notes in Computer Science Artificial Neural Networks and Machine Learning volume, vol 11141. Hovaidi-Ardestani designed, implemented the model and performed all the simulations, analyzed data and wrote the submitted manuscript. Saini supported partially with the design of the model. Martinez supported with writing of submitted manuscript. Giese provided the initial design, data analysis and writing of submitted manuscript.

Chapter III. Hovaidi-Ardestani M., Salatiello A., Giese M.A. (to be submitted). **A Dynamical Generative Model of Social Interactions in Virtual Agents**. Hovaidi-Ardestani designed, implemented the model and performed all the simulations, conducted the psychophysics experiment and wrote the submitted manuscript. Salatiello supported with writing of submitted manuscript. Giese provided data analysis and writing of submitted manuscript.

Following conferences are the selection of conferences in which parts of this work were presented.

1. Hovaidi-Ardestani M., Giese M. A. (2016). Spikingneuron model for the interaction between visual and motor representations of action in premotor cortex. ECVP, 28.Aug-1. Sep, Barcelona, Spain
2. Hovaidi-Ardestani M., Giese M. A. (2016). Spiking model for the interaction between action recognition and action execution. CNS, Jeju, S. Korea

3. Hovaidi-Ardestani M., Giese M. A. (2017). Biophysically plausible neural model for the interaction between visual and motor representations of action. VSS Conference, 19-24 May, St.Petersburg, Florida, Journal of Vision. 2017; 17(10):1167-1167

4. Hovaidi-Ardestani M., Giese M. A. (2018). Neural Model for the Recognition of Agency and Interaction from Motion. VSS Conference, 18-23 May, St.Petersburg, Florida, Journal of Vision September 2018, 18 , 430

5. Hovaidi-Ardestani M., Saini N., Martinez A., Giese, M. A. (2018). Neural model for the visual recognition of animacy and social interaction. 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III, 168-177.

6. Hovaidi-Ardestani M., Saini N., Martinez A., Giese, M. A. (2019). Neural model for the visual recognition of agency and social interaction. ECVP Conference 2019, Perception 48(2S),104

Neurodynamical Model for the Coupling of Action Perception and Execution

(Published In (2017): Lecture Notes in Computer Science Artificial Neural Networks and Machine Learning volume 10613)

Neurodynamical Model for the Coupling of Action Perception and Execution

Mohammad Hovaidi-Ardestani[1,2(✉)], Vittorio Caggiano[3], and Martin Giese[1]

[1] Section of Computational Sensomotorics, Department of Cognitive Neurology,
CIN and HIH, University Clinic Tübingen, Ottfried-Müller-Str. 25,
72076 Tübingen, Germany
Mohammad.Hovaidi-Ardestani@uni-tuebingen.de

[2] IMPRS for Cognitive and Systems Neuroscience, Tübingen, Germany

[3] Computational Biology Center, IBM T.J. Watson Research Center, 1101
Kitchawan Road, Route 134, Room 30-048, 10598 Yorktown Heights, USA

Abstract. In cortical representations action perception and action execution are closely linked, as indicated by the presence of mirror neurons. Experiments show that concurrent action execution and action perception influence each other. We have developed a physiologically-inspired neural model that accounts for the neural encoding of perceived actions and motor plans, and their interactions. The core of the model is a set of coupled neural fields that represent either perceived actions or motor programs. We demonstrate that this model reproduces the results of a variety of quite different experiments investigating the interaction between action perception and execution. It also predicts the emergence and stability of synchronized coordinated behavior of two individuals that observe each other during action execution.

Keywords: Action perception · Motor program · Neural field · Recurrent neural network · Mirror neurons

1 Introduction

Perceptual and motor representations of actions are tightly coupled (e.g. [1]). This is supported by many results from behavioral and functional imaging studies, and physiologically by the existence of mirror neurons, e.g. in premotor and parietal cortex [2,3]. Behavioral and functional imaging studies show influences of motor execution on simultaneous action perception as well as influences in the opposite direction (e.g. [4–6]). Physiological data provides insights in the basis of the encoding of actions at the single-cell level [2,7,8]. This has motivated the development of neural models that account for action perception (e.g. [9,10]) as well as for the neural encoding of motor programs (e.g. [11]). Multiple conceptual models have been proposed that discuss the interaction between action perception and execution (e.g. [12–14]). Some implemented models have been proposed for these interactions in the context of robot systems (e.g. [15]).

A. Lintas et al. (Eds.): ICANN 2017, Part I, LNCS 10613, pp. 19–26, 2017.
https://doi.org/10.1007/978-3-319-68600-4_3

We describe here a model that is based on electrophysiologically plausible mechanisms. It combines mechanisms from previous models that accounted separately for electrophysiological results from action recognition and the neural encoding of motor programs [9,16,17]. We demonstrate that our model provides a unifying account for multiple experiments on the interaction between action execution and action perception. The model might thus provide a starting point for the detailed quantitative investigation how motor plans interact with perceptual action representations at the level of single-cell mechanisms.

2 Model Architecture

The architecture of our model is illustrated in Fig. 1. The core of the model is a set of dynamically coupled neural fields that encode visually perceived actions and motor programs (Fig. 1B). Each encoded action is represented by a pair of neural fields, a *motor field* encoding the associated motor program, and a *vision field* that represents the visually perceived action. Within these fields the evolving action is represented by a stable traveling pulse solution that runs along the field. The different fields are dynamically coupled in a way that enforces a synchronization of the traveling peaks between the vision and motor field that encode the same action. Fields encoding different actions inhibit each other. The vision fields receive a feed-forward input from a visual pathway that recognizes shapes from gray-level images (Fig. 1A). The motor fields are read out by a neural network that models the motor pathway and produces joint angle trajectories that correspond to the evolving action. These angles are used to animate an avatar, which is rendered to produce an image sequence or movie that shows the action (C). The architecture thus models motor execution as well as action recognition. The following sections describe the individual components of the model in further detail.

2.1 Neural Vision and Motor Fields

The model assumes that individual actions can be encoded as visual patterns, or as motor program. Neurally, the patterns are encoded as stable traveling pulse solutions in dynamic neural fields. For the simulations in this paper these fields are defined over periodic spaces ($x, y \in [-\pi, \pi]$). We assume the encoding of M different actions (where M was 2 for the simulations). The vision field that encodes the precept of action m (assuming $1 \leqslant m \leqslant M$) is driven by an input signal distribution $s^m(x,t)$, which is produced by the output neurons of the visual pathway that are tuned for body postures of the action pattern m. The temporal evolution of the activation $u^m(x,t)$ of this visual field is determined by the neural field equation [18]:

$$\frac{\tau \partial u^m(x,t)}{\partial t} = -u^m(x,t) - h + w_u(x) * F(u^m(x,t)) + s^m(x,t) + c_u^m(x,t) \quad (1)$$

with the nonlinear saturationg threshold function $F(u) = d_0 \left(1 - \exp(u^2/2d_1)\right)$ for $u > 0$, and $F(u) = 0$ otherwise, and $h > 0$ determining the resting level activity. As interaction kernel we chose the asymmetric function: $w_u(x) = -a_0 + a_1(\frac{1+\cos(x-a_3)}{2})^{\gamma}$ with $\gamma > 0$. The convolution operator is defined by $f(x) * g(x) = \int_{-\pi}^{\pi} f(x')g(x-x')dx'$. With this kernel for appropriate choice of the parameters, a traveling-pulse input signal $s^m(x,t)$ induces a traveling pulse equilibrium solution that moves synchronously with the input. This solution breaks down if the frames of the input movies appear in inverse or random temporal order [9]. The term $c_u^m(x,t)$ summarizes the inputs from the other fields and is further specified below.

The corresponding motor program is encoded by another neural field without feed forward input. It is defined by the equation:

$$\frac{\tau \partial v^m(y,t)}{\partial t} = -v^m(y,t) - h + w_v(y) * F(v^m(y,t)) + c_v^m(y,t). \tag{2}$$

The form of the interaction kernel w_v is identical to the one of w_u with slightly different parameters, resulting in stronger recurrent feedback. As consequence, once a local activation is established by a 'go signal' a self-stabilizing traveling peak solution emerges that propagates with constant speed along the y-dimension [19]. We associate the values of y with the body poses (joint angles) that emerge during the action, so that the traveling pulse encodes the temporal evolution of a motor program. The term $c_v^m(x,t)$ again specifies inputs from the other fields.

2.2 Coupling Structure

The cross connections between the vision and motor fields encoding the same actions were defined by the kernel function:

$$w_{uv}(x,y) = -b_0 + b_1 \left(\frac{1+\cos(x-y)}{2}\right)^{\gamma} = w_{vu}(y,x). \tag{3}$$

This kernel results in a tendency of the activation peaks in both fields to propagate synchronously. The fields encoding different actions are coupled by the cross-inhibition kernel $w_I(x,y) = -c_0$ with $c_0 > 0$. As consequence the different encoded actions compete in the neural representation. Summarizing, the corresponding interaction terms in Eqs. 1 and 2 are given by the relationships

$$c_u^m(x,t) = w_{uv}(x,y) *_y F(v^m(y,t)) + \sum\nolimits_{m' \neq m} w_I(x,y) *_y (F(u^{m'}(y,t) + F(v^{m'}(y,t)))$$

$$c_v^m(x,t) = w_{vu}(x,y) *_y F(u^m(y,t)) + \sum\nolimits_{m' \neq m} w_I(x,y) *_y (F(u^{m'}(y,t) + F(v^{m'}(y,t)))$$

where the operator $*_y$ indicates the convolution with respect to the variable y.

