Original Paper UDC 159.9.01:165.242.2
Received January 5th, 2014

Massimiliano L. Cappuccio
UAE University, Cognitive Science Program, Room 0018A, E5 Building, Al Ain, United Arab Emirates
m.lorenzo@uaeu.ac.ae

Inference or Familiarity?

The Embodied Roots of Social Cognition*

Abstract

I consider two distinct deflationary theories in social cognition that aim to explain action understanding without demanding meta-representational or mindreading processes: the first one is the 'teleological stance hypothesis' (TSH), claiming that we infer the intended goal of a certain observed action based on the mere perception of its effects and of its situational constraints; I decided to dub the second one 'the embodied familiarity hypothesis' (EFH) to comprise all the theories claiming that we recognize the intended goal of a certain action based on the perceptual or motoric expertise developed within the sensorimotor contingencies associated to that action's context. TSH's main requirement is that the observer could ascribe efficiency, and therefore rationality, to the observed agent's movement, while EFH's main requirement is that the observer were somehow exposed to the perceptual or motoric details of the observed agent's action. I argue that EFH describes a more primitive and fundamental form of action understanding, i.e. one that is necessarily presupposed by TSH: in fact, while recognizing efficiency is neither a necessary nor a sufficient condition for detecting goal-relatedness, some kind of perceptual or motoric familiarity with the details of the observed action's context is always necessary for any ascription of efficiency, and therefore of rationality, to the observed agent. I conclude that, while TSH might certainly be effective in describing certain rational forms of action understanding, it implicitly requires EFH to be true, as its inferential system would be groundless without an assumed familiar background of embodied expertise.

Keywords

social cognition, teleological stance, mirror neurons, rationality, intentional stance, direct matching hypothesis

Introduction: two deflationary approaches to social cognition

Social cognition is the field of interdisciplinary research that studies how we comprehend the others. It is interested, among various things, in how we can understand[1] the goal of an observed action in an effortless, automatic and

* This paper is in debt with the philosophical conversations I had with Corrado Sinigaglia at the University of Milan and with Dan Hutto during my period of research visit at the University of Hertfordshire. I am very thankful to Chiara Brozzo for reading an early version of this paper and for commenting on it. The responsibility of any possible mistake is just mine.

[1] After cautious meditation, I will intentionally ignore the warning, presented and reasonably argued for by various scholars (e.g., Hutto 2013), against any similar use of the concept of 'understanding': even if it were true that in the philosophical literature this word predominantly indicates intellectual comprehension, the expression 'understanding an action' actually has a much broader and less

largely reliable manner, even if we know nothing of what the agent is thinking and very little of her personal background. As opposed to many models of 'mindreading' that have been advanced during the last decades (Baron-Cohen 1994, Goldman 2006), and that demand meta-cognitive representations of the agent's mind, two deflationist theories seem particularly promising today, as they attempt to provide a cognitively parsimonious account of the processes underlying action understanding: both of them assume that in our direct interaction with the others we do not need any intellectual representation of their mental states (i.e. propositional contents such as beliefs, intentions, and desires), because the key information necessary to recognize the action goal is perceptually accessible through observation of the agent's bodily movements (Gallese et al. 2009). Both human infants (Astington & Gopnik 1991; Gopnik, 1993; Wellman 1991) and non-human primates (Call and Tomasello, 2008) manifestly understand the others' actions goals without having propositional or mentalistic representations of the observed agents' states at their disposal, and possibly without being capable of distinguishing their own mental states from the observed agents' (Gallagher and Povinelli 2012). The two aforementioned deflationist accounts attempt to offer different explanations of how low-profile action understanding is available even to subjects that cannot rely on language or "mindreading".

