Review article

Animacy and the prediction of behaviour

Johannes Schultz [a,b,*], Chris D. Frith [c,d,*]

[a] *Center for Economics and Neuroscience, University of Bonn, Bonn, Germany*
[b] *Institute of Experimental Epileptology and Cognition Research, University of Bonn Medical Center, Bonn, Germany*
[c] *Wellcome Centre for Human Neuroimaging, University College London, London, UK*
[d] *Institute of Philosophy, School of Advanced Study, University of London, London, UK*

ARTICLE INFO

ABSTRACT

To survive, all animals need to predict what other agents are going to do next. We review neural mechanisms involved in the steps required for this ability. The first step is to determine whether an object is an agent, and if so, how sophisticated it is. This involves brain regions carrying representations of animate agents. The movements of the agent can then be anticipated in the short term based solely on physical constraints. In the longer term, taking into account the agent's goals and intentions is useful. Observing goal directed behaviour activates the neural action observation network, and predicting future goal directed behaviour is helped by the observer's own action generating mechanisms. Intentions are critically important in determining actions when interacting with other agents, as several intentions can lie behind an action. Here, interpretation is helped by prior beliefs about the agent and the brain's mentalising system is engaged. Biologically-constrained computational models of action recognition exist, but equivalent models for understanding intentional agents remain to be developed.

1. Introduction

All animals need to predict what other agents are going to do next. This is most obviously the case for interactions between predators and prey. We suggest that there is a hierarchy of mechanisms, of increasing sophistication, for making such predictions[1] (Fig. 1). This hierarchy has three levels determined by the properties of agents (Leslie, 1994):

1) Mechanical properties determined by the agent's bodily structure and internal source of energy.
2) Actional properties related to perceiving the environment and pursuing goals.
3) Cognitive properties concerned with the intentions of the agents.

We also suggest that the prediction process is largely based on the organism's own action-generating mechanisms. For example, association learning is a fundamental mechanism that enables agents to make decisions about what to do next based on past experience. So, if we track the past experience of an agent, we can use the same association learning mechanisms to predict what that agent will do next (see e.g Hampton et al., 2008).

This principle creates an arms race.[2] Prey can develop new strategies for escape, but predators can learn about these strategies and take advantage of them. Whether you are prey or predator, you will have an advantage if you take account of how your rival is choosing actions and making predictions. So, the first step is to identify what kind of an agent you are interacting with. The second step is to try to predict what such an agent is likely to do next.

---

* Corresponding authors at: Center for Economics and Neuroscience, University of Bonn, Bonn, Germany.
  E-mail addresses: johannes.schultz@ukbonn.de (J. Schultz), c.frith@ucl.ac.uk (C.D. Frith).
[1] We present the steps in our hierarchy as building on each other, suggesting a sequential temporal order. Some work supports sequential processing of human actions: EEG response patterns evoked when observing human actions reflect the actions' visual, action, and social-affective features at early, intermediate and late response stages, all within 800 ms of the start of the presented action (Dima et al., 2022). Patterns of visual cortex BOLD responses to human actions reflect a gradual, posterior-to-anterior organization of action similarities based on visual appearance, movements similarity, and actors' inferred goals (Tarhan et al., 2021). The steps we describe might however work in parallel, to a degree that might depend on the kind of stimulus. For example, perceiving a goal-directed human arm movement may simultaneously trigger animacy detection and associations with possible actor intentions associated with the action. This is an empirical question that may be studied by methods such as time-resolved representational similarity analyses (Dima et al., 2022).
[2] We note that this arms race may occur over short time periods through associative learning as well as over long time periods through reciprocal evolutionary changes.

2. Agent detection & movement prediction

Agents are typically living creatures and recognising them on the basis of sensory input depends on a process called animacy perception (Scholl and Tremoulet, 2000; Heider and Simmel, 1944). Many sensory events can lead to the detection of animacy, e.g., hearing a person speak or a bird sing, smelling a wet dog, or feeling an insect land on one's arm. We will mainly focus on vision because this sense is particularly useful for agent detection and animacy perception from vision has been extensively studied (Scholl and Tremoulet, 2000; Scholl and Gao, 2013; Gao et al., 2019). For example, neural representations of animate objects in ventral temporal cortex can be related to reaction times on visual animacy categorization tasks (Carlson et al., 2014; Ritchie et al., 2015; Ritchie and de Beeck, 2019). Vision allows detection at a great distance and over a relatively large part of one's surroundings. The shape and motion of things contain cues used for agent detection. Shape cues could be the outline of a face, of a human or animal body, or luminous eyes of a cat at night. However, while shape cues to animacy may help us infer the sophistication of the agent (e.g., human vs cat), they will not contain much information about its future actions. Motion, in contrast, is by definition a dynamic cue, and is therefore more likely to contain information about an agent's next action. Motion cues can also be perceived in a wide expanse of the visual field, thanks to the motion sensitivity of the periphery of the human retina (Mckee and Nakayama, 1984).

Humans have a bias toward detecting agents, apparent, for example, in the over-detection of faces in the environment (Liu et al., 2014) and in the attribution of animacy to the movements of abstract objects (Heider and Simmel, 1944). Such biases appear intuitively advantageous: false positive detections of potential predators are less costly than false negative ones.

Because of their relevance for the judgment of present and future action, we will now concentrate on discussing the motion cues essential for animacy judgments.

2.1. Self-propelled agents (having an internal source of energy)

The movements of self-propelled agents are determined by hidden internal states, rather than by external forces. In other words, they do not show conservation of momentum. Self-propelled agents can suddenly change direction and speed. For example, they change direction just before they hit a barrier (Scholl and Tremoulet, 2000). Many studies have shown that moving objects are judged to be animate when they appear to be self-propelled (Gyulai, 2004; Michotte, 1946; Opfer, 2002; Stewart, 1982; Szego and Rutherford, 2007; Tremoulet and Feldman, 2000). One study developed a parametric, equation-controlled single-dot stimulus based on this phenomenon. The extent to which the motion is self-propelled determines how the stimulus is perceived. When there is little self-propelled motion the stimulus looks like a leaf drifting in the wind. When the motion is highly self-propelled, the stimulus looks like an insect exploring the environment. A greater degree of self-propelled motion increases the likelihood that the stimulus is perceived as animate (Schultz and Bülthoff, 2013).

2.2. Self-propelled agents with efference copy

The emergence of self-propelled animals, some 6 million years ago, was associated with the development of a nervous system, and not long afterwards, the appearance of efference copy in that part of the nervous system concerned with motor control.

Efference copy is an internal copy of an outgoing (efferent), movement-producing signal generated by an organism's motor system (Jeannerod, 2003). It is created during actively generated motion, and enables animals to take account of the effects of their own movements, such as eye movements, on their perception of the world (Sperry, 1950). Efference copy helps a self-propelled agent to distinguish itself from the relatively stable world through which it moves. Efference copy also helps an agent to navigate through space and maintain heading direction. Travelling along a path is thought to rely on a multimodal representation of space, created from a combination of visual input with motion-related interoceptive information (proprioceptive, vestibular, and motor efference copy) into a multimodal representation of space (Burgess et al., 2002; Tcheang et al., 2011). Agents tend to maintain heading direction when moving, and we may use this as a prior expectation during perception: This is presumably why the right panel of Fig. 2, in which the agent changes their orientation with their trajectory, is perceived as more animate than the left panel (Tremoulet and Feldman, 2000).

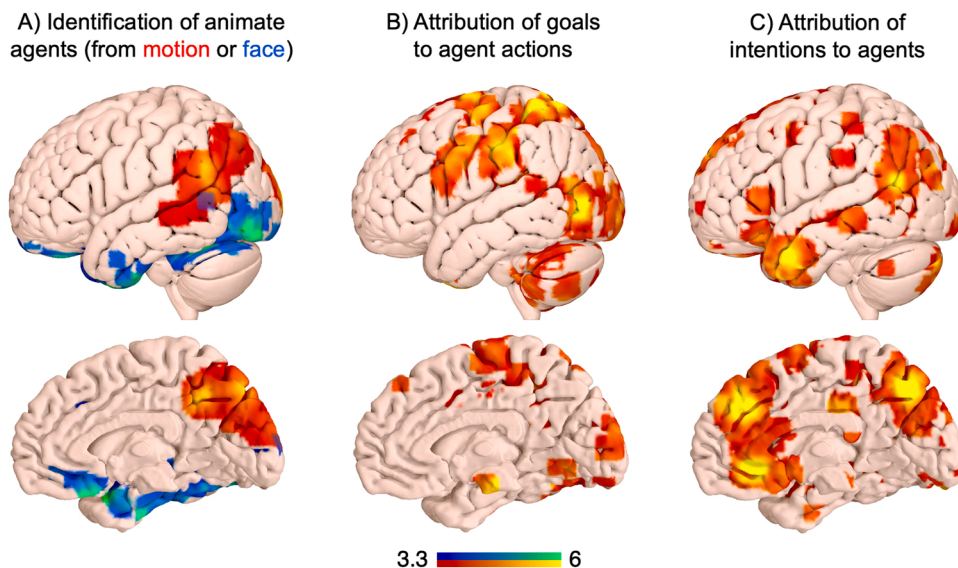Efference copy also allows animals to detect agents and evaluate

