HOW ANIMALS ASCRIBE UNOBSERVABLE GOALS

Angelica Kaufmann

Harousch and Williams (2015) discovered, for the first time, the existence of neurons, in the dorsal anterior cingulate cortex of the macaque's brain, that predict another individual's unobservable, unknown, and future goal. Moreover, they discovered a causal relationship between this activation and the specific enactment of mutually beneficial decisions. The paper explains that one way of framing these findings theoretically is to argue that cingulate neurons get activated, distinctively, when animals foresee their own and others distal intentions. Joint distal intentions are defined as mental representations of goals to which joint plans are directed. Researchers have long known that mirror neurons get activated when an agent performs or observe others performing the very same action (Rizzolatti & Sinigaglia 2010). But is observable goal ascription and unobservable goal ascription served by the same neural circuits?

KEYWORDS: Animal Cognition, Cingulate Neurons; Goal; Intention; Joint Action.

WORD COUNT: 8982 (Including references)

ADDRESS OF CORRESPONDENCE:

Angelica Kaufmann, PhD

Columbia University
Italian Academy for Advanced Studies
1161 Amsterdam Avenue
New York, 10027

Email: angelica.kaufmann@gmail.com
ak3922@columbia.edu

HOW ANIMALS ASCRIBE UNOBSERVABLE GOALS

1. INTRODUCTION

Two adventurers, Laurence and Faisal, are kept prisoners in a stronghold in the middle of the Sahara desert. The enemies who imprisoned them abandoned the base camp leaving the prisoners with only seven gallons of water. Since the jailers left, Laurence and Faisal are optimistically waiting for someone to rescue them. But how long will this wait last? They will eventually run out of water. Laurence knows that his chances to last longer will increase if he manages to get most of the gallons for himself. But his overall chances of surviving will decrease if Faisal dies of thirst, leaving him on his own, in the middle of nowhere. Faisal has the same thought. Their best chance to survive depends on their equal sharing of the gallons. All they need to do is trusting that the other will not take all the gallons for himself. They had better think so, if they want to increase their overall chances of surviving. This depends on thinking ahead, pursuing an equal sharing of the gallons of water, and planning the success of their being eventually rescued. As time passes, the two fellow prisoners have not defeated each other, and this reinforces their motivation to keep an equal sharing of water. Then, one day, Laurence and Faisal hear the sound of an aircraft engine approaching: the rescuers are coming.

This story is not about altruism and cooperation, but it illustrates the difference between pursuing individual and joint distal intentions, and under which circumstances, at least in some cases, like that of Laurence and Faisal, the optimal solution is only achieved by pursuing joint distal intentions.

Distal intentions can be defined as mental representations of goals to which plans are directed. In some social interactions, these mental states include a representation of the unobservable goal of others. In such circumstances we can invoke the notion of joint distal intentions. These can be defined as mental representations of goals to which joint plans are directed (Axelrod, 1984, 1987).

As it will be explained, distal intentions can be differentiated by other goal states. I will draw the distinction between the two states as follows: there is a mechanism known as episodic memory system that is considered responsible for the capacity to foresee one's own future goal-directed actions without exploiting primarily information from perceptual and motor cues (Raby & Clayton, 2012; Ban et al., 2014). Following evidence

for the existence of such mechanism, I argue that the episodic memory system is distinctively activated when distal intentions, i.e. nonperceptual goal states, are instantiated, but not when perceptual goal states are represented. For this reason, I do not analyze the anti-representationalist accounts that could be used to investigate future goal-directed nonhuman animal behaviour (see Gallagher 2001, 2005, 2008; Zahavi 2008; Ratcliffe 2007; Hutto 2007). These accounts are interested in the direct-perceptual based ascription of goal states. But distal intentions are primarily nonperceptual states.

The step further would be to explain what mechanism is responsible for instantiating joint distal intentions specifically.

The aim of the paper is stimulated by the following claim: Some have conjectured that "our capacities for certain forms of shared activity set us apart as a species" (Bratman, 2014, p. 3)[1]. This paper partially counters this view by exploring and analysing what may be the neural underpinnings of joint distal intentions in nonhuman animals; and it encourages to implement recent findings from neuroscience with evidence from the field of cognitive ethology. As said, this paper does not want to argue for nonhuman animal cooperativeness, but simply explore the possibility that nonhuman animals too can ascribe certain unobservable goals to others and act jointly on the basis of this mutual appreciation. This paper supplements the claim that nonhuman animals can master joint distal intentions by exploring the neural underpinnings of both individual and joint action planning. And evidence is analysed for the claim that the same neural mechanism is responsible for the activation of both individual and joint distal intentions.

Researchers have long known that mirror neurons get activated when an agent performs or observes others performing the very same action (Rizzolatti and Sinigaglia, 2010). But evidence for the neural underpinnings of predictive mechanisms that guide animal individual and joint action planning has, so far, only been hypothesised. Harousch and Williams (2015) discovered, for the first time, the existence of neurons, in the dorsal anterior cingulate cortex[2] of the macaque's brain, that predict another individual's unobservable, unknown, and future goal (analogous results in humans are reported by Suzuki et al., 2015). Moreover, they discovered a causal relationship between this activation and the specific enactment of mutually beneficial decisions. Drawing on these results, I argue for the hypothesis that cingulate neurons get activated, distinctively, when

[1] Notably, Michael Bratman's view on shared agency is meant to explain the social dimension of human agency only, however Bratman is sympathetic to the idea that the general structure of his view is broad enough to play an explanatory role for the analysis of much more modest instances of social interactions.

[2] In humans the dACC is known to provide a continuously updated prediction of expected cognitive demand to optimize future behavioural responses (Sheth et al., 2012)

we foresee our own and others' distal intentions. The proposal is that the localization of this circuit and the clarification of its function could enrich our appreciation of the mechanisms and of the degree of complexity of long-term social interactions among nonhuman animals.