2.3 Vision and Motor Pathway

The input module of our model is given by a vision pathway that recognizes shapes from image sequences (Fig. 1A). This module is taken over form a previous model without motor pathway (see [9] for details). In brief, the vision

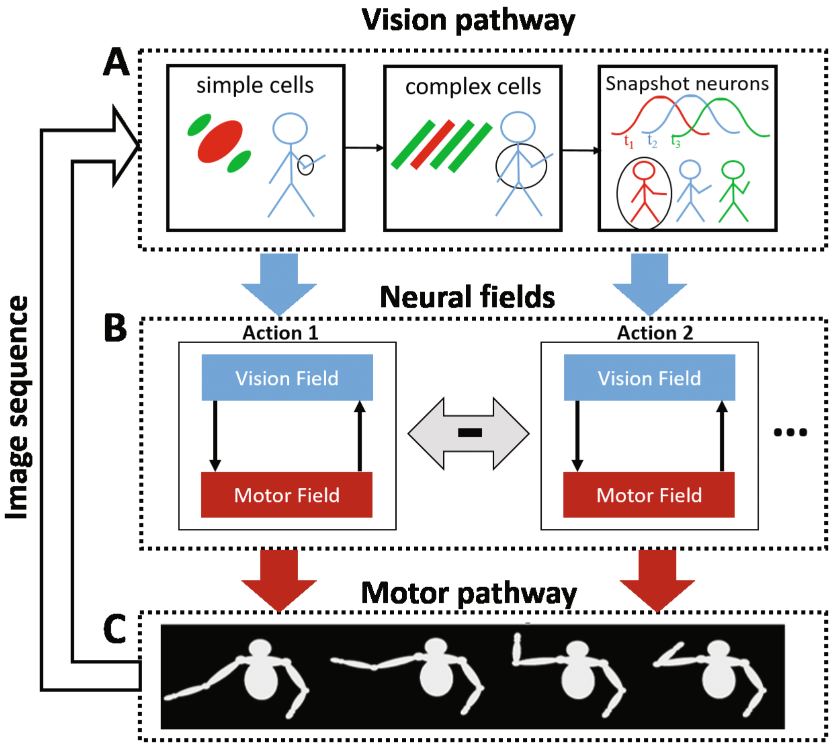


Fig. 1. Overview of the model architecture. **A** The form pathway taken over from a previous neural model [9] drives the input signals for the vision fields from image sequences. **B** The core of the model consists of coupled pairs of vision and motor fields that encode the same action. **C** Motor pathway that reads out the motor fields and generates joint angle trajectories, which are used to animate an avatar, which then can be rendered to produce visual input movies.

pathway consists of a hierarchy of neural shape detectors. The complexity of the extracted features and the position and scaling invariance increase along the hierarchy. The highest level of this pathway is composed from radial basis function (RBF) units that have been trained with snapshots of the learned action movies. These neurons thus detect instantaneous body shapes in image sequences, where the underlying neural network is trained in a supervised manner. Dropping for a moment the index m, assume that the vector $\mathbf{z}(t)$ is formed by the activations of the shape-selective RBF units that encode one particular action pattern at time t, and that the vector $\mathbf{s}(t)$ signifies input signal $s(x, t)$, sampled at a sufficient number of discrete points along the variable x. We learned a linear mapping of the form $\mathbf{s}(t) = \mathbf{R}\mathbf{z}(t)$ between these vectors using sparse regression. Training data pairs consisted of vectors $\mathbf{z}(t)$ of the RBF outputs for equidistantly sampled key frames from the training action movies. Vectors $\mathbf{s}(t)$ were derived from appropriately positioned idealized Gaussian input signals. For learned training patterns the outputs of this linear network define a moving positive input peak, while the input signal $s(x, t)$ remains very small for actions that deviate from the training action. In total, we learned M separate linear mappings from the RBF outputs of the units encoding the keyframes of action m to the corresponding input signal distributions $s^m(x, t)$.

The motor pathway computes joint angles from the position of the activation peak in the motor field along the variable y. This variable parameterizes the temporal evolution of the action. Dropping again the index m, we learned by Support Vector Regression a mapping of the position of the activation peaks $y_{\max}(t) = \arg\max_y v(y,t)$ onto the joint angles of the corresponding body postures. The motor fields encoding different actions compete in a winner-takes-all fashion, and we used only the output of the most activated motor field for the computation of the joint angles. In order to close the loop between action control and perception we used the joint angles to animate an avatar, which then was rendered to produce input movies for the visual pathway.

3 Simulations in Comparison with Experimental Data

We simulated the results of four experiments that studied the interaction between action perception and execution. In the following, simulation results from the model are presented side-by-side with the original data, always using the same model parameters.

(i) Influence of action execution on action perception: In the underlying experiment arm actions were presented as point-light stimuli in noise while the observers performed the same action in a virtual reality setup. The spatio-temporal coherence between the executed and the visually observed action was systematically varied, either by delaying the observed action in time or by rotating it in the image plane relative to the executed action. (See [6] for further details.) Fig. 2A shows a recognition index (RI) that measures the facilitation ($RI > 0$) or inhibition ($RI < 0$) of the visual detection by concurrent motor execution in comparison with a baseline without motor execution. For increasing spatial (Fig. 2A) as well as temporal (Fig. 2B) incoherence between the executed and observed actions the facilitation by concurrent motion execution goes over into an inhibitory interaction. The same behavior is reproduced by our model, simulating the masked point-light stimulus by a noisy traveling input peak (Fig. 2 C, D).

(ii) Influence of action perception on action execution: The underlying experiment measured the variability of motor execution when participants moved their arms periodically in on direction while they saw another person performing a periodic arm movement in the same or in orthogonal direction [4]. As illustrated in Fig. 3A, compared to a baseline without concurrent visual stimulation, the variability of the motor pattern increases when the visually observed arm movement is inconsistent (orthogonal) to the executed pattern. The same increase in variability is obtained from the model (Fig. 3B) (quantified as variability of the timing of the corresponding activation peak in the motor field).

(iii) Spontaneous coordination in multi-person interaction: A classical experiment in interactive sensorimotor control [20] shows that two people that observe each other during the execution of a periodic leg movement tend spontaneously to synchronize their movements. In addition, the variability of the

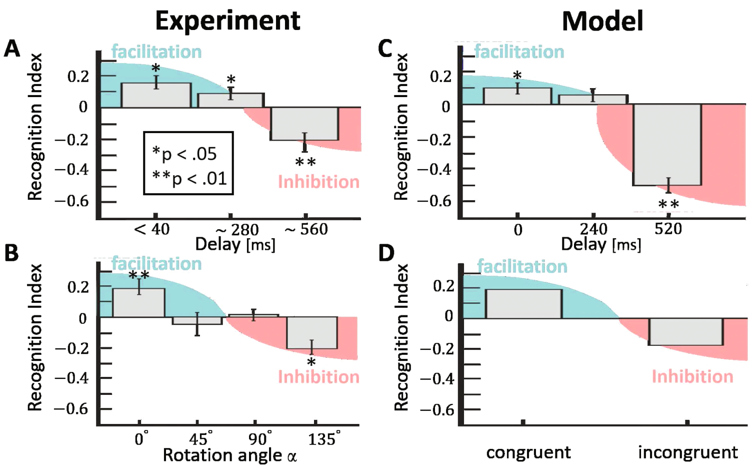


Fig. 2. Influence of concurrent motor execution on the visual detection of action patterns. The experimentally measured Recognition Index (RI) indicates transitions from facilitation to inhibition of visual detection by concurrent motor execution, when the temporal coherence (panel **A**) or the spatial congruence (panel **B**) of the visual pattern with the executed patterns are progressively reduced ([6]). Similar RI computed from the model output shows qualitatively the same behavior (panels **C** and **D**).

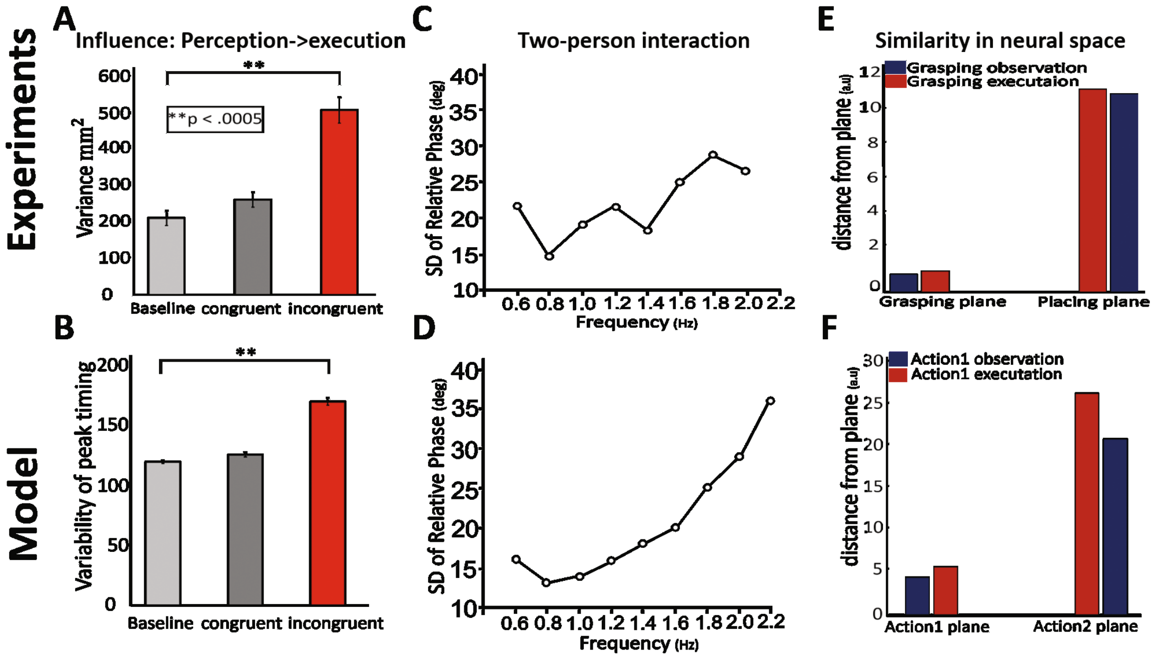


Fig. 3. Reproduction of experimental effects: **A** Motor variability of executed actions increases during observation of incongruent actions [4]. **B** Timing variability of motor peak in the model shows similar behavior. **C** Frequency dependence of standard deviation (SD) of relative phase for the spontaneous synchronization of two agents who observe each other [20]. **D** Corresponding model result derived from activity in motor fields. **E** Neural trajectories for grasping execution and observation are close to 'grasping' plane, but far away from 'placing' plane [8]. **F** Same behavior is observed for the neural trajectories computed from the model neurons. (Details see text.)

relative phase of the synchronized movements is frequency-dependent. Figure 3C shows the original data for the frequency dependence. In order to simulate this interactive behavior of two agents, we implemented two separate models and defined the visual input of either model by the movie that was generated by the motor output of the other. Like in the experiment, the two simulated agents spontaneously synchronize. Figure 3D shows that, in addition, the model predicts correctly frequency dependence of the variability of the relative phase (as consequence of the selectivity of the neural fields for the propagation speed of the moving peaks).

(iv) Reproduction of the population dynamics of F5 mirror neurons: Our last simulation reproduces electrophysiological data from action-selective (mirror) neurons in area F5 [8]. To generate this data, the responses of 489 mirror neurons, relative to the baseline activity, were combined into a population activity vector that varies over time. Using principle components analysis, the dimensionality of the 'neural state space' that is spanned up by these vectors was reduced to three. (Higher-dimensional approximations led to very similar results; see [8] for details.) In this neural state space the trajectories for the execution and observation of a first action ('grasping') were lying close to the same plane, while the trajectory for the observation of another action ('placing') evolved in an orthogonal pane. This is quantified in Fig. 3 E, which illustrates the average distances of the neural trajectories from the planes that fit best the trajectories for the observation of 'grasping' and 'placing'. A very similar topology of the neural trajectories emerges for our model, if we concatenate the activities of all neural field neurons into a population vector and apply the same techniques for dimension reduction (Fig. 3F). Thus neural trajectories for the perception and the execution of the same action are close to the same plane, while neural trajectories for different actions evolve in orthogonal subspaces.

4 Conclusion

The proposed model is consistent with the behavior of action-selective neurons in the superior temporal sulcus and mirror neurons in area F5 of monkeys ([16,17]). It provides a unifying account for a whole spectrum of experiments on the interaction between action perception and execution. Future work needs to give up the strict separation of visual and motor fields, potentially exploiting inhomogeneous neural field models.