Fig. 1. We propose three main stages in the prediction of the behaviour of animate agents and describe some of the human brain regions that might be involved in these stages. To estimate what an animate agent will do in the very near future, the agent first needs to be detected, for example based on visual cues such as biological motion or a face (A). To predict behaviour at a slightly longer time-scale, potential goals behind an agent's actions are computed by the action observation network (B). Predicting behaviour even further in the future involves estimating intentions behind the goals, which engages brain areas associated with mentalizing (C). The regions shown were defined as follows: In A, red-to-yellow colours show T values of a 2nd-level SPM analysis of 38 participants' responses to the contrast biological motion > scrambled biological motion (threshold: P=0.05, family-wise-error-corrected for multiple corrections at the cluster level, cluster-forming threshold, P=0.001; stimuli were point-light walkers); blue-to-green colours show the Z values of an automated meta-analysis association test of 79 studies performed with NeuroSynth.org on 17 June 2022, using the term "face recognition", thresholded at P=0.01, with false-discovery-rate correction. B and C show "action observation" (118 studies) and "mentalizing" (151 studies) NeuroSynth maps with the same settings as used for "face recognition". Data were rendered using Surfice software on the mni152_2009 template brain.
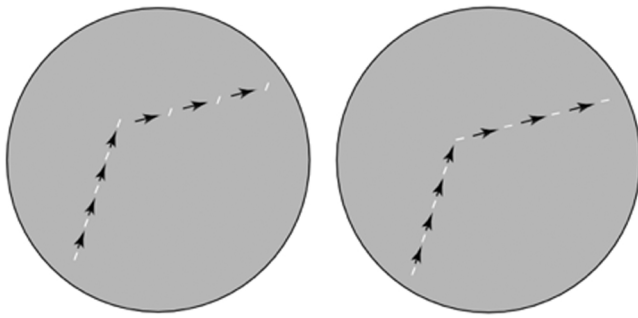
Fig. 2. Two displays of minimal stimuli used to study the attribution of animacy to moving objects. In the left panel, the moving object represented by the white line keeps its orientation unchanged despite changing its trajectory, indicated by the black arrows. In the right panel, the object changes its orientation together with the trajectory; that object is perceived as more animate than the object in the left panel.
Adapted with permission from Tremoulet and Feldman (2000).

their actions: visual motion signals not predicted from efference copy may be caused by the movements of another agent. Such a system appears particularly useful for predators engaged in dynamic pursuit (picture a cheetah chasing an antelope): chasing prey that attempts to escape is only possible if a predator can accurately perceive the prey's movements irrespective of its own movements.

2.3. Biological motion

The motion of animals is constrained by elementary forces such as air or water resistance, gravity, biomechanics, and composition of muscles and skeletons. This leads to the interesting observation that the motion of many land-based vertebrates contains similar motion cues, to which humans appear to be particularly sensitive (Neri et al., 1998). This is typically tested with stimuli displaying so-called "biological motion". Such stimuli can be created by attaching lights to the major joints of a human or other animal. Such dots are immediately perceived as a moving biological agent (Johansson, 1973). The movement of the dots in these displays is a form of embodied movement, since the relative movements of the dots is constrained by the structure of the body, for example the prototypical mammalian body plan (Thurman and Lu, 2013). Indeed, the percept persists when the dots are randomly reassigned anywhere on the limbs of the moving agent, demonstrating independence of local image motion (Beintema and Lappe, 2002). The process by which these moving dots are perceived to be a biological agent is clearly different from that involved in recognising that something is self-propelled. The former process involves extracting form from motion (Blake and Shiffrar, 2007).

2.4. Where will a self-propelled agent go next? Bodily constraints

At the most basic level, the future position of a moving object can be predicted from its trajectory so far. Humans do this when catching a ball (Smeets and Brenner, 1995) and similar mechanisms have been observed in predatory dragonflies, which can focus their attention on a location just in front of a small, moving target (Wiederman et al., 2017).

A self-propelled agent can escape from a predator endowed with such a prediction mechanism by suddenly changing direction. In many fish, for example, this is achieved by the C-start escape response (Fig. 3). When fish detect disturbances in the water generated by striking predators, this escape response enables them to turn very rapidly and swim away in the opposite direction (Eaton et al., 1977). The escape response is set in motion within the first millisecond or two for small fish, and no amount of subsequent stimulation from the other side of the fish can counter the turn. The problem, for the fish, is that this escape route is highly predictable. The tentacle snake, *Erpeton tentaculatus*, takes
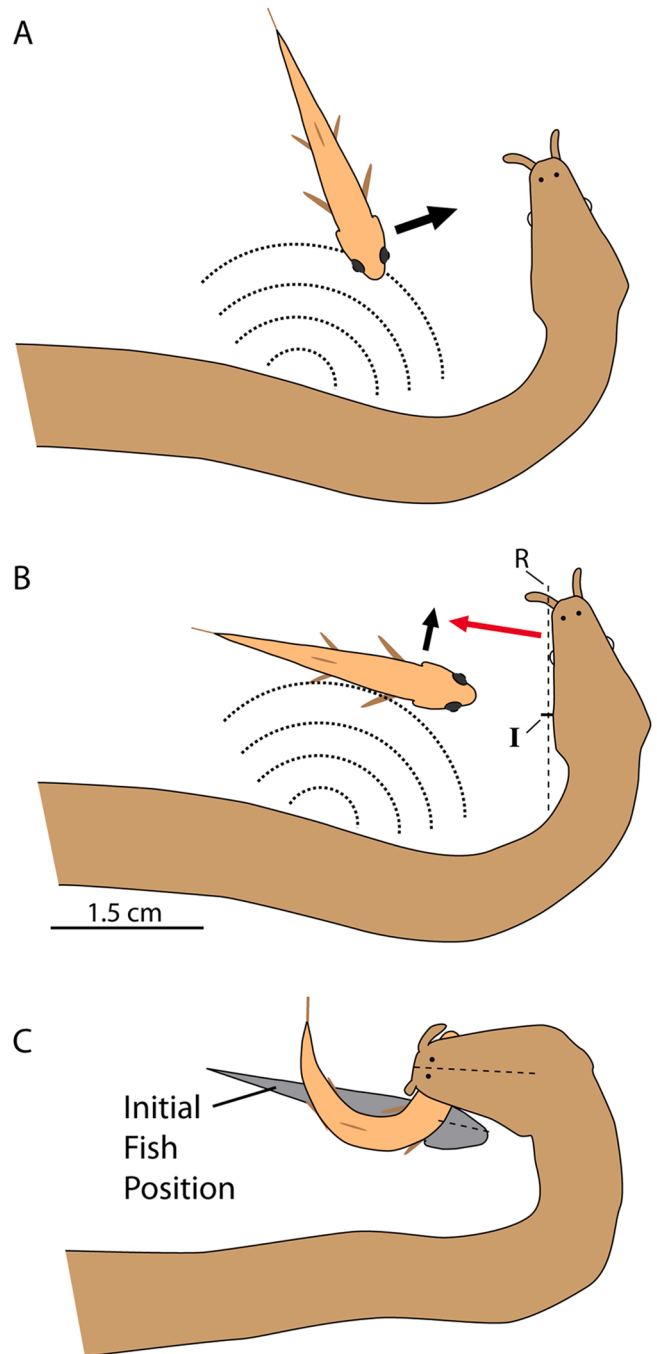

Fig. 3. The tentacle snake, Erpeton tentaculatus, captures fish by provoking a predictable evasion tactic in the fish. (A) The snake induces the C-start response in an unsuspecting fish. (B) The snake anticipates the fish's movement and places its mouth near the future position of the fish. (C) The fish swims directly into the snake's mouth. available under the terms of the CCBY license (https://creativecommons.org/licenses/by/4.0/).
Image reproduced from Catania, 2010 (Catania, 2010), available under the terms of the CCBY license (https://creativecommons.org/licenses/by/4.0/)

advantage of this predictability. The snake curves itself round the fish so that movement of its body on the left side of the fish initiates an escape response to the right. As a result, the fish sometimes swims directly into the snake's mouth (Catania, 2011).[3]

---

[3] As long as such predators are sufficiently rare, as is the case for the tentacled snake, these kinds of escape response remain adaptive.

To overcome this problem, when they detect a possible predator, animals such as cockroaches (Domenici et al., 2008) and praying mantis (Yager et al., 1990) make escape responses in directions that are essentially random. The adaptive value of such behaviour consists of being unpredictable for predators (Brembs, 2011). To avoid triggering evasion tactics, predators attempt to avoid detection by hiding their motion and disguising themselves as stationary. While some animals simply try to move as slowly as possible, others mimic the prey's optic flow of the background in a process known as motion camouflage, so that the attacker does not appear to move when seen by the target (Fig. 4) (Srinivasan and Davey, 1995; Cuthill et al., 2019). For example, dragonflies engage in this strategy by choosing a flight path that allows them to approach the target but also satisfies the constraint of remaining on a line between the target and a stationary landmark point (Mizutani et al., 2003). This way, the only visual change from the target's point of view is the size change of the predator (looming), which is much more difficult to perceive than motion tangential to the target. Interestingly, this strategy is more efficient than moving straight towards the target and is adopted by many animals ranging from hoverflies to falcons. In turn, as looming is the only unavoidable cue emitted by an approaching attacker, it is not surprising that looming is a particularly salient kind of motion that triggers a physiologic response and the perception that an object is approaching (Franconeri and Simons, 2003). Other predators such as leafy sea dragons and stick insects mask their movements by mimicking the motion of the vegetation around them (Bian et al., 2016).