The first account claims that in order to make sense of the other's action we need to infer its goal based on the expected effects of her action and what we see of her current contextual circumstances: this view assumes that the observer's inference is reliable insofar as she realizes that the observed agent knows both what she wants to achieve and what the most efficient means to achieve it are. Gergely and Csibra (2003) refer this to as the 'teleological stance hypothesis', or TSH. The second account claims that our capability to understand the other's action goal relies fundamentally on our own embodied familiarity with the observed action, including the sensorimotor contingencies in which it is situated. I will generically refer to this view, which assumes that we can make sense of the intention underlying another's movements on the basis of our direct acquaintance with them, as 'embodied familiarity hypothesis' (EFH). In this paper I will intentionally characterize this view in a generic way, because – if broadly construed – EFH encompasses various embodied approaches to social cognition, and they do not always entirely agree on the kind of perceptual and motoric requirements that are necessary to recognize movements as goal-oriented (Gallagher 2008). The approaches encompassed by EFH include for example the interactionist and enactivist approaches and the motor cognition hypothesis. On the one side of this approach, interaction theory (Gallagher 2001; Gallagher & Hutto 2008; Gallagher & Zahavi 2008; De Jaegher et al. 2010; Froese and Gallagher 2012) claims that we do not infer goals beyond the others' actions, but we directly perceive them in the others' movements, and that this perceptual function develops with our capability to proficiently adjust our conduct to another's behavior (the goal of the action is perceptually associated to its immediate effects, and these effects gain a perspicuous perceptual meaning during live experience by reason of the history of the unprincipled embodied engagement between interacting agents). On the other side of EFH, the 'motor cognition hypothesis' (Gallese et al. 2009) asserts that it is our capability to perform a bodily action to achieve a certain motor goal that allows us to recognize the movements aiming at the same motor goal, when we see them performed by someone else. Note that both the 'direct-matching hypothesis' (Rizzolatti et al. 2001; Rizzolatti and Sinigaglia 2010) and the 'embodied simulation theory' (Gallese 2005, 2010) offer more specific versions of the motor cognition hypothesis, which we will not consider here. Therefore, according to EFH, goal recognition could either rely on the familiarity that the agent has developed in responding to the observed movements, or on the analogy between the movements observed and the movements that the agent would have produced in similar contexts: the observer maps the observed action in terms of goal-oriented movements that are either complementary or analogous to the ones observed, but in either case – unlike for TSH – it is the involvement of his embodied dispositions (and hence his direct sensorimotor experience) that allows him to make sense of the observed action's goal, not calculative skills or detached capabilities of reasoning and intellectual judgment.

Both TSH and EFH reflect the essence of traditional philosophical doctrines, which today receive extensive empirical support from experimental cognitive psychology: the first of them has been corroborated by a series of psychological experiments on the infants' attention based on a looking-time paradigm (Gergely et al. 1995; Gergely and Csibra 2003); the second received credit from various studies in the field of motor cognition (e.g., the discovery of the so called "mirror neurons", Di Pellegrino et al. 1992), which suggested how embodied skills play a defining role in action understanding.

As both models seem conceptually consistent and grounded in experimental evidence, each of them could convincingly account for at least some instances of action understanding. I do not mean to deny that they could be mutually incompatible, as different cognitive strategies of action understanding probably co-exist to allow us make sense of goals in terms of either practical reasons to act, as argued by THS, or sensorimotor opportunities, as maintained by EFH. However, even if these models were compatible or complementary as alleged, we would still have to determine if one of these two theories could satisfactorily explain the most fundamental and essential processes that underpin action understanding, and if it could do it without assuming the other theory. The question that the two theories have to address, but that possibly not both of them can answer in an equally satisfactory way, is: what is goal attribution in its most primitive and essential form, and what are the minimal cognitive preconditions to have action recognition? While TSH maintains that the capability to infer practical reasons is the most fundamental requirement to characterize an action goal, under the presupposition that the capability to perform a certain action is not required to infer its goal, EFH argues that perceptual habits and motor skills are more fundamental, as it assumes that the capability to rationalize the abstract intention behind another's action is not indispensable to recognize the immediate purpose that shapes her movements. If we agree that rational inference and sensorimotor skills are two distinct and independent sets of capabilities that (at least in principle) do not presuppose one another, it seems that we have to choose between their respective underlying principles: do the roots of action understanding arise from a universal principle of rational inference or rather from embodied dispositions?