Acknowledgments. We thank A. Christensen for helpful comments. Funded by EC, HBP FP7-ICT-2013-FET-F/ 604102, HFSP RGP0036/2016, German Federal Ministry of Education and Research: BMBF, FKZ: 01GQ1002A; Deutsche Forschungsgemeinschaft: DFG GI 305/4-1, DFG GZ: KA 1258/15-1.

References

1. Prinz, W.: Perception and action planning. Eur. J. Cogn. Psychol. **9**, 129–154 (1997)
2. Rizzolatti, G., Fogassi, L., Gallese, V.: Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. **2**, 661–670 (2001)
3. Giese, M.A., Rizzolatti, G.: Neural and computational mechanisms of action processing: Interaction between visual and motor representations. Neuron **88**, 167–180 (2015)
4. Kilner, J.M., Paulignan, Y., Blakemore, S.J.: An interference effect of observed biological movement on action. Curr. Biol. **13**, 522–525 (2003)
5. Calvo-Merino, B., Grèzes, J., Glaser, D.E., Passingham, R.E., Haggard, P.: Seeing or doing? Influence of visual and motor familiarity in action observation. Curr. Biol. **16**, 1905–1910 (2006)
6. Christensen, A., Ilg, W., Giese, M.A.: Spatiotemporal tuning of the facilitation of biological motion perception by concurrent motor execution. J. Neurosci. **31**, 3493–3499 (2011)
7. Barraclough, N.E., Keith, R.H., Xiao, D., Oram, M.W., Perrett, D.I.: Visual adaptation to Goal-directed hand actions. J. Cogn. Neurosci. **21**, 1805–1819 (2009)
8. Caggiano, V., Fleischer, F., Pomper, J.K., Giese, M.A., Thier, P.: Mirror neurons in monkey premotor area F5 show tuning for critical features of visual causality perception. Curr. Biol. **26**, 3077–3082 (2016)
9. Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. **4**, 179–192 (2003)
10. Jhuang, H., Serre, T., Wolf, L., Poggio, T.: A biologically inspired system for action recognition. In: IEEE International Conference on Computer Vision, vol. 1, pp. 1–8 (2007)
11. Chersi, F., Ferrari, P.F., Fogassi, L.: Neuronal chains for actions in the parietal lobe: a computational model. PLoS ONE **6**, e27652 (2011)
12. Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: Codes and their vicissitudes. Behav. Brain Sci. **24**, 910–926 (2001)
13. Wolpert, D.M., Doya, K., Kawato, M.: A unifying computational framework for motor control and social interaction. Philos. Trans. Royal Soc. London B Biol. Sci. **358**, 593–602 (2003)
14. Kilner, J.M., Friston, K.J., Frith, C.D.: The mirror-neuron system: a Bayesian perspective. Neuroreport **18**, 619–623 (2007)
15. Erlhagen, W., Bicho, E.: The dynamic neural field approach to cognitive robotics. J. Neural Eng. **3**, R36 (2006)
16. Cisek, P., Kalaska, J.F.: Neural mechanisms for interacting with a world full of action choices. Annu. Rev. Neurosci. **33**, 269–298 (2010)
17. Fleischer, F., Caggiano, V., Thier, P., Giese, M.A.: Physiologically inspired model for the Visual recognition of transitive hand actions. J. Neurosci. **33**, 6563–6580 (2013)
18. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. Biol. Cybern. **27**, 77–87 (1977)
19. Zhang, K.: Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J. Neurosci. **16**, 2112–2126 (1996)
20. Schmidt, R.C., Carello, C., Turvey, M.T.: Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. J. Exp. Psychol. Hum. Percept. Perform. **16**, 227–247 (1990)

Neural Model for the Visual Recognition of Animacy and Social Interaction

(Published In (2018): Lecture Notes in Computer Science Artificial Neural Networks and Machine Learning volume, vol 11141)

Neural Model for the Visual Recognition of Animacy and Social Interaction

Mohammad Hovaidi-Ardestani[1,2], Nitin Saini[1,2], Aleix M. Martinez[3], and Martin A. Giese[1(✉)]

[1] Section of Computational Sensomotorics, Department of Cognitive Neurology,
CIN and HIH, University Clinic Tübingen,
Ottfried-Müller-Str. 25, 72076 Tübingen, Germany
martin.giese@uni-tuebingen.de

[2] IMPRS for Cognitive and Systems Neuroscience, Tübingen, Germany

[3] Department of Electrical and Computer Engineering, The Ohio State University,
Columbus, OH 43210, USA

Abstract. Humans reliably attribute social interpretations and agency to highly impoverished stimuli, such as interacting geometrical shapes. While it has been proposed that this capability is based on high-level cognitive processes, such as probabilistic reasoning, we demonstrate that it might be accounted for also by rather simple physiologically plausible neural mechanisms. Our model is a hierarchical neural network architecture with two pathways that analyze form and motion features. The highest hierarchy level contains neurons that have learned combinations of relative position-, motion-, and body-axis features. The model reproduces psychophysical results on the dependence of perceived animacy on motion smoothness and the orientation of the body axis. In addition, the model correctly classifies six categories of social interactions that have been frequently tested in the psychophysical literature. For the generation of training data we propose a novel algorithm that is derived from dynamic human navigation models, and which allows to generate arbitrary numbers of abstract social interaction stimuli by self-organization.

Keywords: Hierarchy · Neural network model · Animacy
Social interaction perception

1 Introduction

Humans spontaneously can decode animacy and social interactions from strongly impoverished stimuli. A classical study by Heider and Simmel [1] demonstrated that humans derived very consistently interpretations in terms of social interactions from simple geometrical figures that moved around in the two-dimensional plain. The figures were interpreted as living agents, to which even personality traits were attributed. More recent studies have characterized in more detail which critical features of simple stimuli affect the perception of animacy, that

V. Kůrková et al. (Eds.): ICANN 2018, LNCS 11141, pp. 168–177, 2018.
https://doi.org/10.1007/978-3-030-01424-7_17

is whether the object is perceived as alive [2–4]. Furthermore, detailed studies have focused on the perception of social interactions between multiple moving shapes, e.g. focusing on 'chasing' or 'fighting' [5,6]. Six interaction types have been used in a number of studies [7–9], McAleer and Pollick [9] showed that these categories can be reliably classified from stimuli showing moving circular disks whose movements were derived from real interactions.

Coarse neural substrates of the processing of such stimuli have been identified in fMRI studies. Animacy has been studied, modulating the movement parameters of individual moving shapes [10–12], and stimuli similar to the ones by Heider & Simmel have been frequently used in studies addressing Theory of Mind [13,14]. In fMRI and monkey studies regions like the superior temporal sulcus (STS) and human area TPJ were found to be selective for these stimuli [15–18]. In spite of this localization of relevant cortical areas, the underlying exact neural circuits of this processing remain entirely unclear. Some theories have associated the processing of such abstract stimuli with probabilistic reasoning [19,20], while others have linked them to lower-level visual processing [6]. So far no ideas exist how such functions could be accounted for by physiologically plausible neural circuits.

The goal of this paper is to present a simple neural model that reproduces some of the key observations in psychophysical experiments about the perception of animacy and social interactions from simple abstract stimuli. The model in its present form is simple, but in principle extendable for the processing of more complex stimuli that require also the processing of shape details or shapes in clutter. The model is an extension of classical models of the visual processing stream that account for the processing of object shape and actions [21–24]. However, such models never have been applied to account for the perception of animacy or social interaction. Our attempt to use these types of architectures is motivated by recent work that showed that models of this type for the recognition of hand actions also account for the perception of causality from simple stimulus displays that consist of moving disks [25]. This modeling work predicted also the existence of neurons in macaque cortex that are specifically involved in the visual perception of causality [26]. Here we show that a model based on similar principles accounts for the perception of animacy and social interactions.

In the following section, we first describe how we generated a stimulus set for training of the neural model, devising a generative model for social interaction stimuli that is based on a dynamical systems approach. We then describe the architecture of the model. The following section describes the results, followed by a brief discussion.

2 Stimulus Synthesis

For the training of neural network models a sufficient set of stimuli is required. The problem is that from the classical psychophysical studies only a rather small set of stimuli is publicly available. For a meaningful application of learning-based neural networks approaches thus a sufficiently large training data set with similar properties needs to be generated. In our study we used movies showing individual moving agents, and interaction of 2 agents (chasing, playing, following, flirting, guarding, fighting) described in psychophisical studies [7–9].

In order to model the interaction of two moving agents we exploited a dynamical systems approach, which before was used very successfully for the modeling of human navigation [27]. The underlying idea, originally derived from robotics [28], is to define a dynamical systems or differential equations for the heading directions ϕ_i and the instantaneous propagation speeds v_i of the interacting agents (in our case $i = 1, 2$). The specified movement is dependent on goal and obstacle points in the two dimensional plain, where the other agent can also act as goal or obstacle as well. We modified a model for human steering behavior during walking [29] to reproduce the movements during social interactions.

The resulting dynamics is given by the following differential equations for the heading direction:

$$\begin{aligned}\ddot{\phi}_i &= -b\dot{\phi}_i - k_g(\phi_i - \psi_{\mathrm{g},i})(e^{-c_1 d_{\mathrm{g},i}} + c_2)\\ &\quad + k_o \sum_{n=1}^{N_{\mathrm{obst}}} (\phi_i - \psi_{\mathrm{o},ni})(e^{-c_3|\phi_i - \psi_{\mathrm{o},ni}|})(e^{-c_4 d_{\mathrm{o},ni}}).\end{aligned} \tag{1}$$

The variables $\psi_{\mathrm{g},i}$ and $d_{\mathrm{g},i}$ signify the absolute direction of the actual goal point and the distance of the goal from the agent in the 2D plain. Likewise, $\psi_{\mathrm{o},ni}$ and $d_{\mathrm{o},ni}$ signify the absolute direction and distance from obstacle number n from the agent, where N_{obst} is the number of relevant obstacles, and where k_m and c_m signify constants. The forward speed of the agents is specified by the two stochastic differential equations

$$\tau\dot{v}_i = -v_i + F_i(d_{\mathrm{g},i}) + k_\epsilon \epsilon_i(t), \tag{2}$$

where $\epsilon_i(t)$ is Gaussian white noise. The two functions F_i that specify the distance dependence of the speed dynamics are different for the two agents:

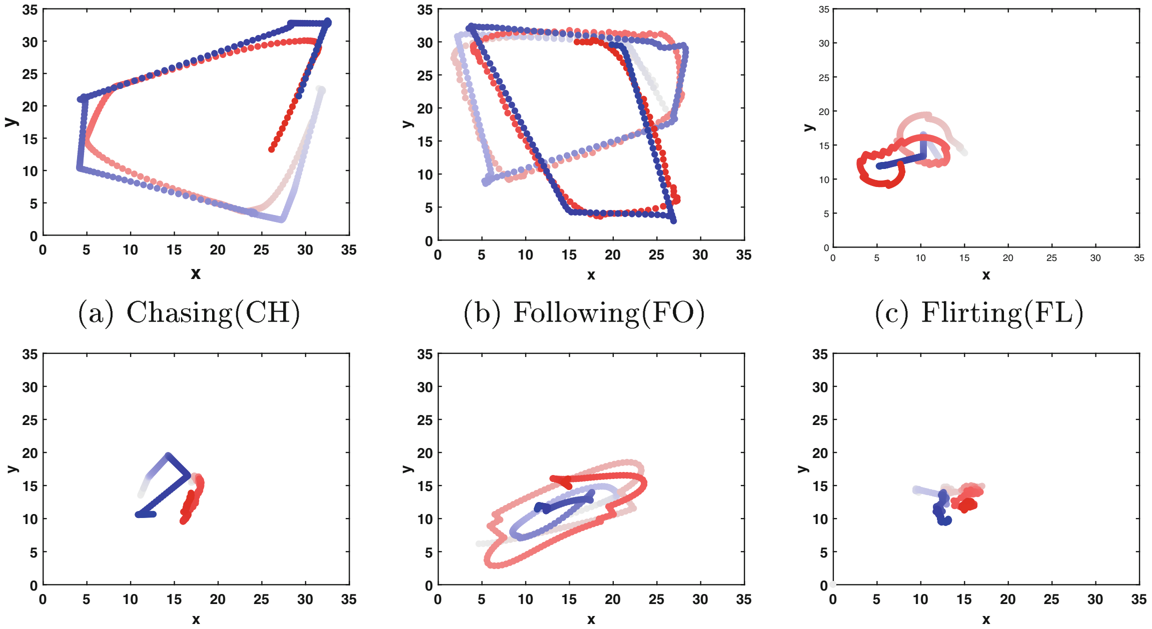
$$F_1(d) = \frac{1}{1 + e^{-c_5(d - c_6)}} - c_7 e^{-kd} \tag{3}$$

$$F_2(d) = \frac{c_8}{1 + e^{-c_9(d - c_{10})}} - c_{11} e^{-kd} + c_{12}. \tag{4}$$

The goal point of the second agent was typically the first agent. The goal points for the first agent was given by a sequence of fixed positions, which were randomly generated by uniformly sampling from the 2D plain and rejecting the samples that were closer than a fixed distance from the last sample. Since it turned out that the influence of the obstacle terms was rather low for the speed dynamics, we dropped the obstacle terms from the speed control dynamics. Table 1 provides an overview of the model parameters for the six simulated behaviors. We generated 50 stimuli for each interaction class. Figure 1 shows examples paths of the agents for the different behaviors for typical simulations.

Table 1. Parameters of simulation algorithm.

	Agent 1					Agent 2						
	k_ε	C_5	C_6	C_7	k	k_ε	C_8	C_9	C_{10}	C_{11}	C_{12}	k
Guarding (Gu)	0	1	5	0	0	0	1	1	3	0	0.5	0
Following (FO)	0	10	7	0	0	0	1	4	4	0	0	0
Fighting (FI)	1	1	3	1	0.1	1	1	1	3	1	0	0.1
Chasing (CH)	0	10	7	0	0	0	1	1	7	0	0	0
Flirting (FL)	0	1	5	0	0	1	0.6	1	2	1	0	0.5
Playing (PL)	0	1	5	0	0	1	1	1	10	0	0.5	0



(a) Chasing(CH)

(b) Following(FO)

(c) Flirting(FL)

(d) Guarding(GU)

(e) Playing(PL)

(f) Fighting(FI)

Fig. 1. Sample trajectories for 6 different social interactions. Colors indicate the positions of the two agents (agent 1: blue, agent 2: red). Color saturation indicates time, the color fading out after long times. (Color figure online)

3 Model Architecture

An overview of the model architecture is shown in Fig. 2. Building on classical biologically-inspired models for shape and action processing [21,22], the model comprises a form and a motion pathway, each consisting of a hierarchy of feature detectors. Presently, these pathways were modelled following these classical papers, which was sufficient for the tested simple stimuli.

Form Pathway: The form pathway of the simple model implementation here comprises only three hierarchy layers. The first is composed from (even and uneven) Gabor filters with 8 different orientations (cf. [22]), whose centers were placed in a grid of 120 by 120 points across the pixel image. The outputs of

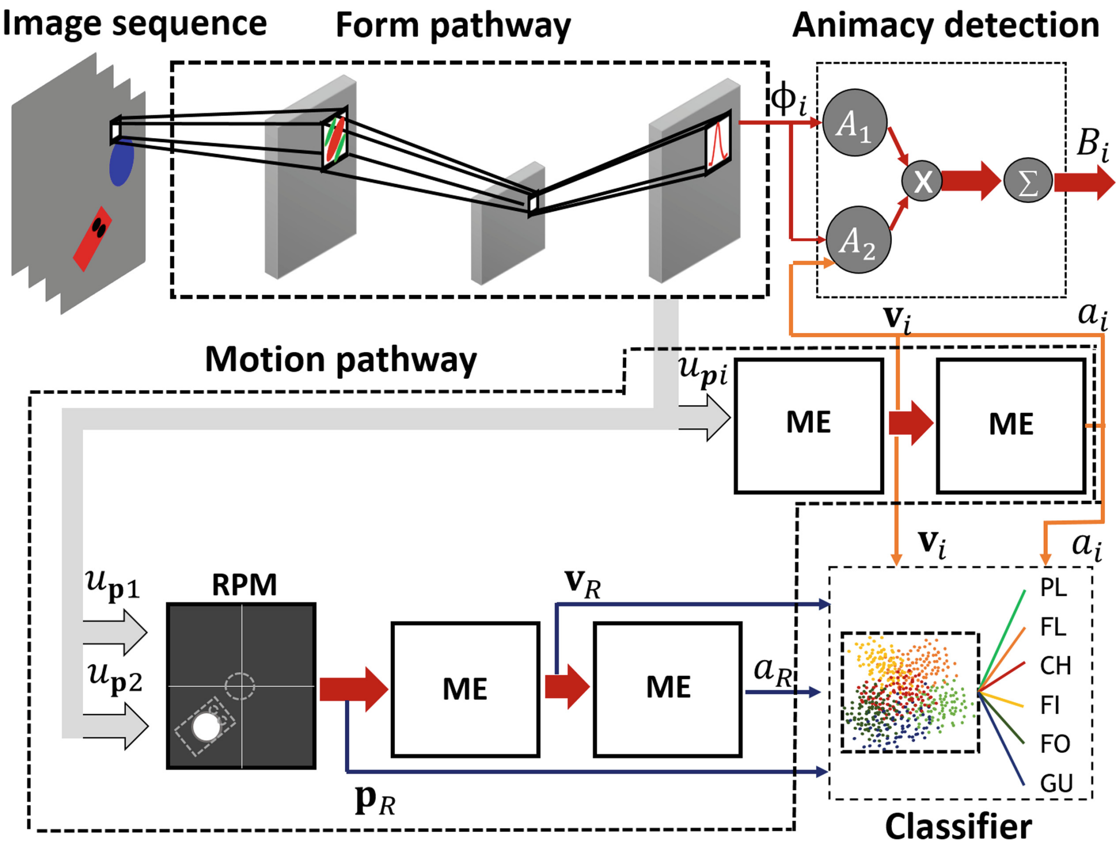


Fig. 2. Model consisting of a form and a motion pathway. ME signifies a layer of motion energy detectors, and RPM the relative position map. The top level of the model is formed by neural detectors for the perceived animacy, and a network that classifies six different types of interactions. (See text for details.)

this Gabor filter array are pooled by the next layer using a maximum operation over a grid of 41 by 41 filters, separately for the different orientations, in order to increase the position-invariance of the representation. The highest layer of the form pathway is formed by Gaussian radial basis function, which are trained with the shapes of the agents in different 2D orientations. Opposed to many other object recognition architectures, these shape-selective neurons have receptive fields of limited size (about 20% of the width of the image), which is consistent with neural data from area IT [30]. The outputs of this layer provide thus information about the identity of the agents, their positions, and their orientation in the image plain. The signal $u_k(\phi, x, y)$ is the output activity of the neural detectors detecting shape k at the 2D position (x, y). Summing this signal over all ϕ provides a neural activity distribution $u_{\mathbf{p}_k}(x, y)$ whose peak signals the position of agent k in the image. This signal is used to compute the velocity and the relative positions of the moving elements or animate objects. Similarly, by summing over the positions one obtains a activity distribution $u_{\phi_k}(\phi)$ over the directions with a peak at ϕ_k.vadjust

Motion Pathway: It analyzes the 2D motion and the relative motion of the moving agents. As input we use the time-dependent signals $u_{\mathbf{p}_k}(x, y)$ for each agent as input to a field of standard motion energy detectors (ME in Fig. 2), resulting in an output that encodes the motion energy in terms of a four-dimensional neural activity distribution (dropping the index k in the following) $u_{\mathbf{v}}(x, y, v_x, v_y, t)$, where $\mathbf{v} = (v_x, v_y)$ is the preferred velocity vector of the motion energy detector. Pooling this output activity distribution over all spatial positions using a maximum operation, a position-invariant neural representation of velocity is obtained. From this a neural representation of motion direction is obtained by pooling this activity distribution over all neurons with the same (similar) motion direction, resulting in a one-dimensional activity distribution $u_\theta(\theta, t)$ over the motion direction θ, from which the direction can be easily estimated by computing a population vector[1]. The same applies to the length of the velocity vector[2] $v = |\mathbf{v}|$. In order to compute also the acceleration of the agents, we transmit the position-invariant activity distribution $u_{\mathbf{v}}(v_x, v_y, t)$ as input to another field of motion energy detectors, which computes from this an energy distribution $u_{\mathbf{a}}(x, y, a_x, a_y, t)$ over the acceleration vectors $\mathbf{a} = (a_x, a_y)$. By pooling over directions, from this an activity distribution over the length of these vectors $a = |\mathbf{a}|$) is computed, and again this parameter can be estimated by a simple population vector. The population estimates of θ, $\mathbf{v}$ and a enter the animacy computation (s.b.).

For analyzing the relative motion of the two agents, following [22], the output distributions $u_{\mathbf{p}_k}(x, y)$ of the form pathway are also fed into a gain field network that computes a representation of the position of the second agent in a coordinate frame that is centered on the first. Its output is computed as convolution-like integral of the form $u_{\mathbf{p}_R}(x, y) = \int_{x', y'} u_{\mathbf{p}_1}(x', y') u_{\mathbf{p}_2}(x + x', y + y')\, \mathrm{d}x' \mathrm{d}y'$. This output defines a neural *relative position map* that represents the position of agent 2 as an activity peak in a coordinate frame that is centered on the first. The integral is taken over a finite region of shifts $|(x, y)| < D$, implying that situations where the agents have a distance substantially larger than D will not produce an output peak. This makes sense since agents that are too distant do not produce the percept of a social interaction. The activity distribution $u_{\mathbf{p}_R}(x, y, t)$ is again processed by a cascade of two levels of motion energy detectors in order to compute the relative speed and acceleration of the two agents. Population estimates of the relative distance $d_R = |\mathbf{p}_R|$, velocity $\mathbf{v}_R$, and the acceleration a_R enter the interaction classifier.