On the other side of the interaction, prey can also attempt to fool the chasing predator by hiding or modifying their motion (Ruxton et al., 2018a), either by playing dead (Skelhorn, 2018; Ruxton et al., 2018b) or, to preserve the possibility of escaping, by mimicking movements of inanimate entities: Mice can roll away from cats towards safety. Such strategies can also be seen in humans during games such as Rugby. The defender can take advantage of the attacker's expectations about the effects of bodily limitations. An example of this is the side-step manoeuvre in Rugby (Fig. 5). The man with the ball (the prey) avoids a tackle by giving the impression that he is going to dodge to his left when he is actually going to dodge to his right.

2.5. Brain systems for interacting with self-propelled agents

In the short-term, the future position of a self-propelled agent can be predicted on the basis of the physical and bodily constraints on its movement. Even the most sophisticated agent cannot overcome these constraints on motion. As we have seen, organisms throughout the
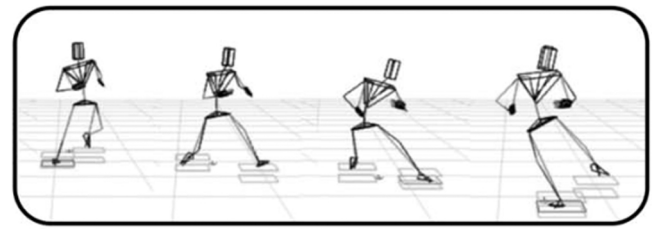

Fig. 4. Attacker trajectories in different kinds of aerial pursuits, including motion camouflage (green). Original image by Chiswick Chap, vectorisation by CheChe, reproduced from Wikimedia Commons (https://en.wikipedia.org/wiki/Motion_camouflage) and available under the CC BY 4.0 license(https://creativecommons.org/licenses/by/4.0/).

Fig. 5. Bodily constraints prevent a rapid change of direction. This is indicated by the trajectory of the centre of mass (the waist), which provides an 'honest' signal indicating that the man will move to the right. However, 'dishonest' signals are provided by the positions of the head, shoulders and left foot. These all suggest that the movement will be to the left. Experts are better than novices at detecting these deceptive movements: Novices attend more to a source of dishonest signals (the shoulders), while experts attend more to the source of honest signals (the waist). In addition, the experts wait longer before making their move (Brault et al., 2012).
Image reproduced from Fig. 5 in Brault et al. (2010), distributed under the terms of the Creative Commons Attribution License.

animal kingdom, from insects and fish to humans, take account of these constrains to anticipate movements, both as predator and prey.

Various parameters, such as position, and velocity, will be available in order to compute the future position of a moving agent. For example, the rate of looming, the expansion in apparent size that occurs as an agent approaches, can be used to estimate the time of contact. Merchant & Georgopoulos (2006) suggest that a very simple nervous system would be able to estimate this contact time. Neurons sensitive to looming are found in several animal species (Merchant and Georgopoulos, 2006) and, in primates, both subcortical and cortical sites respond to looming (Field and Wann, 2005; Maier et al., 2008; Billington et al., 2011).

Interacting with a moving object must recruit the motor system in order to generate a timed motor response, as when catching a ball or avoiding a collision. A recent study reported that Rhesus monkeys, trained to use a joystick to move a dot in order to catch a fleeing square on a computer screen to obtain a reward, accomplished this task by moving toward future positions of the square (Yoo et al., 2020). These future positions were extrapolated on the basis of the square's position, velocity and acceleration. Neurons in dorsal anterior cingulate cortex, a brain region associated with the prediction of action outcomes (Alexander and Brown, 2011; Wittmann et al., 2016), encoded these three variables, and 25% of tested neurons encoded information about the future position of the prey.

But, to predict where an object will move next, it is first necessary to recognise that it is indeed a self-propelled agent. One study showed that when the motion of an object, in this case a single dot, was perceived to be self-propelled (an insect rather than a leaf blowing in the wind), there was a concomitant increase in activity in the right intraparietal cortex of human observers (Schultz and Bülthoff, 2019). This region is part of the 'action observation network', which also includes the posterior part of the superior temporal sulcus (STS) and the ventral premotor cortex (Grossman et al., 2000; Southgate et al., 2008). A part of this network responds both during the observation and the production of a particular action and is called the 'mirror neuron system' (Cook et al., 2014; Kemmerer, 2021; Rizzolatti and Craighero, 2004). Perhaps a result of associative learning between observed and executed actions (Cook et al., 2014; Press, 2011), the system is sensitive to different aspects of biological actions (Kemmerer, 2021) ranging from simple characteristics such as adherence to the "two-thirds power law" (Dayan et al., 2007; Casile et al., 2010) to complex kinematic cues indicative of goals and intentions (Koul et al., 2018; Savaki et al., 2022). Activation of the action observation network and /or the mirror neuron system may thus trigger recognition of the moving object as an agent capable of independent action.

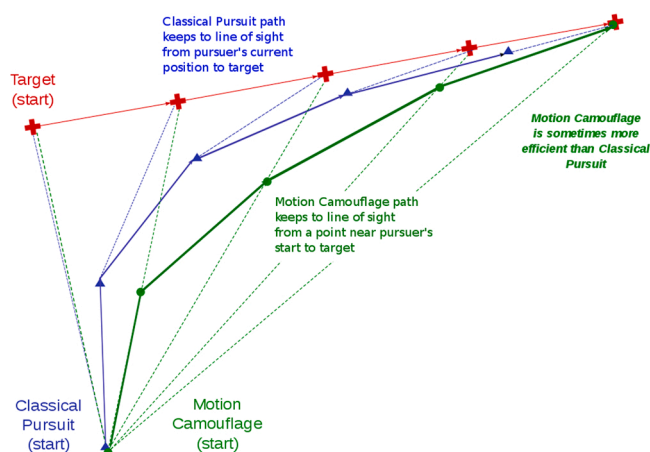Observation of biological motion, in which an agent is reduced to a

system of moving dots, reliably activates the posterior STS (Grossman et al., 2000). This activity is probably related to the processes by which the form, and hence, the identity of the agent (a moving human or some other self-propelled animal) is inferred from the movement (form from motion) (Thurman and Lu, 2013). Such identification provides important clues as to how sophisticated the agent's behaviour is likely to be.

2.6. Summary of section on agent detection and movement prediction

Agents can be identified from their shape and the way they move. In the short-term, the future position of a self-propelled agent is determined by physical and bodily constraints. An arms race occurs in which predators take advantage of these predictions, while prey try to behave unpredictably.

3. Identifying goal-directed agents

In the examples discussed so far, predators successfully predict where self-propelled agents will go next based on the biological constraints on the possible movements of their prey. It is these constraints that determine where the agent will go next in the short-term. Identifying hidden states related to the behaviour may help to predict behaviour in the longer term.

3.1. Goal directed agents minimise the cost of their actions

The behaviour of self-propelled agents can also be determined by goals. So, it is important to know whether we are interacting with a goal-directed agent. Responses of infants to the movements of goal-directed agents have been studied extensively by developmental psychologists. In the scenarios typically used, the agent's goal is to reach a particular spatial location, usually next to some object that the agent 'wants to be next to'. Infants have clear expectations as to how a goal-directed agent should behave. In particular the agent should be rational with regard to action: it should take the most efficient route to the goal, always minimising the cost of its actions (Liu and Spelke, 2017). So, if the agent has to move over a barrier to reach its goal, then its behaviour should change if the barrier is removed, since a new, shorter route is now available to
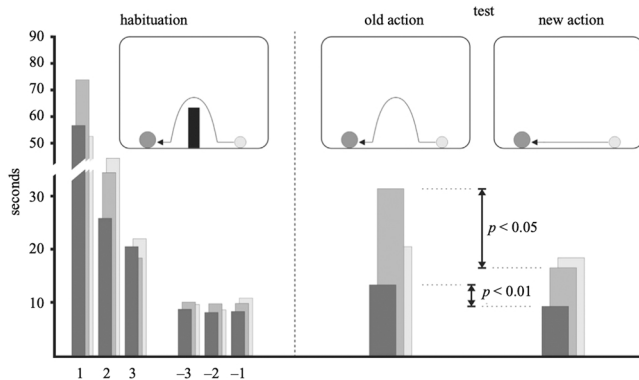

Fig. 6. An experiment revealing infants' expectations about the behaviour of goal-directed agents (Csibra, 2003). First (left), infants are habituated to a display in which an agent jumps over a barrier to reach a target (top inset): looking times decrease from the first to last presentation of the display, for infants of 3 age groups (light grey: 6 months; medium grey: 9 months; dark grey: 12 months). Then, infants are presented with a display of the same action without a barrier, or with a display of a new action in which the agent moves straight to the target. Even though the new action differs from the original action both in terms of barrier presence and trajectory, infants aged 9 and 12 months look longer at the old action without the barrier. This is compatible with the infants being surprised that the agent does not adapt its action to the changed environment without the barrier.
Reproduced with permission from (Csibra, 2003).

reach the goal (Fig. 6). Infants (6.5 m) expect human or animal-like agents to behave in this way and are surprised if they do not (Csibra et al., 1999). These studies suggest that, for infants, movements that are goal-directed are those in which the end point of the movement is more important than the form of the movement. Infants of 6 months attend to the target of an action (object grasped) rather than the form of the action (Woodward, 1998).