To address this question, this paper will focus on the critical discussion of one of the theoretical preconditions of TSH: the necessary attribution of rationality to the observed agent. This critical discussion will not prove this precondition wrong a priori, but will argue that the attribution of rationality is not a

intellectualist use in common language. Considering how difficult it is to do without such expression even when referring to non-human animals and non-linguistic infants, I have the feeling that artificially adopting an extensionally equivalent synonym just to avoid demanding cognitive pre-requisites would be just a pointless lexical exercise.

primitive capability of the infant's mind, as it rests on other contingent, and more primitive, preconditions – of an embodied (not inferential) kind. I believe that this conclusion does not rule out TSH, but suggests the primacy of EFH over TSH, as it leads to the idea that bodily situated enactive capabilities are from the beginning involved in action recognition and goal understanding, and play an active and constitutive role in them.[2]

1. The teleological stance hypothesis

This paper will deny neither that TSH builds on solid experimental data nor that it holds a certain degree of explanatory power. Various bodies of evidence suggest that 12-month old infants understand the goal of an agent's actions employing minimal perceptual cues about the action and its context (Csibra et al. 2003; Gergely et al. 2005). This early capability of understanding precedes folk-psychological knowledge because the infant does not master any such mind reading capability yet, and the observed movements are not sufficient to attribute propositional contents to an agent's state of mind. So how does early goal attribution work at such age? What we know is that generic visual features such as self-propelled movement (Premack 1990), autonomous non-rigid transformations of the moving body's surface (Gibson, Owsley, & Johnston 1978), irregular path of movement (Mandler 1992), and motion caused by spatially and temporally distant events are typically sufficient to indicate an internal source of energy, hence suggesting animacy and goal-directedness. However, these features are insufficient to determine the particular goal behind one's action (Anscombe 1957). TSH aspires to complete the picture, postulating that specific goals can only be attributed to those actions that: 1. display an equifinal structure through time (see Heider 1958: in spite of varying environmental conditions, the patterns of action chosen by the agent consistently result in the same end-state; this suggests, via induction, that the agent's behavior is consistent through time); 2. elicit an attribution of rationality to an agent (see Dennett 1987: the agent's behavior suggests that she has selected the most efficient means to reach a desired goal-state).

Because of these two pre-requisites (i.e. reliability of inductive reasoning and attribution of rationality to the agent), TSH is comparable to the 'intentional stance hypothesis' (or ISH, at first presented by Dennett 1987 and then endorsed by Gergely et al. 1995), and represents its deflationary development (Gergely and Csibra 2003). ISH asserts that goal attribution takes place as a folk-psychological process that presupposes the systemic integration of three kinds of mental representations: the agent's beliefs (e.g., "Paul thinks that there is a banana in the fridge"), desires ("Paul wants a banana"), intentions ("Paul means to grab a banana"). Combined with sufficient perceptual cues ("Paul is reaching for the fridge"), the knowledge of any two of these representations is sufficient to infer the third one, as the known representations work as the premises of an inductive reasoning that has the unknown representation as its conclusion. Since, in this picture, the intention is sufficient to characterize the action's proximal goal, knowing the agent's desires and beliefs can do the job of goal attribution. Beliefs, intentions, and desires are treated as representations because they convey declarative contents, hence the inference drawn from them is mentalistic in character: ISH assumes that the observer must manipulate meta-representations (possibly propositional in format) of the agent's mind in order to guess her reasons and her future behavior. While it is debated whether or not ISH could account for goal attribution in pre-linguistic children and non-human animals (Gergely et al 1995; Gergely and Csibra 2003; Wellman 2002), due to the limited meta-representational capacities of these agents, TSH explicitly claims to be able to explain action understanding in absence of such capacities, as it only requires manipulation of perceptual cues, not meta-representations of the agent's mental states. TSH assumes that, in order to build up such systematic inferences, the observer only needs to consider – instead of the desires [A], the intentions [B], and the beliefs [C] – the physical goal-state at the end of the action movement [A], the observable bodily movements chosen to achieve it [B], and the situational constraints that objectively bound the choice of the movements [C], respectively (Gergely and Csibra 2003). Because all of these elements are perceptually recognizable, and are prerequisites of every embodied action, by simply registering how the agent moves (e.g., in a straight line or a curve), and how the physical surroundings constrain or direct its movements (through obstacles or preferential paths), even an observer unfamiliar with that kind of action can determine to what goal-state the agent's movement is directed. The assumption behind this model of goal attribution via triangulation of the action's end-state is that the agent's movements must display the mark of a rational deliberation.