In the animal kingdom, especially among primates, there are populations of individuals that exhibit complex purposeful group behaviour. Some researchers (Boesch & Boesch-Achermann, 2000) adopted a Rich Account, and argued that nonhuman animals can have a common distal goal and take complementary roles in their joint activities. Some other researchers (Tomasello, 2014) defended a Lean Account, and claimed that, in these group activities, each participant is only attempting to maximize its own chances to pursue a proximal goal. Much of what this paper argues for is motivated by the necessity to deepen our understanding of the mechanisms at the basis of such complex purposeful group behaviour. In this work I try to keep separate the study of the motivational drive behind group activities, and the study of the cognitive capacities that are needed to engage in such group activities. For the concern here is not with what nonhuman animals want to achieve, but with what they can achieve (see also, AUTHOR, FORTHCOMING, 2016).

Now imagine that Laurence and Faisal are not fighting for life, but rather for a brick of apple juice. Laurence and Faisal are two rhesus monkeys (*Macaca mulatta*). They are facing each other, sitting in front of a computer screen. They hold joysticks with which they have to choose over two symbols on the screen. One symbol will deliver more juice to oneself, and the other symbol will deliver an equal sharing of the juice. So, Laurence makes his selection, and Faisal makes his selection, but if both go for the symbol that only delivers juice to themselves, neither of them gets much juice. This is the set-up of the experiment that led Harousch and Williams (2015) to their discovery. They show that, over time, the monkeys tend to favour the equal sharing of the bricks, or what is known as the Pareto optimality condition (Nash, 1950). The prediction of this theory is that two or more agents get the highest payoff, if and only if they both cooperate and maintain this behaviour over time. Harousch and Williams were looking for the distinctive neural circuits responsible for this capacity to understand the unobservable goals of others, that is, I shall argue, dependent on the ability to ascribe distal intentions.

This is the structure of the paper: firstly, I shall outline the distinction between ascribing unobservable goal states, like some distal intentions, and ascribing other observable goal states, like motor representations; secondly, I am going to explain what

neural mechanisms are thought to underpin the ascription of observable goal states, like motor representations, and what neural mechanisms are thought to underpin the ascription of unobservable goal states, like distal intentions. Crucially, the same neural activation is at play during the pursuit of individual and joint distal intentions. Then, I am going to shortly illustrate that the results of Harousch and Williams could enrich the theoretical paradigms developed on evidence for similar behaviour that researchers gained from an ethological approach to the study of social interactions among nonhuman animals (Eaton, 1972; Huffman & Quiatt, 1986). I shall conclude that the capacity to ascribe unobservable goal states depends on the capacity to ascribe joint distal intentions, where joint distal intentions are understood as representations of goals to which joint plans are directed. This ability is served by a specific neural circuit, different from that exploited during goal ascription inferred from observable actions. The proposal is that the localization of this circuit and of its function could enrich our appreciation of the debate between the Rich and the Lean Account about the capacity to entertain social interactions among nonhuman animals (this debate is discussed at full length in AUTHOR, FORTHCOMING, 2016).

2. ASCRIBING DISTAL INTENTIONS AND OTHER GOAL STATES

Some purposeful actions can be explained by appealing to goal states, like motor representations, and some other purposeful actions invoke an additional causal component of goal states, like distal intentions. Evidence points to a neat distinction between the function of goal states and that of distal intentions. Most notably, goal states, such as motor representations, represent outcomes to which actions are directed, and they guide goal-oriented anticipatory behaviour (Butterfill & Sinigaglia, 2014; Jeannerod 1994a, 1997; Pacherie 2000, p. 409; Sinigaglia & Butterfill, 2015)[3]. Distal intentions represent goals to which plans are directed, and they guide action planning. They can represent individual goals or joint goals. In the latter case, that is, in social interactions, motor representations should be responsible for the ascription of

[3] I shall acknowledge that, in the literature, the notion of intention is commonly referred to in relationship to one of the various "dual-intention theories" (Pacherie, 2008). John Searle (1983) distinguishes between prior intentions and intentions in action, Myles Brand (1984) between prospective and immediate intentions, Michael Bratman (1987) between future-directed and present-directed intentions, Alfred Mele (1992) between distal and proximal intentions. Here, I do not assimilate these two notions, but I only deal with the former, i.e. distal intention.

observable goals, whereas distal intentions should be responsible for the ascription of unobservable goals.

My characterization of distal intentions draws on Alfred Mele's analysis: "The causal contribution of intention is traceable both to motivational aspects of intention and to representational features. Intentions *move* us to act in virtue of their motivational properties and *guide* our intentional behaviour in virtue of their representational qualities" (Mele, 1990, p. 289).

According to Mele's description, joint distal intentions are here defined as representations of goals to which joint plans are directed. These mental states have the following representational features: distal intentions (1) are personal-level states, i.e. states of the *other* individual(s) rather than her subsystems; (2) they represent some unobservable goal of the *other* individual(s); and, (3) they have fulfilment conditions, which are satisfied if the intention causes the action that ultimately achieves the goal represented by the *other* individual(s).

Let us give an example. We are back to the desert and Faisal's individual distal intention to drink his gallon of water tomorrow is fulfilled just in case he will actually drink his gallon of water tomorrow (barring all cases where if Faisal fails to drink his share of water it will be because of circumstances that are beyond his control). Otherwise, his intention is simply one among many desires, none of which is significantly predominant with respect to the others. Similarly, Faisal's joint distal intention that Laurence will drink his gallon of water tomorrow is fulfilled if, and only if, Laurence will actually drink is gallon of water tomorrow. In accordance with the Rich Account, mentioned earlier, a joint distal intention is the result of the mutual ascription of individual distal intentions.

By contrast, the ascription of goal states like motor representations is subject to accuracy conditions, which are satisfied if the motor representation accurately guides a correct ascription of the action of the other individual towards the accomplishment of the represented outcome. For example, Faisal could ascribe to Laurence the following goal state: a motor representation of Laurence's outcome "grasp that tank of water", which is a representation of the relation between the size of the tank of water and the degree of openness of Laurence's index finger and of his thumb. This motor representation has to be accurate, not only for Laurence's hand in motion to have a

precision grip over the tank of water, but also for Faisal's capacity to accurately ascribe to Laurence this motor representation.