3.2. Cues indicating that an agent is goal directed

Infants do not expect non-animal-like objects (e.g., wooden boxes) to change their behaviour when the environment changes, but can easily be persuaded that such objects are goal-directed, if they are seen to behave in certain ways. Several aspects of movement play a role in attributing goals to behaviour. *Free choice of action*: A box is treated by infants of 6.5 m as a goal directed agent if its movement is variable, i.e. it gets to the same location by different routes (Csibra, 2008). *Contingent behaviour*: An oblong, furry thing is treated by infants (12 m) as a goal directed agent if the object has previously been seen to interact contingently with another agent (Johnson et al., 2007). *Self-propulsion* of an object does not seem to be necessary for the expectation of goal-directed action. Even objects that do not move in a self-propelled manner can acquire the attribute of being goal-directed (Csibra et al., 1999; Johnson et al., 2007). However, from an early age, infants attribute goal directedness to any entity they perceive as self-propelled (Luo and Baillargeon, 2005). Children attribute more psychological properties (planning, thinking, choosing) to self-propelled objects that seem to follow a goal compared to objects that do not, and both children and adults attribute more biological properties (being alive, growing) to such goal-directed objects (Opfer, 2002). *Cue combination:* Self-propelled motion evokes an animacy percept more readily when combined with contingent behaviour. For example, a moving particle is rated as more animate when it appears to react to its environment compared to when it does not react, and more animate when the reaction is active rather than passive, i.e. explainable by physical forces (Tremoulet and Feldman, 2006). *Biological motion*: Within biological movements, constraint by the human body plan does not seem necessary for the attribution of goal directed action. Articulated non-human biological motion can appear goal-directed and animate to adults (Pyles et al., 2007), and infants attribute goal directedness even to biomechanically impossible actions (Southgate et al., 2008).

3.3. Interacting agents

In the paradigms discussed so far, the goal-directed agents mostly have their movements coupled to a particular position in space. The movements required to reach this position are flexibly adapted to avoid obstacles while minimising effort.

However, the goal in a goal-directed action need not be a fixed position in space. There are many studies of *chasing*, in which one agent (the 'wolf', represented by a disk) chases one among a number of randomly moving disks (the target 'sheep'). Chasing can be detected in displays as long as the wolf always heads almost directly towards the target sheep. Chasing will not be detected if this heading direction varies by more than $30°$. Chasing is also more detectable if the wolf (now represented by a dart shape) continually faces towards the target sheep (Gao et al., 2009). Chasing is more difficult to detect when the wolf's chase is interrupted by periods of random motion (Gao and Scholl, 2011). Human performance can be replicated by a Bayesian ideal observer model that integrates a bottom–up search for motion features with a top–down selection process for possible motion trajectories (Gao et al., 2019).

3.4. Minimal agency

An agent can be goal-directed without having any internal

representation of goals. This has been referred to as *minimal agency,* the conditions for which are only a flexible and adaptive regulation of the agent's coupling with the environment, without the need for mental states (Barandiaran et al., 2009). However, since an agent's actions can appear goal directed even when the acting agent is not representing the goal, there may be an advantage for the observer to representing the agent's goal. This will allow better prediction of future actions by the agent. From the point of view of a predator trying to catch an agent, estimating where it 'wants to go' in the longer term may be more useful than estimating where it will move next in the short term.

3.5. Brain systems concerned with perceiving goal-directed behaviour

Observation of goal-directed action reliably activates the action observation network (Rizzolatti and Craighero, 2004; Cross et al., 2016). For example, observation of a goal-directed, self-propelled object (a ball) jumping over a barrier elicits activity in the inferior parietal lobe [specifically the anterior part of the intraparietal sulcus (AIPS), but not the posterior STS (pSTS)] and the premotor cortex in comparison with a condition in which the same object makes the same movements, but is launched (Stosic et al., 2014).

Pelphrey and colleagues showed observers very simple examples of goal directed behaviour. In one example, an object started flashing and a human avatar either moved her eyes towards this target or away from it. As has been observed before (Puce et al., 1998), the moving eyes, an example of biological motion, elicited activity in pSTS. However, movement away from the target elicited greater activity in pSTS (and also AIPS) than movements toward the target (Pelphrey et al., 2003). The same paradigm was repeated except that grasping movements rather than eye movements were displayed. Again, greater activity was elicited in pSTS and AIPS for the grasp away from the target (Pelphrey et al., 2004). In these examples, greater activity was elicited when the agent did not move towards the expected goal. In another study, pSTS activity increase evoked by changing or missing goals was only observed when a human character or a humanoid robot performed reaching movements, not when simple mechanical devices performed these actions (Carter et al., 2011).

Presumably observers have a strong prior expectation that goal-directed agents, such as humans, will move towards the obvious target. Unexpected movements require updating of the goal of the agent. This is analogous to the surprise shown by infants when an agent does not behave in the expected goal-directed manner. Increased pSTS activation is observed for other kinds of social prediction errors, for example those arising when gamble outcomes do not match predictions given by other people (Behrens et al., 2008) or when the expected degree of influence of a participant's action on another's action does not reflect their actual influence (Hampton et al., 2008). Information about the goal-directedness of the action need not be conveyed by visual motion: Activity in the temporo-parietal junction (TPJ), pSTS & AIPS can even be elicited by verbal descriptions of simple, goal-directed behaviour (Ma et al., 2012).

Observation of agents interacting also robustly activates the pSTS region, especially if a chasing 'wolf' frequently switches goals, from one 'sheep' to another (Gao et al., 2012). The pSTS may be generally involved in assessing goals of moving agents: asking participants to assess the strategy of an agent chasing a moving target leads to increased pSTS activation compared to simply assessing the outcome of the chase (Schultz et al., 2004). Presumably pSTS is particularly active when the observer is detecting which sheep is the goal, i.e. trying to predict where the wolf will move next. Consistent with this interpretation, pSTS is also activated when observers are trying to find chasing in visual displays of completely random motion, when no chasing is actually present (Lee et al., 2014).

In a variation in the chasing paradigm, Schultz and colleagues (Schultz et al., 2005) used a display in which two objects moved simultaneously. The coupling between the two objects varied from no

interaction to a mutual influence on each other's movements. Activity in pSTS increased in relation to the degree of correlated motion between two objects. This was independent of any explicit instruction to look for an interaction. Observers also assigned greater animacy to the objects when their movements were coupled. Interaction between agents seems to be important to elicit this activation: pSTS is less activated when individual agents are observed pursuing goals or interacting with inanimate objects (Isik et al., 2017). Presumably this is because, when there are two agents, there are also two possible goals that need to be tracked.

The involvement of the mirror system is compatible with predictive coding accounts of action understanding (Urgen and Saygin, 2020). According to these accounts, the motor system is involved in drawing inferences from observed actions (Gallese and Goldman, 1998; Friston et al., 2011; Kilner et al., 2007). Specifically, these accounts propose that, based on context, an observer has a prior expectation about the goal of an action, and can use his own motor system to predict movements (what would I do if I had that goal). Deviations from the expected movement (prediction errors) would then be used to update the estimate of the agent's goal (Urgen and Saygin, 2020; Kilner et al., 2007). We note that the extent to which these predictive coding accounts can explain recognition of goals from different kinds of motor acts is still actively debated (Kemmerer, 2021; Vannuscorps and Caramazza, 2016; Heyes and Catmur, 2022). An alternative approach to explain goal recognition from observed actions follows the structure of classic biologically-inspired models for object recognition (Riesenhuber and Poggio, 1999). These models are based on a hierarchical structure of neurons with gradually larger and more complex receptive fields, consistent with neurophysiological findings from the visual system. Replicating classic neural findings, they include parallel and converging pathways for the analysis of shape and motion cues. Their plausibility as models of real neural mechanisms is supported by their abilities to recognize actions in biological motion displays (Giese and Poggio, 2003), detect goal-directed hand actions (Fleischer et al., 2013), and even detect animacy and simplified social interactions between two moving dots (Hovaidi-Ardestani et al., 2018). The fact that the model combines these abilities fits with a proposed similarity between the computational mechanisms of action and causality perception, namely the analysis of sequences of spatiotemporal relationships between interacting stimulus elements (Fleischer et al., 2012; Ullman et al., 2012). In fact, the same brain regions have been found to respond to both filmed naturalistic, goal-directed hand-object interactions and disks interacting in such a way as to replicate the causal relationships of the hand-object interactions (Caggiano et al., 2016).

3.6. Summary of section on goal-directed behaviour

In the longer-term, an agent's actions will be determined by its goals. Several movement features, isolated or combined, can lead to the inference that the agent is goal directed: self-propelled motion; biological motion; contingency between an agent, its movement direction and the assumed target; or signs that an agent adjusts its actions to reach the goal with the least effort. The brain regions that form the action observation network (including TPJ, pSTS & AIPS) are most likely implicated in recognizing goal-directed actions in several kinds of media, from moving objects, through hand and body actions, and even verbal descriptions. pSTS activity reflects prediction error signals that facilitate updating of assumed goals (according to the predictive coding account of action observation) and also appears particularly involved in processing interacting agents. Computational models with biological or cognitive constraints can reproduce goal detection performance, revealing clues about the neural mechanisms underlying this capacity.