Note that, in contrast with more intellectually demanding notions of rational action (e.g., the theory of agency as "planning", as in Bratman 2007), the reasons invoked by TSH are (meant to be) of a minimalist kind: in fact, they simply acknowledge that the means used to achieve a certain goal state were not casual; the agent had practical reasons to pursue that goal with those means, reasons immediately related to reaching that particular goal In an efficient way. Exactly like the teleological stance, also the intentional stance assumes that the observer attributes a capability of deliberation to the agent based on a *principle of rational action*, that is the conjecture of the agent's systematic predisposition to achieve her goal with the most economical means available to her in the given circumstances. In the paradigmatic case of spatial navigation, this means reaching a certain destination through the shortest path. Therefore, like ISH claims that we infer the intention of an action through an inductive reasoning based on the knowledge of the underlying beliefs and desire, TSH assumes that we infer the final direction of a certain movement based on the observation of the movements themselves in relation to the specificities of the environment; therefore, the main difference is that while ISH characterizes goals in terms of intentions (which convey propositional contents), the TSH characterizes them as predictable physical end-states of a movement (which are perceptible states of affaires).

The proponents of TSH have preponderantly focused, with their research, on stimulus properties that are tied only to abstract patterns of the most generic kind of motion, i.e. spatial navigation (Csibra 2003; Csibra and Gergely 1999; Csibra et al. 1999; Csibra et al. 2003; Gergely 2002; Gergely and Csibra 1994, 1997, 2003; Gergely et. 1995). In fact, the data supporting TSH have been collected through the measurement of the looking time of 12-month old babies habituated to computer-animated goal-oriented events performed by non-bio-

2
In this occasion I will not review TSH's main objections against this view (for example, the objection that the capability to comprehend various types of hand actions is commonly exerted without any previous familiarity with those actions): one reason is that I have already addressed these objections somewhere else (Cappuccio 2012); another is that I believe the objections raised here against the primacy of TSH, if proven correct, would already be sufficient to confirm the more fundamental role of EFH.

logical agents (geometrical drawings in motion, such as animated bouncing spheres), according to a violation of expectations paradigm (I refer in particular to Csibra et al. 2003; Gergely and Csibra 2003; Gergely et al. 2005). For example, in one of the conditions a small circle (i.e. the *agent*) approached and then stopped near a large circle (the *end-state*) after "jumping *over*" (the *means act*) a solid obstacle that separates them (*situational constraint*). TSH claims that, in similar scenarios involving the same elements, the child can infer that the goal of the small circle's movement is again to approach the large circle and then stop near it by integrating the available visual cues according to some rational principles. In fact, when the obstacle was removed, the infant's reactions were consistent with the rational principle of parsimony of movement: after removing the situational constraint, the infants had to watch two test displays in which the small circle performed either the same jumping movement as before or a perceptually novel straight-line movement; looking time was significantly longer in the first condition, as the child was surprised to see that the circle "decided" to "jump" even if there was no obstacle that could justify a curved trajectory. According to the experimenters, we can legitimately conclude that the infant expects the small circle to "behave rationally", if we keep in mind that the infant: (1) recognized the action as goal-directed (aiming at the same ending state as the previous condition); (2) inferred the most efficient trajectory (curved, in presence of an obstacle, or straight, without any obstacle) under the new situational constraints; (3) expected the agent to use this trajectory to reach its goal. Alternatively, any two other goal-oriented events (in which the means act is fixed but either the end-state or the situational constraint vary across different conditions) can provide complementary evidence of the integration of the these three types of visual cues by means of inference: in fact, the surprise of the infant indicates that he has inferred the goal of the action when it was non-evident, if the changing trajectories of the agent's movement seem not otherwise explainable in the given environmental contingencies; or that he has inferred the presence of hidden obstacles, if the current trajectory of the agent seems deviate from its original pattern in ways that would not be otherwise compatible to the expected end-state. In all of these cases, the children always look longer to the conditions in which the agent's behavior does not appear as efficient as expected, as they evidently are surprised that the agent is not executing the most rational movements: this proves that the infants have already appropriately inferred the missing element of the triad, and have understood how they were meant to be systematically integrated. Indeed, on the basis of TSH, any inference drawn on the efficiency of the agent "immediately translates in a judgment on her rationality" (Csibra 2003), hence allowing an inductive reasoning on the goal-oriented nature of the action and the missing element of the triad.