In sum, the representational success of the ascription of these different goal states is obtained under different conditions of satisfaction. The successful ascription of some goals can be guided by goal states like motor representations, but the successful attribution of some specific goals can only be guided by joint distal intentions.

For example, Faisal's distal intention to drink his share of water tonight may not achieve representational success due to contingent features of his current environment, or to a competing agent. As an example of the former, he may not have the tank of water in front of himself and thus he may not be able to represent action-properties to ascribe to this object, i.e. the tank of water. An example of the latter may be given by Laurence's stealing his share of the provision, so that accuracy conditions cannot be met. Faisal's distal intention is not subject to accuracy conditions, but rather to fulfilment conditions. Indeed, he cannot have any motor representation such as "to drink the water tonight". On the other hand, a goal state like a motor representation will be very useful when the goal to drink the water 'tonight' will have become the goal to drink the water 'now'. This goal state is subject to accuracy conditions because it achieves representational success in the light of contingent features, i.e. action outcomes of Faisal's current environment.

Further reasons to treat motor representations and distal intentions as distinct goal states come from investigations on the neural mechanisms that underpin social interactions. In these circumstances, we can ask: are the ascription of distal intentions and the ascription of other goal states served by the same neural circuits? Recent evidence points to a negative answer: the mechanisms are different. They might be related in enabling functions, but discernible in preliminary activations.

I want to understand which neural circuits trigger the ascription of distal intentions, and hence deepen our understanding of the ability to plan individual and joint actions. The latter has been defined as the capacity to articulate distal intentions in cross-temporal coordination: "These plans help guide our later conduct and coordinate our activities over time […]—where intentions are typically elements in such coordinating plan" (Bratman, 1984, p. 386).

Notably, Michael Bratman argued that distal intentions, which he specifically calls *plan-states*, are different from other goal states. This is because, he argues, they are subject to special demands for agglomeration and consistency. The way in which plan-states

work is by committing ourselves to action in advance of action execution. Not only that, but distal intentions especially enable the agent to adapt her plan to changing circumstances as time passes.

At the level of the single individual, we can appreciate this distinction in the analysis of the activation of the brain areas constituting the Episodic Memory System (Tulving, 1993; Raby & Clayton, 2012). This system gets activated during action planning, e.g. saving a tank of water for tonight, but not during goal-oriented anticipatory behaviour, e.g. preparing to drink the tank of water now (Raby & Clayton, 2012). Episodic memory is a type of declarative[4] memory that consists in the capacity to, implicitly or explicitly, recall one's own past experiences (either as an observer or as a participant). It is also involved in the mechanism that allows an agent to mentally imagine herself acting in future events (Mullay & Maguire, 2014). Episodic memory not only retrieves memories of past experiences but also produces novel constructs for prospective events. Thus, episodic memory is distinctively activated during action planning, but not during goal-directed anticipatory behaviour driven, for instance, by motor representations. A crucial feature of episodic memory is its being at play when an animal recalls either having observed or having participated in an activity. In short, what episodic memory does is allowing: (1) a flexible deployment of the information available from the event in new situations; (2) a recollection of what happened, where, and when on the basis of a specific past experience; and (3) the formation of an integrated "what-where-when" representation of a past or future state of affairs (see Tulving, 1972). In particular, the activation of the PCC (posterior cingulate cortex) is responsible for the successful recollection of previously experienced events, which are subsequently exploited to generate optimal predictions for action planning. Within Laurence and Faisal's story, we could say that the role of the episodic memory system is that of allowing the reinforcement of mutual trust that they will not defeat each other in the future because their past experience taught them so.

More specifically, the relationship between episodic memory and action planning is the following: Firstly, we have Tulving's (1972) definition of episodic memory, i.e., the "where-what-when" definition. This is the capacity to retrieve and integrate information about the three spatiotemporal features of a single event. Secondly, drawing on the

[4] Declarative memory is the memory about conscious facts and events. There are two type of declarative memory: episodic memory and semantic memory. The latter consists in recalling knowledge for facts that an agent has not personally experienced neither as an agent nor as an observer. Non-declarative memory, instead, is procedural memory, which is expressed through performance rather than recollection of facts.

evidence of patients with memory impairments and neuroimaging studies, it has been, convincingly, hypothesized that if a creature can episodically recall the past, then she should also be capable of planning for the future, since memory is a reconstructive process rather than merely a reproductive one (Clayton, Salwiczek & Dickinson, 2007).

As explained, further evidence for appreciating the role that episodic memory plays in action planning can be obtained from the analysis of fMRI studies on memory impairments and cases of amnesia. Patients with severe memory deficits that display difficulties in tasks that require episodically recalling the past, additionally display difficulties in planning for the future. This correlation has suggested that the same cognitive processes are involved in remembering the past and imagining the future (Atance & O'Neill, 2001; Rosenbaum et al., 2005).

This is evidence for the distinctiveness of the neural system that activates distal intention in the case of individual planning. Arguably, the same claim can be extended to joint planning. One possibility is that there may be a complementary activation between the PCC, which, as said, is responsible for the successful recollection of previously experienced events, and the dACC (dorsal anterior cingulate cortex) which, as shall be explained, gets activated during the recognition of the distal intentions of others. And the whole cingulate cortex, hence including both PCC and dACC, displays a nearly indiscernible connectivity. However interesting, it is an as yet unexplored question to what extent the Episodic Memory System is active during an individual's engagement in joint distal planning.

The claim of this paper is that the capacity for ascribing distal intentions that refer to unobservable future actions is served by a specific neural circuit, different from that exploited during goal ascription inferred from observable actions: the neural activation occurring during unobservable future goal ascription is located in the dorsal anterior cingulate of the brain of the macaque (Harousch and Williams, 2015). This is because during the course of those social interactions that are guided by joint distal intentions, the two or more agents involved need to have some understanding of the other agent's (or agents') non-observable goals. These are future actions that cannot be inferred from perceptual or motor inputs only, but only from the ascription of distal intentions.