Recognition Level: The highest level of the model consists of a circuit that derives the perceived animacy of the two agents, and another one that classifies the perceived interaction class. The neurons detecting instantaneous animacy (dropping again the index k and time) multiply two input derived from the signal of both pathways signals $B = A_1 A_2$. The first signal measures the alignment of

[1] A simple estimate of the encoded angle is given by $\hat{\theta} = \arg\left((\sum_m \exp(i\theta_m)\, u_\theta(\theta_m, t))/(\sum_m u_\theta(\theta_m, t))\right)$, where the θ_m are the preferred directions of the neurons.

[2] Here the estimator is $\hat{v} = \arg\left((\sum_m v_m u_v(v_m, t))/(\sum_m u_v(v_m, t))\right)$, where the v_m are the preferred speeds of the neurons.

the body axis of the moving agent with its direction of its motion. It is just given by the scalar product of the activity distributions over the body axis of the agent $u_\phi(\phi)$ and the motion direction of the agent $u_\theta(\theta)$ in the form $A_1 = \sum_n u_\phi(\theta_n) u_\theta(\theta_n)$. The second signal A_2 linearly combines information about the speed, and the magnitude changes and angular changes of speed, which are given by a and the angular component of $\mathbf{a}$. The linear mixing weights of the animacy neurons were estimated by fitting the psychophysical results from [2]. Final animacy responses were computed as time averages over the whole trajectories.

The second circuit at the top level of the model classifies the different interaction types based on the following features: speeds $\mathbf{v}_i$ and acceleration a_i of the agents, and relative position $\mathbf{p}_R$, velocity $\mathbf{v}_R$, and acceleration a_R of the agents. These features served as inputs of different classifier models, We tested a multi-layer perceptron, linear and nonlinear discriminant analysis (see also [31]), k-nearest neighbor classification, and a linear and a nonlinear support vector machine.

4 Results

Results on animacy detection are shown in Fig. 3. The model reproduces at least qualitatively the dependence of animacy ratings on directions and speed changes [2]. In these experiments an agent shape moved along a straight line and then suddenly changed speed or direction by different amounts. In addition, the model reproduces the fact that a moving figure that has a body axis, like a rectangle, results in stronger perceived animacy than a circle if the movement, and that the rating is highest if the body axis is aligned with the motion than if it is not aligned [2].

Figure 4 shows example results from the application of the different classifier models for the 6 interaction behaviors in the study [9]. The classifiers were trained on movies generated with the stimulus generation algorithm described in Sect. 2. The linear SVM classifier achieves 99% correct classifications on this data set. See Table 2 for the results with the other classifiers. Most importantly, the model achieved also 100 % correct classifications on the example videos from [9], even though these movies were not used for training.

Table 2. Classification results with different classifiers (6 interaction types).

Classifier	Accuracy
Linear SVM	99.0%
Gaussian kernel SVM	96.3%
LDA	94.7%
KNN	94.7%
Nonlinear LDA	94.3%
Neural Network	94.0%

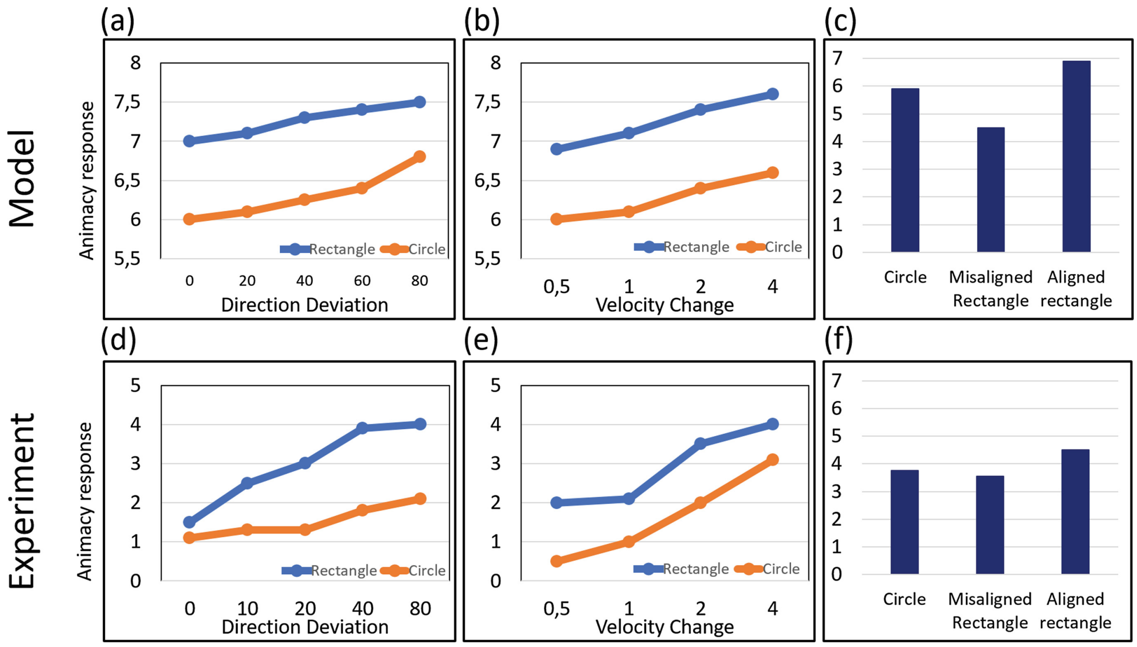
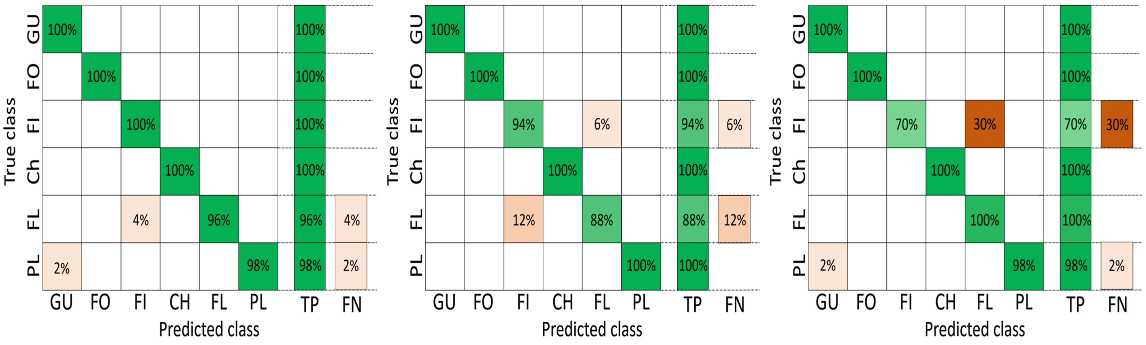


Fig. 3. Simulation results for animacy perception in comparison with experimental results. (a), (d): Dependence of animacy ratings on size of direction change. (b), (e): Dependence of animacy rating on size of speed change. (c), (f): Effect of alignment of body axis with motion direction, compared with moving circle (no body axis).



True class	GU	FO	FI	CH	FL	PL	TP	FN
GU	100%						100%	
FO		100%					100%	
FI			100%				100%	
Ch				100%			100%	
FL			4%		96%		96%	4%
PL	2%					98%	98%	2%

Predicted class

(a) Linear SVM (one-vs-one)

True class	GU	FO	FI	CH	FL	PL	TP	FN
GU	100%						100%	
FO		100%					100%	
FI			94%		6%		94%	6%
Ch				100%			100%	
FL			12%		88%		88%	12%
PL						100%	100%	

Predicted class

(b) Linear SVM (one-vs-all)

True class	GU	FO	FI	CH	FL	PL	TP	FN
GU	100%						100%	
FO		100%					100%	
FI			70%		30%		70%	30%
Ch				100%			100%	
FL					100%		100%	
PL	2%					98%	98%	2%

Predicted class

(c) KNN

Fig. 4. Confusion matrices for the best (Linear SVM) and the worst (KNN) classifier; TP: true positive rate, FN stands for false negative rate. 50 videos per class.

5 Conclusion

Our model accounts by combination of very elementary neural mechanisms for a number of classical results from animacy and social interaction perception from abstract figures. To our knowledge this is the first neural model that can account for such results. Evidently the model is only a proof-of-concept with many shortcomings, a major one being that the accuracy of the form and motion pathway that provide input to the animacy and interaction detection have to be improved. Since the model is in principle consistent with deep architectures for form and action recognition that can achieve high performance level it seems likely that it can be extended to the processing of much more challenging stimulus material. Even in its simple form the model proves that animacy and social interaction judgements partly might be derived by very elementary operations in hierarchical neural vision systems, without a need of sophisticated or accurate probabilistic inference.

Acknowledgments. This work was supported by: HFSP RGP0036/2016; the European Commission HBP FP7-ICT2013-FET-F/604102 and COGIMON H2020-644727, the DFG KA 1258/15-1, and BMBF CRNC FK: 01CQ1704.

References

1. Heider, F., Simmel, M.: An experimental study of apparent behavior. Am. J. Psychol. **57**(2), 243–259 (1944)
2. Tremoulet, P.D., Feldman, J.: Perception of animacy from the motion of a single object. Perception **29**, 943–951 (2000)
3. Tremoulet, P.D., Feldman, J.: The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. Percept. Psychophys. **68**(6), 1047–1058 (2006)
4. Hernik, M., Fearon, P., Csibra, G.: Action anticipation in human infants reveals assumptions about anteroposterior body structure and action. In: Proceedings, Biological Sciences (2014)
5. Scholl, B.J., Tremoulet, P.D.: Perceptual causality and animacy. Trends Cogn. Sci. **4**(8), 299–309 (2000)
6. Gao, T., Scholl, B.J.: Perceiving animacy and intentionality. In: Rutherford, M.D., Kuhlmeier, V.A., (eds.) Social Perception. The MIT Press (2013)
7. Blythe, P., Miller, G.F., Todd, P.M.: How motion reveals intention: categorizing social interactions. In: Gigerenzer, G., Todd, P. (eds.) Simple heuristics that make us smart, pp. 257–285. Oxford University Press, London (1999)
8. Barrett, H.C., Todd, P.M., Miller, G.F., Blythe, P.W.: Accurate judgments of intention from motion cues alone: a cross-cultural study. Evol. Hum. Behav. **26**(4), 313–331 (2005)
9. McAleer, P., Pollick, F.E.: Understanding intention from minimal displays of human activity. Behav. Res. Methods **40**, 830–839 (2008)
10. Schultz, J., Friston, K.J., O'Doherty, J., Wolpert, D.M., Frith, C.D.: Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. Neuron **45**(4), 625–635 (2005)
11. Morito, Y., Tanabe, H.C., Kochiyama, T., Sadato, N.: Neural representation of animacy in the early visual areas: a functional MRI study. Brain Res. Bull. **79**(5), 271–280 (2009)
12. Shultz, S., McCarthy, G.: Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. Neuropsychologia **60**, 115–120 (2014)
13. Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., Decety, J.: The detection of contingency and animacy from simple animations in the human brain. Cereb. Cortex **13**(8), 837–844 (2003)
14. Yang, D.Y.-J., Rosenblau, G., Keifer, C., Pelphrey, K.A.: An integrative neural model of social perception, action observation, and theory of mind. Neurosci. Biobehav. Rev. **51**, 263–275 (2015)