While the methodology used by these experiments is solid, and the present discussion does not aim to question their results, both the extension that is given to the meaning of such results and the assumed way of interpreting their preconditions are somehow conceptually puzzling: the problem is not only that this approach has a limited capability to account for other empirical data (as I will argue later, for referring example to Rochat et al. 2008); it is also that this approach does not capture (and, to some extent, is not consistent with) the inherently embodied dimension of the phenomenology of goal-oriented action in real life. In fact, according to TSH, when infants take the teleological stance to interpret actions as means to goals, they are supposed the evaluate the "relative efficiency of means" by applying the principle of rational action, and generate systematic inferences to identify relevant aspects of the situation that justify the action as an efficient means even when these aspects are not directly visible to them; however, while this hypothesis is in principle sound, the empirical actualities make it factually improbable that the recognition of efficiency is sufficient or even just necessary to attribute a goal to an action. In particular, two assumptions taken for granted by TSH deserve to be critically examined: the first assumption is that non-mentalistic goal attribution is universally based on the calculus of the optimal means (so that efficiency is a necessary condition of goal attribution); the second is that the teleological stance is possible without any other, more fundamental, forms of goal attribution, i.e. without any preconditions or precursors (so that efficiency is a sufficient condition of goal attribution).

There are reasons to question the first assumption: after all, some extremely inefficient behavior leaves no doubts about the goal that motivates them (e.g., humans attempting to grasp a fly with a pair of chopsticks, or dogs running in circle trying to catch their own tails), and actually one could hypothesize that the evident clumsiness and redundancy of the movements deliberately produced by the agents in these circumstances (without any external coercion) is exactly what betrays their goal-oriented intention. Were the optimality of the chosen means the only criterion to determine the action goal, we definitively would not be able to make any such inference. As for the second assumption, its weakness is suggested both by a deeper analysis of the requirements for efficiency, and by a closer consideration of the empirical preconditions that make biological movements more or less efficient. Doubts can be raised as soon as one considers the following two problems: first of all, on what basis do we decide that a movement is efficient? And how efficient are movements supposed to be (or seem) then, in order to be perceived as *the most* efficient? And, second, assuming that we have sufficient criteria and appropriate standards to measure such efficiency, is it true that the movements' efficiency is enough to infer their rationality? In what follows, I will argue that the real role played by efficiency can only be understood through a deep appreciation of the bodily details and bio-mechanical constraints of action execution.

2. Inferring efficient navigation *vs.* the Frame Problem

According to TSH, an observer is capable to tell the most economical movements to achieve a certain action goal even without any previous performative familiarity with that kind of action. TSH supposes this is possible thanks to a general form of mental means/ends calculus that applies both to the trajectories of translational movements and manipulative actions, regardless of whether they are executed by geometrical shapes or human beings: it is assumed that, in either case, parsimony – i.e. riddance of unnecessary or ineffective movements – is the mark of efficiency, and hence of rationality.

However (as noted by Sinigaglia 2008), efficiency evaluation is undoubtedly enormously simplified if we consider only tasks of spatial navigation, like in the experimental conditions appropriately arranged by the experimenters, as the only parameter considered is the length of the path accomplished to reach the endpoint. We cannot fail to notice that this simplification is already the result of a selection deliberately operated to highlight only certain components of the rationality embedded in our actions by excluding the others; moreover, even reducing the kinematic complexity and morphologic variability of the task at hand does not necessarily make it easier to evaluate the efficiency of the observed action. For example, the translational movements done by the whole body do not always follow a straight line: this is because of the physiology of the animal, and its behavioral and ecological specificity (e.g., rabbits' and flies' zigzag trajectories). The reasons behind the anti-economicity and apparent randomness of these movements (e.g., reduce the chances to be caught by a predator) can only be appreciated in relation to the specific evolutionary background to which the organism is adapted. Such background reminds us why moving from a point to another in an ideal Euclidean space could hardly be a distinct goal in and of itself for the animal, but rather a goal embedded in the ecological preconditions that characterize its intrinsic coupling with a specific world-environment.