Before exploring the details of this recent finding and of how the cingulate gyrus generates the representation of distal intentions and the capacity to attribute these mental states to oneself and to the other(s), I will now explain how the activation of the brain

areas that constitute the Mirror Neuron System generates motor representations, and the capacity to attribute observable goals to oneself and to the other(s).

3. THE NEURAL BASIS OF OBSERVABLE GOAL ASCRIPTION

The prefrontal cortex is the area of the human brain that yields intelligence in social behaviour. This is why this area has been called "the organ of civilisation" (Goldberg & Bougakov, 2007). In humans, the larger dimension of this brain area and the high density of paths connecting it with sensory regions has lead scholars in neuroscience to suppose that the prefrontal cortex provides humans with a distinctive capacity to predict others' behaviour. This hypothesis has been widely investigated after the discovery of the properties of neurons in area F5 in the premotor cortex of the macaque's brain (Frith and Frith, 1999; Gallese and Goldman, 1998; Rilling et al., 2004; Sanfey et al., 2006; Vogeley et al., 2001). As we have seen, there is compelling evidence for the claim that nonhuman animals have the capacity to ascribe goals to others by means of observable actions of others.

As explained, motor representations have a good explanatory power for the functioning of observable action recognition and for the ascription of corresponding goals (Flanagan & Johansson, 2003; Wolpert 2003; Ambrosini et al., 2011, 2012; Costantini et al., 2013; Costantini et al., 2014). This is due to the motor properties of certain neurons. In the rostral part of the ventral premotor cortex of the rhesus monkey brain, which is called area F5, a category of neurons with motor properties is located. These neurons discharge in correlation with action performance rather than just in correlation with the particular movement that forms that action. A subclass of these neurons has a more specific property, namely that of discharging not only during the performance of an action but also during the observation of the performance of a similar action. An example is grasping or manipulating an object or observing someone else grasping or manipulating an object. This property consists in a responsiveness to visual stimuli of a certain nature: in order for the neural activation to occur, a specific purposeful action, e.g. grasping (mainly with the hand, less frequently by the mouth—Rizzolatti et al., 1996—or by the foot—Rizzolatti et al., 2009), has to take place. F5 neurons with this property have been named "mirror neurons". More specifically, the area that activates during ascription of goal states such as motor representations, is the VIP (ventral intraparietal area, Ishida et al., 2009). The ventral intraparietal area (VIP) of

the macaque brain is a multimodal cortical region where the activation of bimodal neurons is selective to specific visual and vestibular stimuli (Chen et al., 2013).

It has been proposed that area F5 should be the monkey homolog of Broca's area in the left hemisphere of the human brain (Rizzolatti et al., 1996). Though F5 has been labelled as an area that activates for hand movements, Broca's area is commonly known as being devoted to language. But Broca's area also has motor properties, especially related to hand and arm movements. The hypothesis is that mirror neurons are located in the human brain as well, in Broca's area. Both areas, F5 and Broca's, control oro-laryngeal, oro-facial and brachio-manual movements, and both have a neural structure that links action performance with action perception. These neurons code for some kinds of action understanding, so it has been claimed that the function of mirror neurons is that of building a representation of certain purposive actions (Rizzolatti et al., 1996; Jeannerod et al. 1995). This is a necessary requirement for action imitation and, especially, for goal ascription. As Rizzolatti & Arbib (1998) argue, ascribing a goal from the observation of an action occurs at three levels: (i) recognition of the performance of an action, (ii) identification of a specific action and (iii) performance of another action in accordance with the information obtained from the observed action.

The great interest devoted to mirror neurons is due to the fact that their advocates have proposed that this mechanism provides the neural substrate that enables the understanding of other people's goals and intentions simply through action observation. I am attempting to clarify that ascribing distal intentions means ascribing a very specific type of goal and this capacity is provided by a very specific neural mechanism, not to be confused with that responsible for other instances of goal ascription. A problem in clarifying this has been pointed out, namely the fact that identical movements can be made when performing different actions with different goals. But the same sensory input can have many causes. And there are, at least, four different degrees of description for labelling the same action with different meanings: (1) the kinematic level, that describes the bodily configuration in space and time; (2) the motor signals, that describe the action execution through the corresponding muscle activation; (3) the goal, that describes a short-term objective that is needed in order to achieve a purpose; and (4) the intention, that defines the long-term goal of an action. The idea is that the information passes from the level of (1) kinematic to the level of (2) representation, which can, ultimately be a (3) motor representation, or—before the proposal of an account for identifying motor representations as goal states—(4) a distal

intention (Kilner et al., 2007).

A proposal is that the MNS, Mirror Neurons System, works according to the principles of predictive coding, based on a statistical approach that can be associated with empirical Bayesian inference (Kilner et al., 2007). The function of predictive coding is minimising the possibility of errors between frequent and reciprocal cortical interactions, and therefore minimising the chance of error in the organisation of a hierarchical structure such as that of thoughts. If the prediction about each level of description of an action can be estimated upon the prediction made on the respective lower level, then the expectations based on lower levels can provide a guide to the understanding of the most likely purpose of an action as a whole, i.e. having considered all the three (or four, depending on the interpretation, degrees of description that can be applied to that action) levels. The same models that are exploited to infer motor commands from the observed kinematics produced by others during perceptual inference have analogous computational roles in the domain of goal ascription (Chater et al., 2006).

Taking a closer look at how the MNS works, the motor cortex controls individual synergies, that is, simple movements, and the premotor cortex structures instances of simple motor behaviour into coordinated motor acts. The motor and premotor cortex structure action execution and action perception, imitation and imagination, with neural connections to motor effectors and/or other sensory cortical areas. The premotor system activation can occur in two modalities: in the first modality, neurons fire when the action is performed or imitated, and when the action is observed or imagined, part of the neurons fire showing an activation in the motor system which in this case does not relate to performing, as it is not connected to the motor system, but rather just to understanding that there is a purposeful action going on. In the second modality, the system deactivates those of its functions that concern performance and observation of an action. By means of this uncoupling, the premotor system engages with primarily non-sensorimotor areas, namely those areas in the dorsal prefrontal cortex that regulate the hierarchical structuring of though. So, both types of awareness, that of one's own and that of the other's goal, correlate with the same network activation but by different patterns of the brain.