4. Identifying intentional agents

What are the mechanisms underlying our ability to predict the actions and the reactions of other agents when we take account of their

intentions? For what kind of behaviour is the representation of mental states necessary, what triggers the intentional stance (Mar and Macrae, 2008; Dennett, 1987)? Presumably behaviour that is constrained by mental states over and above any effects of the physical environment. But how intentional behaviour is detected is still far less well understood than the mechanisms underlying our ability to parse movement kinematics or infer goals.

4.1. Perceiving movements as caused by mental states

As described earlier, children attribute both intentions and goals to self-propelled entities. Are there any characteristics of movements that make them appear intentional, rather than simply goal directed? Evidence that this is the case comes from animations of the kind originally created by Heider and Simmel (Heider and Simmel, 1944). When observers watch the geometric shapes in these animations moving about and interacting with one another, they tell remarkably consistent stories about what they think is happening. In these narratives, the behaviour of the triangles is described as being determined by mental states. For example, *"The circle is trying to persuade the large triangle. The large triangle is jealous. The small triangle is trying to coax the circle. The large triangle is manipulating the circle's feelings. The two small shapes are celebrating their deception of the large triangle."*

What the observers see in these animations is very artificial. The interactions are seen from above and the movement of the shapes is not biological in the sense of being constrained by bodily structure. There are no facial expressions and there is no sound. So, what is it about these animations that reveals that the behaviour of the triangles is intentional? Two requirements seem to be that contingent movements and mutual interactions are shown between two or more agents and that the time frame is reasonably long. But these factors are clearly not enough. How does movement caused by intentions differ from goal directed movement? Goal-directed behaviour is marked by rationality. The agent reaches his/her goal via the shortest or least effortful path. So, what do we infer when the movement is not constrained by the principle of least effort? Is this a marker of intentionality?

4.2. Movement as communication

If we see agents jumping unnecessarily high over a barrier, we will spontaneously ask, 'why they are behaving like this'? One possibility is that this is more than just action. This could be a deliberate signal, a communication. Vesper and Richardson (Vesper and Richardson, 2014) studied a situation in which two participants had to synchronise their hand movements when tapping on different targets (Fig. 7). However, only one of the two actors, the leader, knew which was the target on each trial. If the follower could see the hand movements of the leader, then the leader made exaggerated movements which indicated which was the current target. Deliberate signalling of this kind would be an indication that the movements were being generated by an intentional agent. The behaviour of the triangles in the Heider and Simmel experiment (Heider and Simmel, 1944) frequently includes movements that

seem irrational in terms of goal directed behaviour. For example, near the beginning of the sequence, the large triangle rotates from side to side in front of the small triangle. This makes no sense in terms of moving to reach a goal. But it makes sense as a signal, communicating 'no' to the small triangle. Once such signalling has been inferred, the observer will assume that the objects are intentional agents and will seek to understand their behaviour as being caused by hidden mental states, and will look for a story that explains the interaction.

4.3. The ambiguity of movements as cues to mental states

Movements and other forms of behaviour are interpreted by combining perceptual signals with prior beliefs about the agent. Many actions contain kinematic cues about intentions (Cavallo et al., 2016), and regions of the action observation network represent this information (Koul et al., 2018; Patri et al., 2020). However, a major problem for inferring intentionality occurs when identical behaviour must be interpreted differently depending on the most probable mental states of the agent: Does the doctor pick up the scalpel to help or to harm (Jekyll vs Hyde)? (Jacob and Jeannerod, 2005). To infer the intention behind the goal in such a case, we need some prior knowledge about the agent. When observing people grasping a cup in order to drink or pour, attribution of these intentions depends on both expectations about the agent and informativeness of their movement (Koul et al., 2019). Social brain regions of participants watching and imagining simple object movements were more active when the movements were presented in a visual context suggesting an animate interpretation rather than an inanimate interpretation (Wheatley et al., 2007).

There are many cases where our beliefs affect our interactions with others. For example, observing an actor's movements can interfere with the movements of an observer. However, these interfering effects are not determined by whether observed movements are biologically plausible or not, but by whether the observer believes they are made by a human or a computer (Stanley et al., 2007). People accused of wrongdoing are blamed less when they are believed to be victims rather than heroes (Gray and Wegner, 2011). Less trustworthy-looking participants receive smaller investments in a trust game (van 't Wout and Sanfey, 2008). Beliefs held at different levels (human or not, goal of current action, current mood, personality, social group…) allow action prediction and corresponding response selection at different time scales (next move, overall outcome of the game, ultimate goal of the interaction…).

Prior beliefs thus play a very important role when we try to understand intentional behaviour. This large influence of beliefs about other agents on the interpretation of their actions may stem from the difficulty in detecting intentional agents simply on the basis of their behaviour.

4.4. The problem of recursion

Another major problem for our interactions with intentional agents is that, while we are trying to infer the mental states of our opponent, that opponent is trying to infer our mental states. This can lead to an infinite recursion: should I try to infer what you think I think you think my
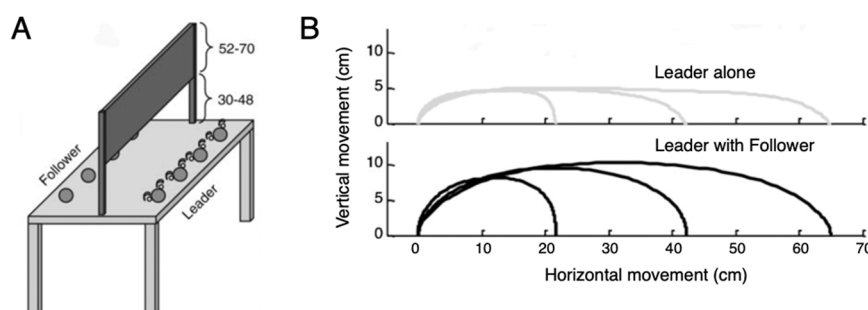
Fig. 7. Communicative intentions change movements. (A) In this experiment, two participants seated on opposite sides of a table are asked to synchronously tap on grey disks. The Leader receives information about which disks to tap on, and the Follower attempts to synchronize with them. (B) The Leader's hand movements are amplified when acting with the Follower (black trajectories, bottom panel) compared to a control condition in which the Leader acts alone (light grey trajectories, top panel).
Modified, with permission, from (Vesper and Richardson, 2014).

mental states is …? Fortunately, the rationality of most people is severely bounded so that typically only two recursions are considered. Computational modelling of dyadic games based on bounded rationality models of game interactions (Camerer et al., 2004) allows us to quantify the sophistication of a game partner's strategies (Hampton et al., 2008; Coricelli and Nagel, 2009). This allows us to simulate actions of game partners adopting such strategies, because it is not enough to detect that we are interacting with an intentional agent, we must also assess how sophisticated the agent is. To win in a competitive interaction, it is advantageous to be more sophisticated than our opponent.

In such interactions we also find a major influence of our prior beliefs about our opponent. Devaine and colleagues (Devaine et al., 2014a) had participants play against a computer agent in two conditions. In one condition, participants believed they were playing against a person, while in the other they believed they were playing against a complex one-armed-bandit. In reality the behaviour of the opponent was identical in the two conditions. When they believed that they were playing against a human, they could hold their own against a sophisticated opponent. But when they believed that they were playing against the machine, they were beaten by that same sophisticated opponent. Apparently, they were unable to detect the degree of sophistication from the behaviour of the opponent. Reciprocal influences between agents engaged in collective decision-making also disappear if people believe that they are working with a computer partner (Mahmoodi et al., 2018).

Confronted with complex behaviour, we face a dilemma. Should we adopt an intentional strategy which is cognitively demanding, or less demanding strategy which may lead us to lose against this opponent? Our default assumption seems to be that most opponents are not sophisticated. This may be, in part, because sophisticated strategies may not always be ideal: in competitive situations, sophisticated strategies can be too costly, while in cooperative situations, less sophisticated strategies work well (Devaine et al., 2014b).

4.5. Brain systems associated with detecting intentional behaviour

In addition to the action observation network, observation and detection of intentional behaviour brings into play the brain's mentalising network (Frith and Frith, 2003). This network includes medial prefrontal cortex (mPFC), temporoparietal junction (TPJ) and precuneus (Schurz et al., 2014, 2021).

mPFC has a role in evaluating the intentions that lie behind the goals of an agent's action. This brain region is active when people think about mental states of others or themselves (Amodio and Frith, 2006), for example when participants consider false beliefs of others (Fletcher et al., 1995), or when they select a cartoon character's most likely next action (Brunet et al., 2000). Neurons in this region represent the contents of other people's beliefs and accurately predict whether they are true or false (Jamali et al., 2021). This region is also active when playing games such as stone-paper-scissors (Gallagher et al., 2002) or economic games (Sanfey et al., 2003) against a human or a similarly complex opponent. mPFC appears to be generally involved in modelling the choices of agents (Nicolle et al., 2012; Apps et al., 2016; Báez-Mendoza et al., 2021) and responds to different kinds of simulated reward prediction errors (i.e. teaching signals) experienced by the other player during strategic games (Lee and Seo, 2016).