Now, if the efficiency of the movements is correlated with trajectories that are likely to be more complex than a straight line, as is the case, because it depends on the bio-mechanical and behavioral specificity of the organism rather than Euclidean norms, then how could we judge some movements as more or less efficient in achieving their goal? The assessment of the efficiency of an action cannot possibly be dissociated from the consideration of its specific bio-mechanical and ecological constraints in so far they are not just contingent accidents or unnecessary complications, but constitutive motoric components that allow the realization of the action itself: the motoric specifications of the action execution (e.g., the degrees of freedom or the elasticity of the tendons in a rabbit leg) are not obstacles or extrinsic circumstances that would mark the distance between the actual action and an ideal standard; they are the medium itself in which the very possibility to efficaciously reach an embodied/situated motor goal is concretely inscribed.

That is why, outside of abstracted geometrical scenarios, like those used in the experiments of Csibra and colleagues, the distance travelled is just one of the many relevant factors that we might need to consider when we judge the efficiency of a certain action. In many cases, the shortest path is neither the fastest nor the most economical, for example in terms of energy expenditure: Rosenbaum (2008) shows that a fixed path length can be perceived as more or less demanding depending on whether or not the subject is walking in it to reach a certain object (and whether or not the object to be reached is conveniently located in relation to the subject' handedness). The psychological perception of the energetic cost of a movement is not an objective measure, but one that depends on the particular situation of the subject and her goal; however, even if we relied on a more objective calculation of the energy expenditure necessary to complete a certain translational movement, the most energetically economical movement would not necessarily be the most convenient one. If I need to reach my office at the first floor of the Philosophy Department building, I might have to choose: *taking the stairs* (fast, but long and energy-consuming); calling, waiting for, and then *riding the old elevator* (energy-preserving, but slow and long); or even *climbing the external wall* and enter from the window (short, but slow and energy-consuming). The choice largely depends on my contingent motivations (priorities) and embodied dispositions. All of these priorities and dispositions must somehow play a role in shaping a truly rational decision in so far they determine the whole range of the agent's actual possibilities to reach her goal in ways that could be deemed efficient. Note that the agent does not need to explicitly consider all these factors when she produces her movements to achieve her goals; it is sufficient for her to rely on her embodied expertise, i.e. her unaware skillful predisposition to interpret certain environmental circumstances as viable opportunities for action and cope with them in an effective (I am implicitly building here on the theory of absorbed coping, Dreyfus 2002). But how could anyone – as a mere observer

– assess which of the observed agent's priorities and dispositions could best characterize her choice as a rational one, in a given occurrent scenario?

In my example, the evaluation of the best means act to reach the office cannot leave out of consideration a combination of independent variables such as whether I am late, I have a bad health condition, or I am not keen on meeting a certain colleague in the building. The observers who witness the efforts of an agent to achieve a certain goal can of course decide to focus only on some of these variables, prioritizing them in the description of the scenario that we are using, in the attempt to judge whether the agent is effectively behaving in a way that maximizes the benefits of its decisions based on an ad hoc multidimensional matrix. But, is it not true that also this decision to focus only on a limited number of elements is itself a simplification made out of a selection of the complex background preconditions that might underpin the agent's choice? After all, when I have to decide how to reach my office, there is no a priori reason for which, say, the distance to my office would be a parameter objectively more important than the amount of tiredness I am experiencing in my legs. According to Dreyfus (1992), any selection of some parameters over the others is arbitrary if it is not embedded in the concrete familiarity with the transcendent complexity of a real-life scenario, which is of course very difficult to approximate by means of a third-person simulation. Only an agent that is practically familiar with that scenario could tell what decisional parameters are really relevant in those circumstances, and one could doubt that there is such a thing as a rational understanding of what could objectively and universally count as relevant without such familiarity. A totally disembodied and alien system, even if intelligent and generically enabled to take rational decisions, could hardly decide what decisions are rational without any familiarity with the context, regardless of how computationally powerful that system is. The programmers of GPS navigation apps know that path length, travel duration, and fatigue are just some of the key factors that inform our simplest decisions on "how to reach X". But these factors prove immensely complex when we attempt to simulate the bottomless background of massively interconnected preconditions that underlie the smallest rational decision in real life circumstances, including any choice of the best route to drive back home, which might require to consider an indefinite number of contingencies such as: the gasoline level in our tank, the road conditions, the cost of tolls, the weather, the amount of traffic, the aesthetic value of a scenic route, the proximity of a certain preferred restaurant or shop, the level of restlessness of the kids in the back seat, etc. The set of the relevant contingencies that might affect our simplest decisions can neither be fully represented nor reduced to a smaller set of elements, if we want to preserve the sensitivity to the real context of our intelligent processes. This is the well-known philosophical version of the "frame problem" of artificial intelligence (Dreyfus 1992; Wheeler 2008): the persistent difficulty to analytically spell out the determining factors behind any rational procedure, due to the impossibility to exhaustively list all (and only) the context-sensitive variables that would apply a priori to a certain procedure-based decision.