In summary, mirror neurons appear to be very sensitive to very small differences between some nearly analogous observed actions, even when the kinematics of the movements is very similar. But, as I shall explain in the following section, in order to

increase the effectiveness of predictions it turns out to be quite useful to learn to interpret others' distal intentions, and not simply others' observable goals. In particular when goal ascription is non-perceptual or primarily non-motorically driven, and it is distal in time and complexity, exploiting these predictive skills is very effective—especially when we consider, as I did, that the same movement can have different outcomes induced by different purposes.

It seems that there is a mismatch between, on the one hand, the continuity in primate evolution of social cognition, and, on the other hand, the discontinuity among primates in the capacity to process complex recursive structures in the way humans do. The continuity consists in the capacity of both human and non-human primates to ascribe goals on the basis of observable behaviour. The discontinuity is due to the higher computational power of the human premotor cortex that was thought to be necessary to understand also unobservable goals. For this reason, before the recent discovery of the function of cingulate neurons that I am about to discuss, the hypothesis that nonhuman animals could appreciate and ascribe distal intentions was merely speculative. But the recent discovery of neurons with primarily non-motor properties that are responsible for the ascription of unobservable future goals of others pushes the debate onto a new ground: nonhuman animals, while lacking the higher computational power of the human premotor cortex, may be exploiting other neural resources to appreciate the distal intentions of others, and pursue joint plans on the basis of this appreciation. In the following section I discuss evidence for this proposal.

4. THE NEURAL BASIS OF UNOBSERVABLE GOAL ASCRIPTION

"Two individuals can each either cooperate or defect. The payoff to a player affects its reproductive success. No matter what the other does, the selfish choice of defection yields a higher payoff than cooperation. But if both defect, both do worse than if both had cooperated." (Axelrod, 1987, p.2). This is the prediction of the optimal expected solution to the so-called "Prisoner's Dilemma" (Skyrms, 2004), a version of which was used to tell the story of Laurence and Faisal.

Haroush and Williams's (2015) investigation concerns the neural underpinnings of social interactions. This, as we will see, is a very different investigation with respect to that concerning the neural underpinnings of observable goal ascription. We know that some outcomes can be motorically represented by certain goal states, but some other

outcomes cannot be just motorically represented. I am now concerned with the latter.

Harousch and Williams claim that social interactions of the sort depicted in the case of the Prisoner's Dilemma (at least in the variation of the Dilemma that they are testing) involve awareness of one's own decision and of which outcomes one's own decision will cause. But they also involve awareness of the causal function of the other's decision and of which outcomes the other's decision will cause. A Prisoner's Dilemma-like scenario occurs when two or more agents are capable of anticipating the other's unobservable future actions, i.e. those actions that are caused by their distal intentions. Evidence shows that the neural underpinnings of this capacity have been located in the dACC, dorsal anterior cingulate cortex, of the rhesus monkey's brain (*Macaca mulatta*). Cingulate neurons get activated when the macaque foresees its own and others' unobservable future goal, i.e. and therefore distal intentions, insofar as they cause such goal (as opposed to mirror neurons that get activated when the macaque performs or observes others performing the very same observable action).

This is a description of the experimental set up: two rhesus monkeys—call them Laurence and Faisal—are facing each other, sitting in front of a computer screen. They hold joysticks with which they have to choose over two symbols on the screen. One symbol (a blue triangle) will deliver more apple juice to oneself, and the other symbol (a red hexagon) will lead to sharing the apple juice evenly. Their respective choice is hidden to the other prior to selection. *Only* after both Laurence and Faisal have made their choice, their respective selection becomes visible to the other. So, Laurence makes his selection, and Faisal makes his selection. There are three possible results: (1) if one of the two—say, Laurence—deceives and the other cooperates, Laurence gets six bricks of apple juice, and Faisal only gets one brick of apple juice. (2) If both go for the symbol that only delivers juice to themselves, neither of them gets much apple juice, i.e. two bricks each. (3) If both go for the symbol that delivers juice to both, they each receive four bricks of apple juice.

In this study, Harousch and Williams (2015) show that, over time, monkeys learn to favour the equal sharing, or what is known as the Pareto optimality condition (Nash, 1950). The prediction of this theory is that two or more agents get the highest payoff if and only if they both cooperate. One way of reading these results is to claim that the success in the game depends on the participant's ability to anticipate the distal intentions of the other because one's own expected reward can only be predicted in relation to the choice made by the other: arguably, the success of the game depends on the pursuit of

joint distal intentions. Harousch and Williams (2015) registered the neural activation of the dorsal anterior cingulate cortex under such circumstances, and showed how individual cingulate neurons represent – what we may call – another's unobservable goal, i.e. a distal intention. In 79% of the trials, the activity of these neurons, when one monkey made the decision, matched the activation of the same neurons for the same decision in the other monkey. Further confirmations were obtained courtesy of a wide range of control tasks—for example, by putting the same monkeys to play against a computer rather than another monkey[5]. In this scenario, the results were very different: monkeys were significantly less willing to cooperate. This demonstrates the existence of a causal link between cingulate activity and the specific enactment of mutually beneficial decisions. The prediction is based on memory of past experience, that is, on a process of familiarization with the choice made by the others over time. As mentioned earlier, these results nicely fit the role ascribed to the Episodic Memory System in explaining individual action planning. The episodic memory system includes the posterior cingulate cortex (PCC). The social interaction system includes the dorsal anterior cingulate cortex, and, possibly, the posterior cingulate cortex as well.

What is now known is that the successful recollection of experienced events is associated with the activity of cortical regions (ventral posterior parietal cortex, and medial prefrontal cortex) that include retrosplenial/posterior cingulate cortex. The angular gyrus activates at the time of retrieval.