Interestingly, several studies have shown that the mPFC is more active when people *believe* that they are playing a game against a person rather than a computer, even when the opponent's action are the same (Gallagher et al., 2002; Takahashi et al., 2015; Kircher et al., 2009). This effect appears already during action observation: mPFC is more active when watching an action performed by a human character rather than a robot, despite the movements being identical (Carter et al., 2011). A number of studies show that the action observation network is activated when observing both human and robot actions, while the mentalising network is only activated when participants believe that the agent is human (see Cross et al., 2016, for a review). For example, mPFC is

activated when observers believe that biological motion is human generated, whether the actual movements are realistic or not (Stanley et al., 2010).

Economic games are particularly interesting when studying more advanced attribution of intentions, as the complexity of players' strategies can be quantified, increasing from the simplest to the more complex: step 0 = choose randomly, step 1 = choose under the assumption that the other plays at step 0, step 2 = choose under the assumption that the other plays at step 1 or below, and so on (Camerer et al., 2004). mPFC activity increases with the complexity of a player's recursive thinking (Coricelli and Nagel, 2009), is higher in social games requiring a higher level of strategic thinking (Nagel et al., 2018), and reflects the degree to which participants follow a successful strategy (Konovalov et al., 2021). In sum, as predicting other agents' actions based on their intentions becomes more difficult, activity in mPFC increases. Whether mPFC activity is driven more by strategy complexity or the "human-likeness" of an opponent appears to depend on the specific study settings (Konovalov et al., 2021; Takahashi et al., 2014).

Similar effects have been observed in TPJ. In a study by Carter and colleagues (Carter et al., 2012), participants played simplified poker against humans or computers, deciding on each trial to bet or fold. The activation pattern in TPJ predicted betting or folding, much better than any other region, but only when playing against humans, and only in participants who judged the human opponent to play better than the computer. In another study, participants played an economic game against a human opponent, an artificial agent that followed a fixed strategy, and an artificial agent that adjusted its choices through a machine-learning algorithm (Ogawa and Kameda, 2019). Left TPJ was correlated with choice behaviour against the human opponent and the sophisticated artificial agent only, while the right TPJ showed this effect only against the human opponent. Disrupting TPJ activity with transcranial magnetic stimulation disrupted mentalizing-based decision-making in a social game (Hill et al., 2017). TPJ interacts with mPFC during strategic interactions (Lee and Seo, 2016; Griessinger and Coricelli, 2015). TPJ could be implicated in identifying optimal strategies, which are then implemented by the mPFC (Konovalov et al., 2021).

The major role of the mentalising system is to represent other people (and perhaps other similarly complex agents) in terms of their mental states (Thornton et al., 2019a). Estimates of current mental states of other people helps to predict their future mental states (Thornton et al., 2019b) and hence their future actions. Dimensions underlying the representation of these mental states and transition probabilities between mental states have been proposed (Tamir and Thornton, 2018), and computational models of mentalizing based on game theory, predictive coding or Bayesian inference are being developed (for a review, see Kliemann and Adolphs, 2018). Such models attempt to dissociate the specific computations involved in mentalizing (Kliemann and Adolphs, 2018; Baker et al., 2017; Konovalov et al., 2018; Charpentier and O'Doherty, 2018; Suzuki and O'Doherty, 2020; Schaafsma et al., 2015). Some of these computations might also be applied in non-social contexts, calling into question these computations' domain-specificity (Konovalov et al., 2021; Lockwood et al., 2020; Ramsey and Ward, 2020; Heyes, 2014). Whether experimental situations trigger these computations and their underlying neural correlates probably depends on the sophistication of the interaction partner and the beliefs that participants have about them. How mental states link to observable behaviour, how external situational factors influence the transitions between metal states, and to what extent computational models can predict observable, real-world behaviour is likely to remain a focus of intense research in the foreseeable future.

4.6. Summary of the section on brain and intentionality

People spontaneously attribute intentions to certain movements. On what basis are these intentions inferred? First, there are motion cues, such as those signalling that actions deviate from those costing the least

effort (indicative of a communicative intention). Then, there are prior expectations about the nature of the agent performing the action: Is it human? How sophisticated is it? The mentalising network of the brain, including mPFC and TPJ, come into play when the agent is believed to be intentional. mPFC, in particular, appears to play a role in estimating the sophistication of the agents we are interacting with.

4.7. Epilogue

We propose a hierarchy of mechanisms implicated in predicting the behaviour of animate agents. Complementing recent reviews discussing behaviour prediction (e.g., Lee and Seo, 2016); Tamir and Thornton, 2018; Kliemann and Adolphs, 2018), we discuss how different mechanisms are triggered depending on cues about the agent's animacy, the kind and the time-scale of the behaviour to be predicted. A number of neural mechanisms are in place for predicting the behaviour of self-propelled and goal directed agents, but dealing with intentional agents is much more difficult. It is hard to detect that an agent is intentional from behavioural cues alone, and we depend to large extent on our prior beliefs about which kinds of agent are intentional. Our deep interest in novels, sitcoms and gossip may reflect our need to attain a greater understanding of intentional behaviour. Likewise, our fascination with stories about atypical agents such as zombies, psychopaths, robots or other kinds of artificial intelligences may reflect our difficulty in defining the category of intentional agents.

But today, such agents are no longer restricted to the imagination. Humanoid robots and self-driving cars have arrived among us. Our prior expectations about the behaviour of non-human agents are going to have to be adjusted, and we need to develop robots and artificial intelligence systems that match our expectations in order to benefit from interacting with them (Henschel et al., 2020). Through development and study of these novel agents we have a new opportunity to understand intentional agents and, thus, ourselves much better.

Funding

Declarations of interest

None.

Data Availability

The data used to create Figure 1 are available at https://osf.io/957cv/.

References

van 't Wout, M., Sanfey, A.G., 2008. Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. Cognition 108, 796–803.

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. 14, 1338–1344.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.

Apps, MAJ, Rushworth, MFS, Chang, SWC, 2016. The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. Neuron 90 (4), 692–707.

Báez-Mendoza, R., Mastrobattista, E.P., Wang, A.J., Williams, Z.M., 2021. Social agent identity cells in the prefrontal cortex of interacting groups of primates. Science 374, eabb4149.

Baker, C.L., Jara-Ettinger, J., Saxe, R., Tenenbaum, J.B., 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat. Hum. Behav. 1, 1–10.

Barandiaran, X., Di Paolo, E., Rohde, M., 2009. Individuality, normativity, asymmetry and spatio-temporality in action. J. Adapt. Behav. 1–13.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2008. Associative learning of social value. Nature 456, 245–249.

Beintema, J.A., Lappe, M., 2002. Perception of biological motion without local image motion. Proc. Natl. Acad. Sci. U.S.A. 99, 5661–5663.

Bian, X., Elgar, M.A., Peters, R.A., 2016. The swaying behavior of extatosoma tiaratum: motion camouflage in a stick insect? Behav. Ecol. 27, 83–92.

Billington, J., Wilkie, R.M., Field, D.T., Wann, J.P., 2011. Neural processing of imminent collision in humans. Proc. R. Soc. B Biol. Sci. 278, 1476–1481.

Blake, R., Shiffrar, M., 2007. Perception of human motion. Annu. Rev. Psychol. 58, 47–73.

Brault, S., Bideau, B., Kulpa, R., Craig, C.M., 2012. Detecting deception in movement: the case of the side-step in rugby. PLoS One 7, e37494.

Brembs, B., 2011. Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 278, 930–939.

Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. NeuroImage 11, 157–166.

Burgess, N., Maguire, E.A., O'Keefe, J., 2002. The human hippocampus and spatial and episodic memory. Neuron 35, 625–641.

Caggiano, V., Fleischer, F., Pomper, J.K., Giese, M.A., Thier, P., 2016. Mirror neurons in monkey premotor area F5 show tuning for critical features of visual causality perception. Curr. Biol. 26, 3077–3082.

Camerer, C., Ho, T.H., Chong, J.K., 2004. A cognitive hierarchy model of games. Q. J. Econ. 113, 861–898.

Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., Ma, J., 2014. Reaction time for object categorization is predicted by representational distance. J. Cogn. Neurosci. 26, 132–142.

Carter, E.J., Hodgins, J.K., Rakison, D.H., 2011. Exploring the neural correlates of goal-directed action and intention understanding. NeuroImage 54, 1634–1642.

Carter, R.M.K., Bowling, D.L., Reeck, C., Huettel, S.A., 2012. A distinct role of the temporal-parietal junction in predicting socially guided decisions. Science 336, 109–111.

Casile, A., 2010. Neuronal encoding of human kinematic invariants during action observation. Cereb. Cortex 20, 1647–1655.

Catania, K.C., 2010. Born knowing: tentacled snakes innately predict future prey behavior. PLoS ONE 5, e10953.

Catania, K.C., 2011. The brain and behavior of the tentacled snake. Ann. N. Y. Acad. Sci. 1225, 83–89.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., Becchio, C., 2016. Decoding intentions from movement kinematics. Sci. Rep. 6, 37036.

Charpentier, C.J., O'Doherty, J.P., 2018. The application of computational models to social neuroscience: promises and pitfalls. Soc. Neurosci. 13, 637–647.

Cook, R., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: from origin to function. Behav. Brain Sci. 37, 177–192.

Coricelli, G., Nagel, R., 2009. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 106, 9163–9168.