As the frame problem threatens any system that derives its intelligence from representation, it is not just a theoretical impasse for the cognitivist dream to build thinking machines based on internal models of the world and rules of thumb; it is also a serious difficulty for the rationalistic approaches to social cognition, insofar as they presuppose an attribution of rationality to the agent's decisional processes according to an abstractly universal standard of objective economy and parsimony. The problem with this kind of intellectualist approach to action understanding is twofold: first, how can we judge that an observed agent's action is rational, if we cannot choose the appropriate parameters to judge whether it is economic/efficient or not? And, second, how can we choose the parameters to assess the efficiency of an action, if we do not know the horizon of real-life contingencies that motivates its goal yet? Note that in order to *infer* – as postulated by TSH – whether an action is efficient or not it is not sufficient to have an intuitive grasp of, or some kind of pre-reflective holistic sensitivity to, the motivational background of that action; on the contrary, in order to make a legitimate inference on the other's rational decision it is indispensable to hold a fully explicit model of the premises of the inference – i.e. an explicit representation of the preconditions and motivational factors of that decision. In any other case, if we could realize someone's goal without disposing of any such representation, then it is probably not an inference that allowed us to draw such conclusion. The main problem of any inferential theory of basic action understanding – such as TSH – is that, ultimately, the background preconditions of our concrete rational decisions exceed our possibility of explicit representation and enumeration: this is not a problem for the embodied theories of action understanding that are not based on inferences – such as EFH – because such basic understanding does not require representations at all, as it only relies on a perceptual sensitivity to contextual opportunities for actions and a fluid responsiveness to the way fluctuating environmental conditions can affect us. The concreteness of real-life contexts is a problem for the inferential forms of action understanding because there are always too many parameters that would be in principle relevant to determining whether an action is actually efficient or not, and because the same action could well be efficient according to certain representations of the world and simultaneously inefficient according to others.

If the competing representations of the parameters of efficiency, with their overwhelming complexity, are the problem of basic action understanding, then there could only be two solutions: the first is that the agent's actions actually conform to certain standardized representations of his goals and priorities, and that the observer were instructed in this representation, being informed in advance of the real motivational background of the action. This way the observer could explicitly tell whether the action she is observing is economical or not. But this would not really help explain how goal recognition occurs in the first place, as it assumes exactly what TSH is supposed to explain: the attribution of a teleological valence to the observed action by means of an inference. Assuming that the agent *were told* by any of his world-representations what the goal of the agent is (as opposed to *inferring such goal by himself*), defeats the purpose of TSH as an explanatory psychological model of action understanding.

Alternatively, the observer should be familiar with the material non-representational preconditions and specific embodied constraints that shape the agent's action possibilities: in that case, it would not be indispensable to hold a representation of her goal in order to recognize whether her action is efficient or not; in order to judge whether the other is carrying out movements that are efficient or not it would be both necessary and sufficient, for the observer, to rely on his own practical competence, that is his embodied capability to perform those very movements first hand. But note that, if it is the case that goal understanding is based on such embodied familiarity with the sensorimotor circumstances of action execution, then the movements' goal would not need to be *inferred* anymore, as it is not by means of an inference that the goal is intrinsically associated to fine-grained patterns of characteristic coordinated movements. In this scenario, the goal of the action would not be inferred at all, as the form of understanding brought about by the required embodied familiarity is completely different from an inference: while an inference is an explicit reasoning that derives logical conclusions from known premises, the recognition of the actual goal based on embodied familiarity is purely based on the original associative connection, in the perceptual experience, of the movements and their motivating intention (Butterfill and Sinigaglia 2014).