15. Lahnakoski, J.M., et al.: Naturalistic FMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. Front. Hum. Neurosci. **6**, 233 (2012)
16. Isik, L., Koldewyn, K., Beeler, D., Kanwisher, N.: Perceiving social interactions in the posterior superior temporal sulcus. PNAS **114**, E9145–E9152 (2017)
17. Sliwa, J., Freiwald, W.A.: A dedicated network for social interaction processing in the primate brain. Science **356**(6339), 745–749 (2017)
18. Walbrin, J., Downing, P., Koldewyn, K.: Neural responses to visually observed social interactions. Neuropsychologia **112**, 31–39 (2018)
19. Baker, C.L., Saxe, R., Tenenbaum, J.B.: Action understanding as inverse planning. Cogn. Reinf. Learn. High. Cogn. **113**, 329–349 (2009)
20. Shu, T., Peng, Y., Fan, L., Lu, H., Zhu, S.-C.: Perception of human interaction based on motion trajectories: from aerial videos to decontextualized animations. Top. Cogn. Sci. **10**(1), 225–241 (2018)
21. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nat. Neurosci. **2**, 1019–1025 (1999)
22. Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. **4**, 179–192 (2003)
23. Jhuang, H., Serre, T., Wolf, L., Poggio, T.: A biologically inspired system for action recognition. In: IEEE 11th International Conference on Computer Vision (2007)
24. Fleischer, F., Caggiano, V., Thier, P., Giese, M.A.: Physiologically inspired model for the visual recognition of transitive hand actions. J. Neurosci. **15**(33), 6563–80 (2013)
25. Fleischer, F., Christensen, A., Caggiano, V., Thier, P., Giese, M.A.: Neural theory for the perception of causal actions. Psychol. Res. **76**(4), 476–493 (2012)
26. Caggiano, V., Fleischer, F., Pomper, J.K., Giese, M.A., Thier, P.: Mirror neurons in Monkey premotor area F5 show tuning for critical features of visual causality perception. Current Biology **26**(22), 3077–3082 (2016)
27. Warren, W.H.: The dynamics of perception and action. Psychol. Rev. **113**(2), 358–389 (2006)
28. Schner, G., Dose, M.: A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion. Robot. Auton. Syst. **10**(4), 253–267 (1992)
29. Fajen, B.R., Warren, W.H.: Behavioral dynamics of steering, obstacle avoidance, and route selection. J. Exp. Psycholology Hum. Percept. Perform. **1**(3), 184–184 (2003)
30. di Carlo, J.J., Zoccolan, D., Rust, N.C.: How does the brain solve visual object recognition? Neuron **73**(3), 415–434 (2012)
31. You, D., Hamsici, O.C., Martinez, A.M.: Kernel optimization in discriminant analysis. IEEE Trans. Pattern Anal. Mach. Intell. **33**(3), 631–638 (2011)

A Dynamical Generative Model of Social Interactions in Virtual Agents

(To Be Submitted)

A Dynamical Generative Model of Social Interactions in Virtual Agents

Mohammad Hovaidi-Ardestani
Section for Computational Sensomotorics, Department of Cognitive Neurology, Centre for Integrative Neuroscience, Hertie Institute for Clinical Brain Research, University Clinic Tübingen
Tübingen, Germany
mohammad.hovaidi-ardestani@uni-tuebingen.de

Alessandro Salatiello
Section for Computational Sensomotorics, Department of Cognitive Neurology, Centre for Integrative Neuroscience, Hertie Institute for Clinical Brain Research, University Clinic Tübingen
Tübingen, Germany
International Max Planck Research School for Intelligent Systems
Tübingen, Germany
alessandro.salatiello@uni-tuebingen.de

Martin A. Giese
Section for Computational Sensomotorics, Department of Cognitive Neurology, Centre for Integrative Neuroscience, Hertie Institute for Clinical Brain Research, University Clinic Tübingen
Tübingen, Germany
martin.giese@uni-tuebingen.de

Abstract

Humans reliably attribute social interpretations and agency to highly impoverished stimuli, such as interacting geometrical shapes. The computational mechanisms underlying this visual function are unknown and of high interest for numerous technical applications, such as driver-assistance systems or visual scene analysis. Only few psychological stimulus sets are available for testing of this visual function, way not enough for the training of machine vision systems. The automatic generation of such stimulus sets is thus an important technical problem. We introduce here a novel framework for the modelling different classes of social interaction between virtual agents. The algorithm for the simulation of these interactions has been derived from dynamic models of human navigation. We validate our model in three psychophysical experiments where participants had to categorize the animations generated with our model. We were able to isolate 12 interaction classes that were classified consistently by human participants. The remaining confusions between these categories were largely explained by the semantic similarity of the labels used for characterizing the different classes. The proposed methods provides a basis for the development of machine vision algorithms and neural models that classify social interactions from video sequences.

CCS Concepts: • **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks →** Network reliability.

Keywords: social interactions, generative model, perception

ACM Reference Format:
Mohammad Hovaidi-Ardestani, Alessandro Salatiello, and Martin A. Giese. 2020. A Dynamical Generative Model of Social Interactions in Virtual Agents. In *Glasgow '20: ACM International Conference on Intelligent Virtual Agents, October 19–23, 2020, Glasgow, UK.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 Introduction

In their seminal study [5] Heider and Simmel demonstrated that humans can reliably decode animacy and social interactions from strongly impoverished stimuli. Specifically, the participants attributed anthropomorphic qualities to simple geometrical figures that moved around in the two-dimensional plane. Moreover, the figures, were interpreted as animate agents endowed with intentions, emotions, and even personality traits. This study raises the question abut the computational mechanisms underlying this visual function, and its findings were replicated in other studies using similar videos for both human adults ([8, 9, 12]) and preschoolers as young as 5 years old ([18]). Moreover, recent studies have characterized in more detail which critical features of simple stimuli affect the perception of animacy, that is whether the object is perceived as alive [21, 22]. Since Heider and Simmel ([5]), several researchers have tried to reveal which visual cues of motion promote the perception of animacy. Violations of the conservation energy principle, such as heading and acceleration [13], and Newtonian laws of motion [6], together with speed and trajectory direction changes [13, 19, 20] are among the minimal kinematic cues for the emergence of animacy in moving shapes. Later

Glasgow '20, October 19–23, 2020, Glasgow, UK

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

studies [3, 13, 22] used more controlled stimuli and systematically examined what factors can impact the perception of goal-directed actions in a decontextualized animation. These findings provided converging evidence that the perception of human-like interactions relies on some critical low-level motion cues, such as speed and motion direction. In order to further investigate the fundamental basis of the neural encoding of social intent and semantics, the creation of appropriate stimulus sets for humans and monkeys is unavoidable. Specifically the development of neural models for this function, and also the development of computer vision algorithms requires larger stimulus sets as training data. The automatic generation of such stimuli to investigate the mechanisms underlying perception of animacy and social interactions is a challenging task. Handmade animations like those of Heider and Simmel have rich motion features but are not amenable to parametric control. Furthermore, more advanced experiments generally rely on the usage of several figures interacting in different ways. For this reason, the handmade production of such animations is generally unfeasible. In this work, we present a dynamical model that can generate different classes of social interactions controlling the dynamics of the most important factors of social interaction perception, namely speed and motion direction. We validate our model with three separate experiments, where we demonstrate that participants are able to consistently attribute the intended interaction class to animations generated with our model. Our experiments thus show that that artificial displays are rich enough to capture natural looking motions and interactions, unlike what has been recently claimed [7]. Our displays have the additional advantage that the motion cues are quantizable and can be controlled precisely by modulating the model's parameters; more importantly our model allows the automatic generation of arbitrary numbers of videos per social interaction type.

2 Methods

2.1 The Original Approach

In 1976 Reichardt and Poggio [10], provided a quantitative analysis of navigation model that describes mathematically how a fly steers toward moving targets which they chase as part of their mating behavior. Following this seminal work that described the orientation behavior of an autonomous agent using a dynamical system with only an attractor at the direction in which targets lie, several other studies showed that a detailed navigation behavior cannot be described based only on target acquisition. To address this problem, Schöner and Dose ([14, 15]) provided a dynamical system framework that integrates the target acquisition and obstacle avoidance for navigation and exploration of an autonomous agent. In order to model the interaction of two moving agents we exploited this dynamical systems approach, which before was used very successfully for the modelling of human navigation. The original approach focuses on mathematical formalization of reactive control for autonomous robots using differential equations that specify attractors and repellors for behavioral variables that control the agent's heading direction and speed [1]. This framework of integration of target acquisition and obstacle avoidance has been used to implement navigation successfully in an unknown environment both for vehicles and robotic arms [11].

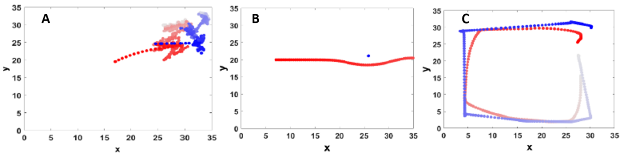


Figure 1. Trajectories of three example social interactions. (A) Fighting; (B) Avoiding; (C) Chasing. Colors indicate agent identity; agent 1: blue; agent 2: red. Color saturation indicates time: darker colors indicate recent time samples.

2.2 The Generative Model

Here, we derived the original idea and defined a dynamical systems or differential equations for the heading directions $\phi_i(t)$ and the instantaneous propagation speed $v_i(t)$ of the interacting agent i. The specified movement is dependent on goal and obstacle points in the two dimensional plane, where the other agent can also act as goal or obstacle as well. We modified a model for human steering behaviour during walking [4] to reproduce the movements during social interactions. The resulting dynamics is governed by the following differential equations for the heading direction

$$\ddot{\phi}_i(t) = -b\dot{\phi}_i(t) + R(\phi_i(t)) + S(\phi_i(t)) \tag{1}$$

where

$$\begin{aligned} R(\phi_i(t)) &= -k^g(\phi_i(t) - \psi_i^g(t))(e^{-c_1 d_i^g(t)} + c_2) \\ S(\phi_i(t)) &= k^o \sum_{n=1}^{N_{obst}} s^{o_n}(\phi_i(t)) \end{aligned} \tag{2}$$

and

$$s^{o_n}(\phi_i(t)) = (\phi_i(t) - \psi_i^{o_n}(t))(e^{-c_3|\phi_i(t)-\psi_i^{o_n}(t)|})(e^{-c_4 d_i^{o_n}(t)}) \tag{3}$$

The variables $\psi_i^g(t)$ and $d_i^g(t)$ represent the instantaneous goal direction of agent i, and the euclidean distance between the agent and its goal. Likewise, $\psi_i^{o_n}$ and $d_i^{o_n}$ represent the instantaneous direction of obstacle n, and its euclidean distance from the agent. Moreover, N_{obst} is the number of relevant obstacles, and k_j and c_j are constants. The forward speed of the agents is specified by the following stochastic differential equation:

$$\tau \dot{v}_i(t) = -v_i(t) + F_i(d_i^g(t)) + k_\epsilon \epsilon_i(t) \tag{4}$$

where $\epsilon_i(t)$ is Gaussian white noise. The two functions F_i that specify the distance dependence of the speed dynamics are different for the two agents. Specifically,

$$F_1(d) = \frac{1}{1 + e^{-c_5(d-c_6)}} - c_7 e^{-kd} \tag{5}$$

$$F_2(d) = \frac{c_8}{1 + e^{-c_9(d-c_{10})}} - c_{11} e^{-kd} + c_{12} \tag{6}$$

To generate the trajectories, we first randomly sample a series of goal points for agent one from a two-dimensional uniform distribution over the 2D plane of action. We then use the instantaneous position of the agent one as goal position for agent two. Samples that are too close to the current agent's position are rejected. Representative trajectories for three example social interactions are illustrated in Figure 1. Note that the the speed control dynamics is not influenced by the presence of obstacles, since their effect does not play a role in our psychological experiments.