It is worth pointing out that I am by no means suggesting that humans or other animals are natural born rational maximisers of expected utility, but rather that it is a fact of life that in many social interactions nobody can have it all, and that we are more likely to get the most by giving up a little. Becoming aware of this requires experience, or — to say it with the relevant experimental jargon — familiarization: these results emerged after a preliminary familiarization with the task. The experimenters registered the neural activation of the dorsal anterior cingulate cortex of the macaque, under the circumstances that I just described, and they showed how individual neurons implement the representation of another's unknown distal intentions.

Harousch and Williams focused on the activity of the dorsal region of the anterior cingulate cortex (dACC). dACC neurons encode the monkey's decision to defect or cooperate. The highest activity is registered both during self-cooperation and other-cooperation. The increasing rate of correct predictions gets encoded by the dACC

[5] Interestingly, similar results were obtained in a study on human's brain neural responsiveness to the interaction with an actual partner, vs. an avatar, vs. an inanimate object (see, Costantini et al., 2011).

neurons and this explains the capacity to anticipate the decision of the other, and the ability to, eventually, settle on a stable cooperation.

I have explained that the capacity to ascribe joint distal intentions is served by a neural circuit different from that exploited during goal ascription inferred from observable actions. This is an additional reason to appreciate the difference between motor representations and distal intentions, and to urge an empirically informed account of the latter. A theory of joint distal intentions in nonhuman animals is needed in order to understand nonhuman animals' individual and collective planning skills. Many instances of complex group behaviour flourished from the literature in cognitive ethology, and the findings of Harousch and Williams can likewise provide and gain benefit from ethological findings. I do not have enough room for an extensive overview of the relevant literature from cognitive ethology, so now I will only review some data on the unique patterns of the group behaviour of the *Macaca fuscata*, best known as "snow monkeys"—for reasons that will shortly become clear. These animals and the *Macaca mulatta*, the latter being the subjects of the juice brick experiment, are closely related species of the *Macaca fascicularis* species group (Chu, Lin, & Wu, 2007). Therefore, neurocognitive results on rhesus monkeys should be consistently applied to inform ecological results on *Macaca fuscata*.

5. EVIDENCE FROM THE FIELD

I have analysed experimental data for the claim that nonhuman animals exploit their capacity for ascribing joint distal intentions to others when it comes to obtain a higher payoff over time. The animal kingdom is full of instances of sophisticated group activities, but I will use one case study in particular: snowball play-fight by Japanese rhesus monkey (*Macaca fuscata*). This is one of the few instances of joint action where the play-fight is the only overt purpose of the interaction. There is no reason to invoke an explanation that appeals to motivations such as competitive drive or dominance display. As such, we should, unambiguously, talk about the appreciation of joint distal intentions on the part of the macaques. The specific conditions in which this behaviour occur are not triggered by any vital needs; they are spontaneous and the playing scenario is most likely a genuine joint venture.

Japanese macaques (*Macaca fuscata*) display several instances of purposeful behaviour (Huffman & Quiatt, 1986). Among these activities, researchers acknowledge

stone handling of various types, most notably washing sand off sweet potatoes in water and passing this behaviour on to others (Leca et al., 2007c, 2010b). These actions are classified as pre-cultural, for they are regularly pursued and spread among the majority of the individuals within a troop (Kawai, 1965). Call these activities *SH* (stone handling) *patterns* (Leca et al., 2012). As far as individual planning is concerned, the most notable example is stone hurling (Osvath, 2009). On top of this, Japanese macaques living in Nagano have been observed engaging not only in individual SH patterns, but also in joint ones(for another joint case, see also van Hooff & Lukkenaar, 2015)

As mentioned, Japanese macaques are commonly known by the nickname of "snow monkeys". Both in wild and feral groups, they have been observed to construct snowballs and engage in play-fights (Eaton, 1972; Huffman & Quiatt, 1986).

This is one of the few instances of joint action where the play is the only overt purpose of the interaction. It takes two or more individuals to make snowball play-fight meaningful. You cannot play on your own: you need a target to hit and you need a shooter to hide from. So, as said, there is no reason to invoke an explanation that appeals to a motivational force like competitive drive, or dominance display, as argued by the defenders of the Lean Account. Sarah Hankerson comments on the crucial role that snowball play-fighting practices have in revealing the nature of nonhuman animals' social interactions: "Non-human primates can tell us a lot about the basic structure of behaviour in group settings. We can look at the rudimentary way individuals handle conflict and affiliation. Being highly social animals, Japanese macaques can serve as models of group dynamics. This study looks at play behaviour, which may seem a non-functional activity, but infants (both human and non-human) develop skills, improve physical strength and dexterity, and learn a lot about the world around them and their place in it by engaging in play behaviour" (Engebretson, 2013).

SH patterns, among which we can include play-fighting, have been classified as instances of planned activity (Osvath, 2009a). On a very broad reading, planned activities are time-structured intentional actions, which can be further divided into sub-phases or sub-plans (Bratman, 2014). The *Macaca fuscata's* practice respects the following threefold structure: (1) selection, (2) manufacturing, (3) use (Osvath & Karvonen, 2012). Macaques have to (1) select the right kind of snow with the right texture and compactness, in order to (2) manufacture the balls by rolling the snow, and lastly, (3) throw the snowball to the play-fighting fellows.

As noted by Leca and colleagues (2012), even if SH patterns are primarily

instantiated in the form of individual activities, their social character should not be overlooked. Firstly, they say, because SH patterns are socially transmitted; secondly—and crucially for our interest in joint distal intentions—because these comparative studies showed that group size correlates with the number of group members engaging in SH patters simultaneously; and thirdly, because SH are producing social interactions. As we argued, the findings on cingulate neurons may illuminate the nature of play-fight, and other studies on joint activities. In the past researchers studying the behaviour of non-human animals, especially primates, shown that they can perform actions that are guided by certain mental states (Tomasello and Call, 1997; Tomasello, 1999; Tomasello et al., 2012). This paper argues that these mental states are distal intentions: representations of goals to which plans are directed. In the past, this brought into question another issue: do non-human primates understand the distal intentions of others (Povinelli, 1986; Call & Tomasello, 1998, p. 192)? The causalistic explanation had it that non-human primates can predict others' behaviour on the basis of their past experience. When an event occurs repeatedly, they can recall they had a similar experience in the past and know how to behave accordingly (Tomasello & Call, 1997). The other line of analysis emphasizes the role that action prediction can play in the understanding of the intentions of others (Premack & Woodruff, 1978). A number of studies (Hare et al., 2000; Hare et al., 2001; Suddendorf & Withen, 2001; Call et al., 2004) led to the common conclusion that non-human primates have some knowledge about others' psychological states. But, still, this knowledge is determined by the direct perception of others' actions. This is also the case for a recent experiment on strategic future planning, which for the first time investigates whether chimpanzees can actively hide objects from others (Karg et al., 2015).