Cross, E.S., Ramsey, R., Liepelt, R., Prinz, W., Hamilton, de C., 2016. The shaping of social perception by stimulus and knowledge cues to human animacy. Philos. Trans. R. Soc. B Biol. Sci. 371, 20150075.

Csibra, G., 2003. Teleological and referential understanding of action in infancy. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358, 447–458.

Csibra, G., 2008. Goal attribution to inanimate agents by 6.5-month-old infants. Cognition 107, 705–717.

Csibra, G., Gergely, G., Biro, S., Koos, O., Brockbank, M., 1999. Goal attribution without agency cues: the perception of 'pure reason' in infancy. Cognition 72, 237–267.

Cuthill, I.C., Matchette, S.R., Scott-Samuel, N.E., 2019. Camouflage in a dynamic world. Curr. Opin. Behav. Sci. 30, 109–115.

Dayan, E., 2007. Neural representations of kinematic laws of motion: Evidence for action-perception coupling. Proc. Natl. Acad. Sci. 104, 20582–20587.

Dennett, D.C. The intentional stance. (MIT Press, 1987).

Devaine, M., Hollard, G., Daunizeau, J., 2014a. The social bayesian brain: does mentalizing make a difference when we learn? PLoS Comput. Biol. 10.

Devaine, M., Hollard, G., Daunizeau, J., 2014b. Theory of mind: did evolution fool us? PloS One 9, e87619.

Dima, D.C., Tomita, T.M., Honey, C.J., Isik, L., 2022. Social-affective features drive human representations of observed actions. eLife 11, e75027.

Domenici, P., Booth, D., Blagburn, J.M., Bacon, J.P., 2008. Cockroaches keep predators guessing by using preferred escape trajectories. Curr. Biol. CB 18, 1792–1796.

Eaton, R.C., Bombardieri, R.A., Meyer, D.L., 1977. The Mauthner-initiated startle response in teleost fish. J. Exp. Biol. 66, 65–81.

Field, D.T., Wann, J.P., 2005. Perceiving time to collision activates the sensorimotor cortex. Curr. Biol. 15, 453–458.

Fleischer, F., Christensen, A., Caggiano, V., Thier, P., Giese, M.A., 2012. Neural theory for the perception of causal actions. Psych. Res 1–47.

Fleischer, F., Caggiano, V., Thier, P., 2013. M.A. Giese Physiol. inspired Model Vis. Recognit. Transit. hand Actions, 33, pp. 6563–6580.

Fletcher, P.C., 1995. Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. Cognition 44, 283–296.

Franconeri, S.L., Simons, D.J., 2003. Moving and looming stimuli capture attention. Percept. Psychophys. 65, 999–1010.

Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. Biol. Cybern. 104, 137–160.

Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358, 459–473.

Gallagher, H., Jack, A., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. NeuroImage 16, 814–821.

Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2, 493–501.

Gao, T., Scholl, B.J., 2011. Chasing vs. stalking: Interrupting the perception of animacy. J. Exp. Psychol. Hum. Percept. Perform. 37, 669–684.

Gao, T., Newman, G.E., Scholl, B.J., 2009. The psychophysics of chasing: a case study in the perception of animacy. Cogn. Psychol. 59, 154–179.

Gao, T., Scholl, B.J., McCarthy, G., 2012. Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *J. Neurosci. Off*. J. Soc. Neurosci. 32, 14276–14280.

Gao, T., Baker, C.L., Tang, N., Xu, H., Tenenbaum, J.B., 2019. The cognitive architecture of perceived animacy: intention, attention, and memory. Cogn. Sci. 43, e12775.

Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. 4, 179–192.

Gray, K., Wegner, D.M., 2011. To escape blame, don't be a hero-Be a victim. J. Exp. Soc. Psychol. 47, 516–519.

Griessinger, T., Coricelli, G., 2015. The neuroeconomics of strategic interaction. Curr. Opin. Behav. Sci. 3, 73–79.

Grossman, E., et al., 2000. Brain areas involved in perception of biological motion. J. Cogn. Neurosci. 12, 711–720.

Gyulai, E., 2004. Considerations on perception of 'animacy' in the motion of a single object. Percept. Mot. Skills 99, 1014–1026.

Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc. Natl. Acad. Sci. U.S.A. 105, 6741–6746.

Heider, F., Simmel, M., 1944. An experimental study of apparent behaviour. Am. J. Psychol. 57, 243–259.

Henschel, A., Hortensius, R., Cross, E.S., 2020. Social cognition in the age of human–robot interaction. Trends Neurosci. 43, 373–384.

Heyes, C., 2014. Submentalizing: i am not really reading your mind. Perspect. Psychol. Sci. 9, 131–143.

Heyes, C., Catmur, C., 2022. What happened to mirror neurons? Perspect. Psychol. Sci. 17, 153–168.

Hill, C.A., et al., 2017. A causal account of the brain network computations underlying strategic social behavior. Nat. Neurosci. 20, 1142–1149.

Hovaidi-Ardestani, M., Saini, N., Martinez, A.M., Giese, M.A., 2018. Neural model for the visual recognition of animacy and social interaction. In: *Artificial Neural Networks and Machine Learning {\textendash} ICANN*. Springer, Cham, pp. 168–177.

Isik, L., Koldewyn, K., Beeler, D., Kanwisher, N., 2017. Perceiving social interactions in the posterior superior temporal sulcus. PNAS 114, E9145–E9152.

Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: a critique. Trends Cogn. Sci. 9, 21–25.

Jamali, M., et al., 2021. Single-neuronal predictions of others' beliefs in humans. Nature 591, 610–614.

Jeannerod, M., 2003 (MIT Press). In: Michael, Arbib (Ed.), Action Monitoring and Forward Control of Movements, vol. 1. The Handbook Of Brain Theory And Neural Networks, pp. 83–85.

Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. Percept. Psychophys. 14, 201–211.

Johnson, S.C., Shimizu, Y.A., Ok, S.-J., 2007. Actors and actions: the role of agent behavior in infants' attribution of goals. Cogn. Dev. 22, 310–322.

Kemmerer, D., 2021. What modulates the mirror neuron system during action observation? Prog. Neurobiol. 205, 102128.

Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. Cogn. Process. 8, 159–166.

Kircher, T., 2009. Online mentalising investigated with functional MRI. Neurosci. Lett. 454, 176–181.

Kliemann, D., Adolphs, R., 2018. The social neuroscience of mentalizing: challenges and recommendations. Curr. Opin. Psychol. 24, 1–6.

Konovalov, A., Hu, J., Ruff, C.C., 2018. Neurocomputational approaches to social behavior. Curr. Opin. Psychol. 24, 41–47.

Konovalov, A., Hill, C., Daunizeau, J., Ruff, C.C., 2021. Dissecting functional contributions of the social brain to strategic behavior. Neuron 109 (3323–3337), e5.

Koul, A., 2018. Action observation areas represent intentions from subtle kinematic features. Cereb. Cortex 28, 2647–2654.

Koul, A., Soriano, M., Tversky, B., Becchio, C., Cavallo, A., 2019. The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. Cognition 182, 213–219.

Lee, D., Seo, H., 2016. Neural basis of strategic decision making. Trends Neurosci. 39, 40–48.

Lee, S.M., Gao, T., McCarthy, G., 2014. Attributing intentions to random motion engages the posterior superior temporal sulcus. Soc. Cogn. Affect. Neurosci. 9, 81–87.

Leslie, A.To.M.M., 1994. ToBy, and Agency: core architecture and domain specificity. In: Hirschfeld, L.A., Gelman, S.A. (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press.

Liu, J., 2014. Seeing jesus in toast: neural and behavioral correlates of face pareidolia. Cortex 53, 60–77.

Liu, S., Spelke, E.S., 2017. Six-month-old infants expect agents to minimize the cost of their actions. Cognition 160, 35–42.

Lockwood, P.L., Apps, M.A.J., Chang, S.W.C., 2020. Is there a 'social' brain? Implementations and algorithms. Trends Cogn. Sci. 24, 802–813.

Luo, Y., Baillargeon, R., 2005. Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. Psychol. Sci. 16, 601–608.

Ma, N., Vandekerckhove, M., Van Hoeck, N., Van Overwalle, F., 2012. Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. Soc. Neurosci. 7, 591–605.

Mahmoodi, A., Bahrami, B., Mehring, C., 2018. Reciprocity of social influence. Nat. Commun. 9, 1–9.

Maier, J.X., Chandrasekaran, C., Ghazanfar, A.A., 2008. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. Curr. Biol. 18, 963–968.

Mar, R.A., Macrae, C.N., 2008. Triggering the intentional stance. In: Bock, G., Goode, J. (Eds.), *Novartis Foundation Symposia*. John Wiley & Sons Ltd, pp. 111–133. https://doi.org/10.1002/9780470030585.ch9.

Mckee, S.P., Nakayama, K., 1984. The detection of motion in the peripheral visual field. Vis. Res. 24, 25–32.

Merchant, H., Georgopoulos, A.P., 2006. Neurophysiology of perceptual and motor aspects of interception. J. Neurophysiol. 95, 1–13.

Michotte, A. La perception de la causalité. (Institut superieur de philosophie, 1946).

Mizutani, A., Chahl, J.S., Srinivasan, M.V., 2003. Motion camouflage in dragonflies. Nature 423, 604–606.