2.3 Model Validation

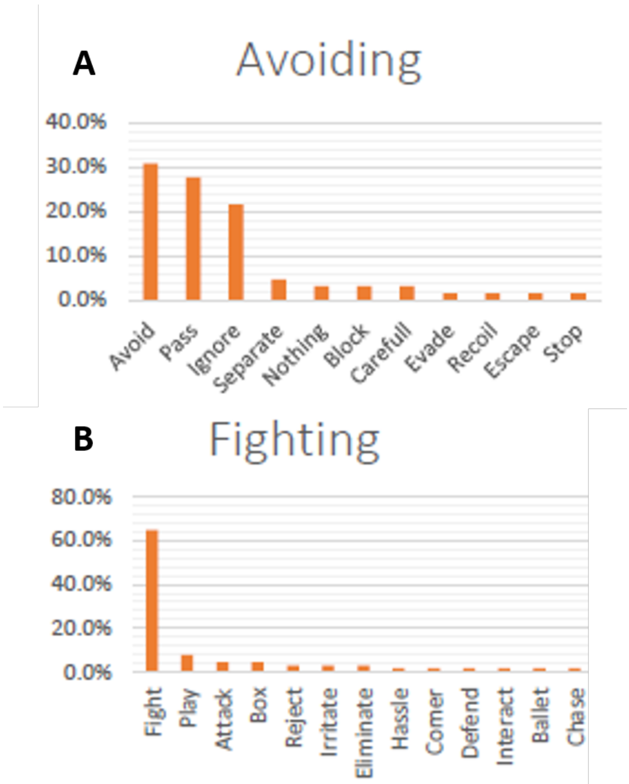


Figure 2. Histograms of reported labels for three example social interactions. True classes: (A) Avoiding, (B) Fighting.

2.3.1 Subjects.

In order to validate our generative model, we carried out three behavioral experiments. A total of 39 subjects with normal or corrected vision took part in the experiments: 13 in experiment 1 (9 females, 4 males), ten in experiment 2 (5 females, 5 males), and 16 in experiment 3 (9 females, 7 males). All participants provided written informed consent before the experiments. All experiments were in full compliance with the Declaration of Helsinki. Participants were naïve regarding the purpose of the study and were compensated for their participation.

2.3.2 Setup.

During the first and the second experiment participants were seated in a dimly lit room in front of an LCD monitor (resolution: 1920x1080, refresh rate: 60 Hz). The computer screen was placed $0.6m$ in front of the participants while their heads were supported by a headstand. Each experiment started with a short training period during which the subjects became familiar with the experimental setup. After the familiarization session, the participants started the main experiment, which consisted of five blocks of watching animated videos and labelling them accordingly separated by five minutes of rest. The animated videos showed two agents (a blue circle and a red rectangle) moving in a 2D plane. The trajectories followed by the agents were generated offline with our model and presented in pseudo-randomized order. Specifically, we heuristically determined 12 parameter sets, where each modeled a distinct class of social interaction. we refer to Figure 6A for a complete list.

2.3.3 Experiment 1.

The first experiment was aimed at assessing whether subjects would perceive the motion of virtual agents generated with our model as a social interaction. A second goal of this experiment was to identify unequivocal labels for the interaction classes generated with our model. To this aim, we asked participants to watch a total of 14 different videos of interaction in 5 variations, each of them shown at most three times. Participants could choose to skip the video after the first or second trial in case they were confident about their interpretation. The total experiment was preceded by one training block that aimed to familiarize participants with the task and the stimuli as well as to direct their focus to the perceptual properties as opposed to the semantic properties of the stimuli. After watching the videos, subjects were asked to provide their interpretations either in a few sentences or ideally by one label. In addition, subjects were asked to report about the percentage of animacy that they perceived for each agent, together with the percentage of naturalness of the interactions they saw in the video with respect to the social interactions. The labels that were most commonly reported by the participants to describe each video in this experiment were used as *ground-truth* interaction labels for the remaining experiments.

2.3.4 Experiment 2.

. The second experiment was aimed at further studying the social interaction classes perceived by the participants while watching our animated videos. To this aim, new subjects were exposed to videos generated with our model. For this experiment we chose 12 of the most distinctive classes of interactions from the previous experiment and showed them to participants in 5 variations 3 times per block. Critically, unlike in experiment 1, after watching the videos, participants were asked to describe the videos by choosing up to three labels, among those selected in experiment 1. Moreover, subjects were also asked to indicate the perceived animacy of each agent.

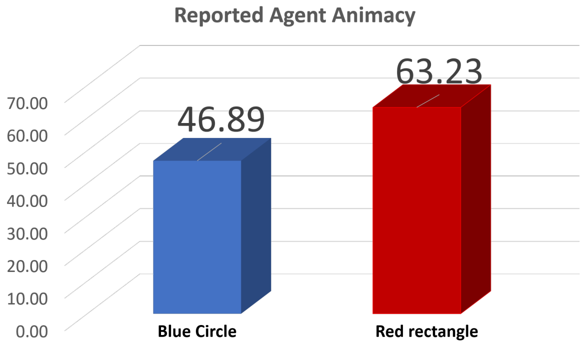


Figure 3. Reported Agent Animacy. The results are average across interactions classes ans subjects

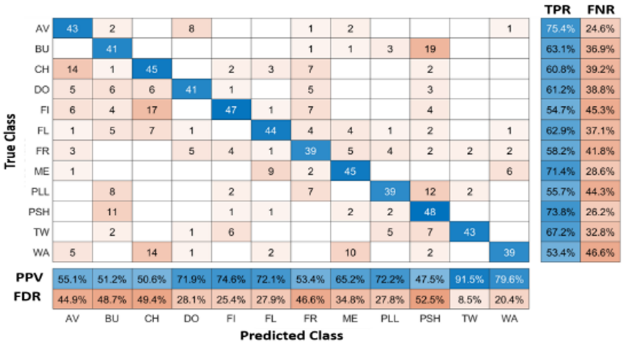
2.3.5 Experiment 3.

Our last research question was whether there are interpretable semantic distinctions among labels. Some misclassified social interaction classes in the previous experiment, suggests that either the generated animated videos are not distinctive enough or these classes semantically overlap with each other. To validate that this confusion is due to inherent similarity of these classes, we ran a semantic survey test by new set of participants. Participants in this experiment did not watch any video but were explained the definitions of each social interaction class label clearly. Having provided the definitions to the subjects we asked them to indicate the level of semantic similarity for each pair of labels by giving rates ranging from 0 to 10 for dissimilar labels to the definitive similar ones. This means that subjects gave rating of 10 for the similarity of each label with itself. In order to quantify the extent to which subjects perceive label meaning to be like another, the data from participants were subjected to multi-dimensional scaling (MDS) ([16, 17]) that provides a spatial representation of underlying relational structures contained in similarity data.

3 Results

3.1 Experiment 1

As mentioned above, participants in this experiment were completely free to either give their own labels or explain their interpretations from the video that they observed. For each class of video all the definitions and labels were pooled together, and the most frequent ones were nominated for the class label. Figure 2 reports two example histograms of reported labels for the classes *Avoiding* and *Fighting*. Avoiding is described by 3 different semantically related labels while for the category fighting mostly one label is used. Although, this was not the case all the time and some classes were named interchangeably depending on from which perspective subjects reported their interpretation about the videos. Also multiple labeling sometimes resulting by characterizing the action of the one or the other agent. For example, pushing and pulling were the classes that their labels were used in both. Besides, some labels (for instance bumping and pushing) were often also misclassified regardless of the perspective from which subjects might have observed the videos. Reported animacy ratings also show that both agents have been perceived animate due to the fulfilment of behavioural cues. Self-propulsion [2], goal directedness [23], and being reactive to social contingencies [3] as the most important behavioural cues and direction together with acceleration [21], and speed [19] are the most discussed motion cues behind the animacy perception. In addition, regardless of social interaction type, the red rectangle has been always perceived more animate which is also compatible with the findings about animacy perception [21].



True Class	AV	BU	CH	DO	FI	FL	FR	ME	PLL	PSH	TW	WA	TPR	FNR
AV	43	2		8			1	2				1	75.4%	24.6%
BU		41					1	1	3	19			63.1%	36.9%
CH	14	1	45		2	3	7			2			60.8%	39.2%
DO	5	6	6	41	1		5			3			61.2%	38.8%
FI	6	4	17		47	1	7			4			54.7%	45.3%
FL	1	5	7	1		44	4	4	1	2		1	62.9%	37.1%
FR	3			5	4	1	39	5	4	2	2	2	58.2%	41.8%
ME	1					9	2	45				6	71.4%	28.6%
PLL		8			2		7		39	12	2		55.7%	44.3%
PSH		11			1	1		2	2	48			73.8%	26.2%
TW		2		1	6				5	7	43		67.2%	32.8%
WA	5		14	1		2		10		2		39	53.4%	46.6%
PPV	55.1%	51.2%	50.6%	71.9%	74.6%	72.1%	53.4%	65.2%	72.2%	47.5%	91.5%	79.6%		
FDR	44.9%	48.7%	49.4%	28.1%	25.4%	27.9%	46.6%	34.8%	27.8%	52.5%	8.5%	20.4%		

Figure 4. Confusion matrix of the classification task. Rows represent the true interaction class and columns the interaction class reported by the participants. TPR: True Positive Rate (a.k.a. hit rate), FNR: False Negative Rate (a.k.a. miss rate); PPV: Positive Predictive Value (a.k.a. precision); FDR: False Discovery Rate

3.2 Experiment 2

Figure 4 shows the total confusion matrix of the classification task of labelling 60 videos per subject (12 classes with 5 variations). As it can be easily observed by the true positive rate or hit rate (TPR) and false negative rate or miss rate (FNR), even the worst class achieved 53.4% of TPR which is considerably higher than the baseline level of 8.3% hit rate for this multiclass of classification task with 12 different classes.

Moreover, the most distinctive classes (*Avoiding*, *Meeting*, and *Pushing*) scored more than 71% of TPR. Nonetheless, there are obviously some misclassifications namely bumping and pushing, chasing and fighting, or walking and chasing. One reason for this misclassification could be the fact that these labels are semantically and intrinsically similar and even real videos of these types of social interactions could be mislabelled.