What the experiments of Harousch and Williams show is that monkeys interpret others' behaviour, they do not just perceive it. But, before these findings, it was commonly thought that nonhuman animals' capacity for goal ascription was limited to the directly perceivable behaviour of others, i.e. to their observable goals, and not to their non-observable goals, i.e. those than can be inferred through the ascription of distal intentions.

6. CONCLUSION

We have analysed the first instance of empirical evidence for what has been hypothesised to be neural underpinnings of predictive behaviour that guides animal social interactions (Frith and Frith, 1999; Rilling et al., 2004; Sanfey et al., 2006; Vogeley et al., 2001).

This paper has attempted to shed light on how the ascription of distal intentions to others, that is, the ascription of future goals on the basis of unobservable cues, is made possible by a specific neural circuit, different from that exploited during goal ascription inferred from observable cues, that are primarily perceptual or motor. Nonhuman animals, in particular macaques, can ascribe joint distal intentions to others, and act jointly on the basis of this appreciation of the context of social interactions.

On this issue, the debate is far from settled, and the investigation is much in its infancy. At the same time I tried to show that it is a research ground worth to be explored for its promising explanatory potential.

REFERENCES

Atance, C. M., & O'Neill, D. K. (2001). "Episodic Future Thinking", *Trends in Cognitive Sciences*, 5(12), pp. 533-539.

Axelrod, R. (1984). *The Evolution of Cooperation*, Basic Books, New York.

Axelrod, R. (1987). "The Evolution of Strategies in the Iterated Prisoner's Dilemma", in *Genetic algorithms and simulated annealing*, L. Davis, New York.

Ban, S.D., Boesch, C., & Janmaat, K. R. L., (2014), "Tai chimpanzees anticipate revisiting high-valued fruit trees from further distances", *Animal Cognition*, 17(6), pp. 1353-1364.

Boesch, C., & Boesch - Achermann, H., (2000). *The Chimpanzees of the Tai Forest*, Oxford University Press, Oxford.

Bratman, E. M. (1984). "Two Faces of Intention", *The Philosophical Review*, 93(3), pp.375-405.

Bratman, E. M. (2014). *Shared Agency: A Planning Theory of Acting Together*, Oxford University Press.

Butterfill, S., & Sinigaglia, C. (2014). "Intention and motor representation in purposive action", *Philosophy and Phenomenological Research*, 88(1), pp. 119-145.

Call, J., & Tomasello, M. (1998). "Distinguishing Intentional from Accidental Actions in Orangutans (Pongo pygmaeus), Chimpanzees (Pan troglodytes), Human Children (Homo sapiens)", *Journal of Comparative Psychology*, 112(2), pp. 192-206.

Call, J. Hare, B., Carpenter, M., & Tomasello, M., (2004), " 'Unwilling' versus 'unable': chimpanzees'understanding of human intentional action", *Developmental Science*, 7(4), pp. 488- 498.

Chater, N., & Manning, C. D. (2006). "Probabilistic models of language processing and acquisition", *Trends in Cognitive Sciences*, 10 (7), pp. 335-344.

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). " Ready Both to Your and to My Hands: Mapping the Action Space of Others", *PloS ONE*, 6 (4).

Chu, J. H, Lin, Y. S., & Wu, H. Y. (2006), "Evolution and dispersal of three closely related macaque species, Macaca mulatta, M. cyclopis, and M. fuscata, in eastern Asia", *Mol. Phylogenet. Evol.,* 43(2), pp. 418-529.

Clayton, N. S., Salwiczek, L. H., Dickinson, A, (2007), "Episodic Memory", *Current Biology*, 17, pp. 189- 190.

Eaton, G. (1972)."Snowball construction by feral troop of Japanese macaques (Macaca fuscata) living under seminatural conditions", *Primates*, 13(4), pp. 411-414.

Frith, C.D., and Frith, U. (1999). "Interacting minds—a biological basis", *Science*, 286, 1692–1695.

Gallese, V., & Goldman, A. (1998). "Mirror Neurons and the Simulation Theory of Mind- reading", *Trends in Cognitive Sciences*, 2 (12), pp. 493-501.

Goldberg, E., & Bougakov, D., 2007, "Goals, Executive Control, and Action", (Eds.) B. J. Baars & N. M. Gage, *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, London, Academic Press, pp. 339-420.

Engebretson, K. (2013). "Student Study finds Snow Monkeys Just Wanna Have Fun",

http://www.stthomas.edu/news/student-study-finds-snow-monkeys-just-wanna-have-fun/.

Hare, B., Call, J., Agnetta, B., & Tomasello, M., (2000), "Chimpanzees know what conspecifics do and do not see", Animal Behavior, 59, pp. 771-785.

Hare, B., Call, J., & Tomasello, M. (2001). "Do chimpanzees know what concpecifics know and do not know?", *Animal Behavior*, 61, pp. 139-151.

Hare, B., & Tomasello, M. (2004). "Chimpanzees are more skilful in competitive than in cooperative cognitive tasks", *Animal Behaviour*, 68, pp. 571-581.

Haroush K., & Williams, Z. M., (2015). "Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates", *Cell*.

Hobaiter, C., Poisot, T., Zuberbuhler, K., Hoppitt, W., & Gruber, T. (2014). "Social Network Analysis Shows Direct Evidence for Social Transmission of Tool Use in Wild Chimpanzees", *PLoS Biology*.