Nagel, R., Brovelli, A., Heinemann, F., Coricelli, G., 2018. Neural mechanisms mediating degrees of strategic uncertainty. Soc. Cogn. Affect. Neurosci. 13, 52–62.

Neri, P., Morrone, M.C., Burr, D.C., 1998. Seeing biological motion. Nature 395, 894–896.

Nicolle, A., 2012. An agent independent axis for executed and modeled choice in medial prefrontal cortex. Neuron 75, 1114–1121.

Ogawa, A., Kameda, T., 2019. Dissociable roles of left and right temporoparietal junction in strategic competitive interaction. Soc. Cogn. Affect. Neurosci. 14, 1037–1048.

Opfer, J.E., 2002. Identifying living and sentient kinds from dynamic information: the case of goal-directed versus aimless autonomous movement in conceptual change. Cognition 86, 97–122.

Patri, J.-F., 2020. Transient disruption of the inferior parietal lobule impairs the ability to attribute intention to action. Curr. Biol. 30 (4594–4605), e7.

Pelphrey, K., Singerman, J., Allison, T., McCarthy, G., 2003. Brain activation evoked by perception of gaze shifts: the influence of context. Neuropsychologia 41, 156–170.

Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J. Cogn. Neurosci. 16, 1706–1716.

Press, C., 2011. Action observation and robotic agents: learning and anthropomorphism. Neurosci. Biobehav. Rev. 35, 1410–1418.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. 18, 2188–2199.

Pyles, J.A., Garcia, J.O., Hoffman, D.D., Grossman, E.D., 2007. Visual perception and neural correlates of novel 'biological motion'. Vis. Res 47, 2786–2797.

Ramsey, R., Ward, R., 2020. Putting the nonsocial into social neuroscience: a role for domain-general priority maps during social interactions. Perspect. Psychol. Sci. 15, 1076–1094.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025.

Ritchie, J.B., de Beeck, Op, 2019. Using neural distance to predict reaction time for categorizing the animacy, shape, and abstract properties of objects. Sci. Rep. 9, 13201.

Ritchie, J.B., Tovar, D.A., Carlson, T.A., 2015. Emerging object representations in the visual system predict reaction times for categorization. PLoS Comput. Biol. 11, e1004316.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., Speed, M.P., 2018a. Avoiding Attack: The Evolutionary Ecology of Crypsis. Aposematism, and Mimicry. Oxford University Press. https://doi.org/10.1093/oso/9780199688678.001.0001.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., Speed, M.P., 2018b. Thanatosis. *Avoiding Attack*. Oxford University Press. https://doi.org/10.1093/oso/9780199688678.003.0014.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision making. Science 300, 1755–1758.

Savaki, H.E., Kavroulakis, E., Papadaki, E., Maris, T.G., Simos, P.G., 2022. Action observation responses are influenced by movement kinematics and target identity. Cereb. Cortex 32, 490–503.

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. Trends Cogn. Sci. 19, 65–72.

Scholl, B.J., Tremoulet, P.D., 2000. Perceptual causality and animacy. Trends Cogn. Sci. 4, 299–309.

Scholl, B.J., Gao, T., 2013. Perceiving animacy and intentionality: visual processing or higher-level judgment. in social perception: detection and interpretation of a*nimacy*. In: Rutherford, M.D., Kuhlmeier, V.A. (Eds.), Agency, and Intention. MIT Press, Cambridge, MA, pp. 197–231.

Schultz, J., Bülthoff, H.H., 2013. Parametric animacy percept evoked by a single moving dot mimicking natural stimuli. J. Vis. 13, 15.

Schultz, J., Bülthoff, H.H., 2019. Perceiving animacy purely from visual motion cues involves intraparietal sulcus. NeuroImage 197, 120–132.

Schultz, J., Imamizu, H., Kawato, M., Frith, C.D., 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. J. Cogn. Neurosci. 16, 1695–1705.

Schultz, J., Friston, K.J., O'Doherty, J., Wolpert, D.M., Frith, C.D., 2005. Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. Neuron 45, 625–635.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34.

Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Sallet, J., Kanske, P., 2021. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol. Bull. 147, 293–327. https://doi.org/10.1037/bul0000303.

Skelhorn, J., 2018. Avoiding death by feigning death. Curr. Biol. 28, R1135–R1136.

Smeets, J.B.J., Brenner, E., 1995. Perception and action are based on the same visual information: distinction between position and velocity. J. Exp. Psychol. Hum. Percept. Perform. 21, 19–31.

Southgate, V., Johnson, M.H., Csibra, G., 2008. Infants attribute goals even to biomechanically impossible actions. Cognition 107, 1059–1069.

Sperry, R.W., 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. J. Comp. Physiol. Psychol. 43, 482–489.

Srinivasan, M.V., Davey, M., 1995. Strategies for active camouflage of motion. Proc. R. Soc. B Biol. Sci. 259, 19–25.

Stanley, J., Gowen, E., Miall, R.C., 2007. Effects of agency on movement interference during observation of a moving dot stimulus. J. Exp. Psychol. Hum. Percept. Perform. 33, 915–926.

Stanley, J., Gowen, E., Miall, R.C., 2010. How instructions modify perception: an fMRI study investigating brain areas involved in attributing human agency. NeuroImage 52, 389–400.

Stewart, J.A. Perception of Animacy. (Unpublished PhD Thesis, 1982).

Stosic, M., Brass, M., Van Hoeck, N., Ma, N., Van Overwalle, F., 2014. Brain activation related to the perception of minimal agency cues: The role of the mirror system. NeuroImage 86, 364–369.

Suzuki, S., O'Doherty, J.P., 2020. Breaking human social decision making into multiple components and then putting them together again. Cortex 127, 221–230.

Szego, P.A., Rutherford, M.D., 2007. Actual and illusory differences in constant speed influence the perception of animacy similarly. J. Vis. 7, 1–7.

Takahashi, H., et al., 2014. Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain. Cortex 58, 289–300.

Takahashi, H., Izuma, K., Matsumoto, M., Matsumoto, K., Omori, T., 2015. The anterior insula tracks behavioral entropy during an interpersonal competitive game. PLoS One 10, e0123329.

Tamir, D.I., Thornton, M.A., 2018. Modeling the predictive social mind. Trends Cogn. Sci. 22, 201–212.

Tarhan, L., De Freitas, J., Konkle, T., 2021. Behavioral and neural representations en route to intuitive action understanding. Neuropsychologia 163, 108048.

Tcheang, L., Bülthoff, H.H., Burgess, N., 2011. Visual influence on path integration in darkness indicates a multimodal representation of large-scale space. Proc. Natl. Acad. Sci. U.S.A. 108, 1152–1157.

Thornton, M.A., Weaverdyck, M.E., Tamir, D.I., 2019a. The brain represents people as the mental states they habitually experience. Nat. Commun. 10, 2291.

Thornton, M.A., Weaverdyck, M.E., Tamir, D.I., 2019b. The social brain automatically predicts others' future mental states. J. Neurosci. 39, 140–148.

Thurman, S.M., Lu, H., 2013. Physical and biological constraints govern perceived animacy of scrambled human forms. Psychol. Sci. 24, 1133–1141.

Tremoulet, P.D., Feldman, J., 2000. Perception of animacy from the motion of a single object. Perception 29, 943–951.

Tremoulet, P.D., Feldman, J., 2006. The influence of spatial context and the role of Intentionality in the Interpretation of animacy from motion. Percept. Psychophys. 68, 1047–1058.

Ullman, S., Harari, D., Dorfman, N., 2012. From simple innate biases to complex visual concepts. Proc. Natl. Acad. Sci. U.S.A. 109, 18215–18220.

Urgen, B.A., Saygin, A.P., 2020. Predictive processing account of action perception: Evidence from effective connectivity in the action observation network. Cortex 128, 132–142.

Vannuscorps, G., Caramazza, A., 2016. Typical action perception and interpretation without motor simulation. Proc. Natl. Acad. Sci. U.S.A. 113, 86–91.

Vesper, C., Richardson, M.J., 2014. Strategic communication and behavioral coupling in asymmetric joint action. Exp. Brain Res. 232, 2945–2956.

Wheatley, T., Milleville, S.C., Martin, A., 2007. Understanding animate agents: distinct roles for the social network and mirror system. Psychol. Sci. 18, 469–474.

Wiederman, S.D., Fabian, J.M., Dunbier, J.R., O'Carroll, D.C., 2017. A predictive focus of gain modulation encodes target trajectories in insect vision. eLife 6, 1–19.

Wittmann, M.K., Kolling, N., Akaishi, R., Chau, B.K.H., Brown, J.W., Nelissen, N., Rushworth, M.F.S., 2016. Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nat. Commun. 7 (1), 12327 https://doi.org/10.1038/ncomms12327.

Woodward, A.L., 1998. Infants selectively encode the goal object of an actor's reach. Cognition 69, 1–34.

Yager, D.D., May, M.L., Fenton, M.B., 1990. Ultrasound-triggered, flight-gated evasive maneuvers in the praying mantis Parasphendale agrionina. I. Free flight. J. Exp. Biol. 152, 17–39.

Yoo, S.B.M., Tu, J.C., Piantadosi, S.T., Hayden, B.Y., 2020. The neural basis of predictive pursuit. Nat. Neurosci. 23.