3.3 Experiment 3

The output from the MDS process is a similarity map that quantifies the pairwise semantic similarity of labels. Since MDS is inherently spatial, items that were rated as being highly like one another are close to one another in the final output. To the degree that any two items were rated as dissimilar, the distance between them have been grown and the similarity matrix in Fig. 6C shows this in more detail. The positions of the 12 stimuli in a 2D spaces generated by MDS are depicted in a 2-dimensional MDS map (Fig. 6A) that shows which classes of interactions are even semantically close together.

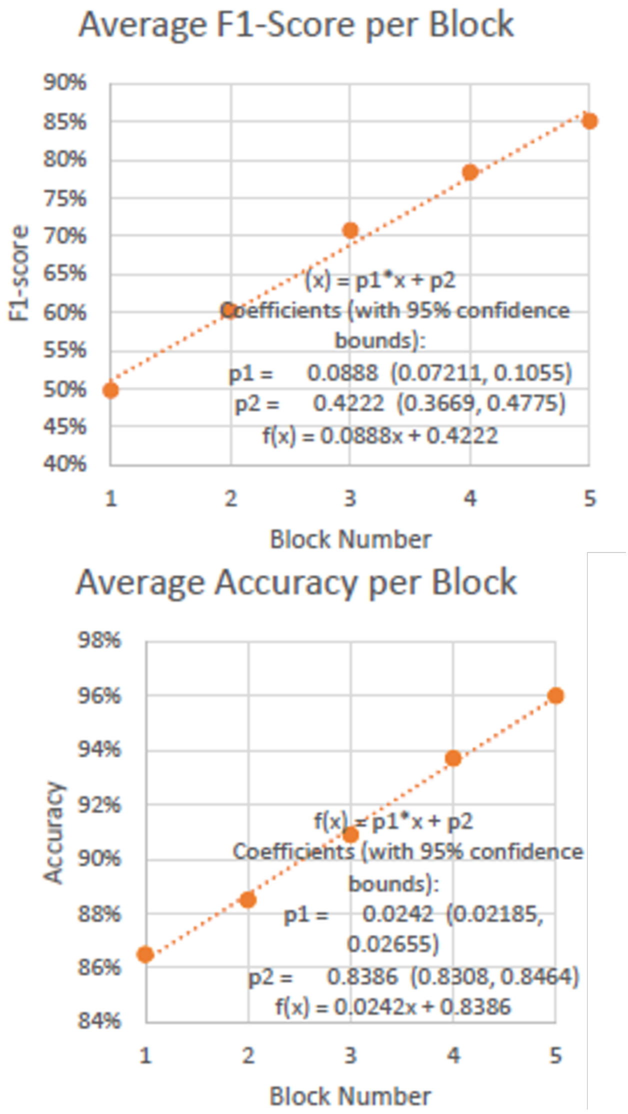


Figure 5. Average F1-score and classification accuracy across blocks. Insets show the fitted linear models together with estimated parameters and confidence intervals

A hierarchical clustering depicted in (Fig. 6B) illustrates how these labels are allocated to different clusters. This shows that misclassified classes are even semantically similar and to some extent we cannot avoid having confusion for these cluster of labels. It can be observed in (Fig. 6) that the semantic similarity does not justify the whole confusion of classification task. However, for some cases i.e., *Pushing* VS *Bumping*, *Walking* VS *Meeting*, *Avoiding* VS *Dodging*, it shows that the confusion was made mostly because of the semantic similarities of these classes. Here we do not claim that all the confusions in the previous task were due to inherent semantic similarities of the classes, but we want to speculate that even this reasonable result of classification task (53.4% TPR of the worst case in 12 classes classification) could have been better if the classes were more sharply distinct. To summarize, our analysis of semantic similarity shows that in one hand some of the confusions are due to semantic similarities and on the other hand demonstrates that some of the similarly semantic classes became less confusing after watching the videos. (e.g., *Tug of War* VS *Pulling*, *Frightening* VS *Avoiding*, *Fighting* VS *Pushing*, etc). (Fig. 5) shows positive linear relationships in accuracy and F1 score of confusion matrix, meaning that after each block of experiment subjects were more confident in choosing their labels for different variations of videos.

4 Conclusion

In this work, we introduced a novel framework for the automatic generation of videos of socially interacting virtual agents. The interactions are defined by two dimensional trajectories, the heading direction, and the speed of each agent generated from the the simulation of a dynamical system. The system is composed of two decoupled differential equations, which define the agents' heading directions and speeds. Moreover, we showed that our model is able to generate as many as 12 different interaction classes, defined by different parameter sets. Finally, we validated our model with three different behavioral experiments, in which participants were able to consistently identify the intended interaction class. Our model is thus suitable for the automatic generation of animations of socially interacting agents, which can be used for instance in experiments in psychology, neuroscience, and similar fields to uncover the neural mechanisms of the perception of social interactions.

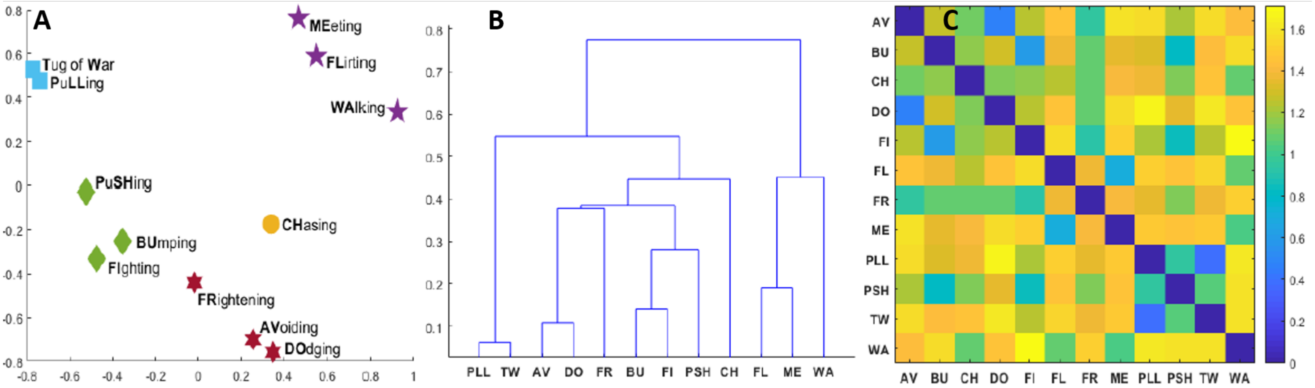


Figure 6. Results of cluster analysis. (A) MDS of semantic labeling (B) Hierarchical clustering of the labels; (C) Matrix of the distances between labels. The label abbreviations stand for: 1. PLL: Pulling, 2. TW: Tug of War, 3.AV: Avoiding, 4.DO: Dodging 5.FR: Frightening, 6. BU: Bumping, 7. FI: Fighting, 8. PSH: Pushing, 9. CH: Chasing, 10. FL: Flirting, 11. ME: Meeting, 12. WA: Walking

Acknowledgments

This work was supported by: HFSP RGP0036/2016; the European Commission HBP FP7-ICT2013-FET-F/ 604102 and COGIMON H2020-644727, and the DFG GZ: GI 305/4-1 and KA 1258/15-1. The authors would like to thank the participants who took part in the study.

References

[1] Estela Bicho, Pierre Mallet, and Gregor Schöner. 2000. Target representation on an autonomous vehicle with low-level sensors. *The International Journal of Robotics Research* 19, 5 (2000), 424–447.

[2] Gergely Csibra. 2008. Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition* 107, 2 (2008), 705–717.

[3] Winand H Dittrich and Stephen EG Lea. 1994. Visual perception of intentional motion. *Perception* 23, 3 (1994), 253–268.

[4] Brett R Fajen and William H Warren. 2003. Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 343.

[5] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *The American journal of psychology* 57, 2 (1944), 243–259.

[6] Katharina Kaduk, Birgit Elsner, and Vincent M Reid. 2013. Discrimination of animate and inanimate motion in 9-month-old infants: an ERP study. *Developmental cognitive neuroscience* 6 (2013), 14–22.

[7] Phil McAleer, Jim W Kay, Frank E Pollick, and MD Rutherford. 2011. Intention perception in high functioning people with autism spectrum disorders using animacy displays derived from human actions. *Journal of autism and developmental disorders* 41, 8 (2011), 1053–1063.

[8] Albert Michotte. 2017. *The perception of causality*. Vol. 21. Routledge.

[9] Keith Oatley and Nicola Yuill. 1985. Perception of personal and interpersonal action in a cartoon film. *British Journal of Social Psychology* 24, 2 (1985), 115–124.

[10] Werner Reichardt and Tomaso Poggio. 1976. Visual control of orientation behaviour in the fly: Part I. A quantitative analysis. *Quarterly reviews of biophysics* 9, 3 (1976), 311–375.

[11] Hendrik Reimann, Ioannis Iossifidis, and Gregor Schöner. 2011. Autonomous movement generation for manipulators with multiple simultaneous constraints using the attractor dynamics approach. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 5470–5477.

[12] Bernard Rimé, Bernadette Boulanger, Philippe Laubin, Marc Richir, and Kathleen Stroobants. 1985. The perception of interpersonal emotions originated by patterns of movement. *Motivation and emotion* 9, 3 (1985), 241–260.

[13] Brian J Scholl and Patrice D Tremoulet. 2000. Perceptual causality and animacy. *Trends in cognitive sciences* 4, 8 (2000), 299–309.

[14] Gregor Schöner and Michael Dose. 1992. A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion. *Robotics and Autonomous systems* 10, 4 (1992), 253–267.

[15] Gregor Schöner, Michael Dose, and Christoph Engels. 1995. Dynamics of behavior: Theory and applications for autonomous robot architectures. *Robotics and autonomous systems* 16, 2-4 (1995), 213–245.

[16] Roger N Shepard. 1962. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* 27, 2 (1962), 125–140.

[17] Roger N Shepard. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* 27, 3 (1962), 219–246.

[18] Ken Springer, Jo A Meier, and Diane S Berry. 1996. Nonverbal bases of social perception: Developmental change in sensitivity to patterns of motion that reveal interpersonal events. *Journal of Nonverbal Behavior* 20, 4 (1996), 199–211.

[19] Paul A Szego and Mel D Rutherford. 2008. Dissociating the perception of speed and the perception of animacy: A functional approach. *Evolution and Human Behavior* 29, 5 (2008), 335–342.

[20] Birgit Träuble, Sabina Pauen, and Diane Poulin-Dubois. 2014. Speed and direction changes induce the perception of animacy in 7-month-old infants. *Frontiers in psychology* 5 (2014), 1141.

[21] Patrice D Tremoulet and Jacob Feldman. 2000. Perception of animacy from the motion of a single object. *Perception* 29, 8 (2000), 943–951.

[22] Patrice D Tremoulet and Jacob Feldman. 2006. The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. *Perception & psychophysics* 68, 6 (2006), 1047–1058.

[23] Benjamin van Buren, Stefan Uddenberg, and Brian J Scholl. 2016. The automaticity of perceiving animacy: Goal-directed motion in simple shapes influences visuomotor behavior even when task-irrelevant. *Psychonomic Bulletin & Review* 23, 3 (2016), 797–802.