Huffman, M. A., & Quiatt, D. (1986). "Stone handling by japanese macaques (Macaca fuscata): Implications for tool use of stones", *Primates*, 27, pp. 413-423.

Jeannerod, M. (1994a). "The representing brain: neural correlates of motor intention and imagery", *Behavioral and Brain Sciences*, 17, pp. 187–246.

Jeannerod, M., (1995), "Mental Imagery in the Motor Cortex", *Neuropsychologia*, 33 (11), pp. 1419-1432.

Jeannerod, M. (1997). *The Cognitive Neuroscience of Action*, Oxford: Blackwell.

Karg, K., Schmelz, M., Call, J., Tomasello, M. (2015). "Chimpanzees strategically manipulate what others can see", *Animal Cognition*, 18(5), pp. 1069-76.

Kawai, M. (1965). "Newly acquird pre-cultural behaviour of a natural troop of Japanese monkeys on Koshima Island", *Primates*, 6, pp. 1-30.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). "The mirror-neuron system: a Bayesian Perspective", *NeuroReport*, 18 (6), pp. 619-623.

Leca, J.-B., Gunst, N., & Huffman, M. A. (2007c). "Japanese macaque cultures:Inter- and intra-troop behavioural variability of stone handling patterns across 10 troops", *Behaviour*, 144, pp. 251- 281.

Leca, J.-B., Gunst, N., & Huffman, M. A. (2010b). "Indirect social influence in the maintainance of the stone handling tradition in Japanese macaques (Macaca fuscata)", *Animal Behaviour*, 79, pp. 117-126.

Leca, J.-B., Gunst, N., & Huffman, M. A., (2012). "Thirty years of stone handling tradition in Arashiyama macaques: implications for cumulative culture and tool use in non-human primates", *The Monkeys of Stormy Mountain: 60 years of Primatological Research on the Japanese Maaques of Arashiyama*, (Eds., J.B. Leca, M.A. Huffmann & P.L. Vasey), Cambridge Press University.

Mele, A. R. (1990). "Exciting Intentions", *Philosophical Studies*, 59(3), pp. 289-312.

Mullay, S. L., & Maguire, E. A., (2014), "Memory, Imagination, and Predicting the future: A Common Brain Mechanism?", *The Neuroscientist*, 20(3), pp. 220-234.

Nash, J.F. (1950). "Equilibrium points in N-person games", *Proc. Natl. Acad. Sci. USA*, 36, 48–49.

Osvath, M. (2009). "Spontaneous planning for future stone throwing by male chimpanzee", *Current Biology*, 19(5), pp. 190-191.

Osvath, M., & Karvonen, E., (2012), "Spontaneous Innovation for Future Deception in a Male Chimpanzee", *PloS ONE*, 7(5), pp. 1-8.

Pacherie, E., (2000), "The Content of Intentions", *Mind & Language*, 15 (4), pp. 400-432.

Premack, D., Woodruff, G., (1978), " Does the chimpanzee have a theory of mind?", *Behavioural and Brain Sciences*, 4(4), pp. 515-629.

Raby, C. R., & Clayton, N. (2012). "Episodic Memory and Planning", in (Eds., T. K. Shackelford and J. Vonk) *The Oxford Handbook of Comparative Evolutionary Psychology*, Oxford Handbooks Online.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). "A neural basis for social cooperation", *Neuron*, 35, 395–405.

Rizzolatti, G., Fadiga, L., Matelli, M., & Bettinardi, V. (1996). "Localization of grasp representations in humans by PET", *Exp. Brain Res.,* 111, pp. 246-252.

Rizzolatti, G., & Arbib, M. A. (1998). "Language within our grasp", *Trends Neurosci.,* 21, pp. 188-194.

Rizzolatti, G., Fabbri-Destro, M., & Cattaneo, L., (2009), "Mirror neurons and their clinical relevance", *Nature Clinical Practice Neurology*, 5, pp. 24-34.

Rizzolatti, G. & Sinigaglia, C. (2010). "The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations", *Nature reviews. Neuroscience*, 11(4), pp. 264-274.

Rosenbaum, R. S., Kohler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., & Black, S. E., (2005), "The case of KC: Contributions of a memory-impaired person to memory theory", *Neuropsychologia*, 43(7), pp. 989-1021.

Sanfey, A.G., Loewenstein, G., McClure, S.M., and Cohen, J.D. (2006). "Neuro-economics: cross- currents in research on decision-making", *Trends Cogn. Sci.* 10, pp. 108–116.

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., Bush, G., & Eskandar, E. N., (2012), "Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation", *Nature*, 488(7410), pp. 218-21.

Sinigaglia, C., & Butterfill, S., (2015), "A puzzle about the relations between thought, experience, and the motoric", *Synthese*, 192 (6), pp. 1923-1936.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*, Cambridge University Press.

Suddendorf, T., & Whiten, A. (2001). "Mental evolution and development: evidence for secondary representation in children, great apes, and other animals", *Psychol. Bull.,* 127(5), pp. 629-650.

Suzuki, S., Adachi, R., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2015). "Neural Mechanisms Underlying Human Consensus Decision-Making*", Neuron*, 86, pp. 1-12.

Tomasello, M, & Call, J. (1997). *Primate cognition*, Oxford University Press.

Tomasello, M., Melis, A., Tennie, C., Wyman, E., Herrmann, E. (2012). "Two key steps in the evolution of human cooperation: the Interdependency Hypothesis", *Current Anthropology*.

Tomasello, M.. (2014). *A Natural History of Human Thinking*, Harvard University Press.

Tulving, E.. (1972). "Episodic and Semantic Memory", in (Eds., E. Tulving, and W. Donaldson), *Organization of Memory*, Academic Press, pp. 381-402.

Tulving E. (1993). "What is episodic memory?", *Current Directions in Psychological Science*, 2(67).

Van Hooff, J. R. A. M., & Lukkenaar, B. (2015) "Captive chimpanzee takes down a drone: tool use toward a flying object", *Primates*, online first: 10.1007/s10329-015-0482-2.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe , F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., and Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self- perspective. *Neuroimage,* 14, pp. 170–181.