Therefore, the evaluation of the efficiency of the movements, according to the perspective of embodied familiarity, is not only necessary and sufficient to recognize the movements' goal, but also concretely possible in the absence of an explicit representation of the reasoning parameters that contextually define whether the action is economical to achieve a certain goal: being capable to perform that action first-hand necessarily implies being familiar with both the implementational details of its execution in typical real-life circumstances and the goal that motivates the appropriate coordination of the movements, in a way that the two of them are intrinsically correlated and mutually presupposing one another. The knowledge of the parameters that make an action efficient would be an operative one, i.e. a pragmatic and habitual capability, not a set of information to be stored and computed in a declarative format. It would be based on the performative familiarity that the observer has matured with respect to that action and to its typical purpose, not on an intellectual calculus of means to be compared with an amodal representation of an abstract goal.

3. Inferring the goal of hand-actions *vs.* neurocomputational complexity

So far we have been considering the efficiency of translational movements only but, if we consider goal-oriented hand actions, then rational inference seems an even weaker option to explain goal attribution. Proponents of TSH mention Woodward and Sommerville (2000) as evidence that one-year-olds expect a hand to grab a target object with the most direct means action available, and that the target object is attributed as a goal of the action only if the hand acted efficiently to obtain the object. Furthermore, Gergely et al. (2002) claims that evaluating the rationality of actions is not restricted to passive observational contexts, because infants modulate their imitative behavior according to the justifiability of the goal-directed actions performed by a model.

This empirical datum is interesting per se, but does not necessarily back an inferential view on social cognition. In fact, similar experiments suggest that it is harder for an observer to recognize the goal of the agent's action, if the former has no expertise in performing the latter's movements. In accordance with TSH, Rochat et al. (2008) demonstrate that monkeys look longer at grasping actions whose trajectories are less economical than those typically followed in the same context with the same target; but this study also shows that monkeys' appreciation of the means-ends adequacy depended on their sensitivity to the goal relatedness of the observed movements, because observation of incongruent hand trajectories in actions that lack actual goals (when the target object is absent) do not evoke any attentional enhancement, even if the action in question looks very similar to grasping. Additionally, no attentional enhancement was observed when the monkey observed actions that, while visually similar to grasping, were actually motivated by goals that the monkey had never attempted to achieve before, such as lifting an object with the thumb. This suggests that the recognition of economical/efficient behaviors can support an inference on the nature of the action's goal only when a certain familiarity with the goal of the action already exists.

Therefore, either we conclude that the attribution of rationality is not indispensable to the mechanism of goal recognition, or that this mechanism involves attribution of rationality only insofar as the efficiency of the observed action is measured through the observer's acquaintance with the movements observed. This acquaintance would then be the most fundamental premise of goal attribution. While admittedly Rochat et al. (2008) do not allow specifying whether the required acquaintance must be merely visual or also performative in nature, it provides a serious reason to doubt that – in either case – a general system to calculate the movements' efficiency could work without relying on a repertoire of hand actions mastered by the agent. If we consider transitive actions with effectors (e.g., interaction with objects), the most efficacious movements (i.e. the movements that are well structured to perform a certain transformative function) are not necessarily the most bio-mechanically parsimonious: due to the composition and functional organization of muscles and joints, reaching and manipulative gestures (e.g., hand grasping, pulling, etc.) are usually not structured according to a simple geometry, i.e. they are not designed to reduce unnecessary component movements so that, for example, grasping between two fingers always requires the simultaneous coordination of five. However, importantly, the higher complexity of the morphological description of these movements does not imply a higher complexity of the cognitive information required for producing/controlling them. At least in some cases the opposite could actually be true, as the functional organization of the frontal and pre-frontal cortex suggests that motor areas map structured action goals and not isolated movements (Rizzolatti et al. 1988, Gentilucci and Rizzolatti 1990). In facts, movements intentionally produced in isolation without a motor goal are less frequent and not automatized, and that is why the voluntary independent movement of one single finger of the hand is neuro-computationally more demanding than a rich pattern of organized movements of the whole hand/arm/torso in a typical goal-oriented motor chain such as reaching-for-grasping-for-bringing-to-the-mouth. The modes of the cortical initialization and on-line control of these movement patterns reflect how, in real life interactions with the environment, the fine-grained coordination of the rich details of the movement (morphology, kinematics, velocity, etc.) are closely associated to the goal of the action itself. So closely that recognizing the former can be sufficient to immediately understand the latter, even just from the very initial movements of the hand, as complex patterns of functionally structured movements carry the distinctive signatures of either immediately transformative or social goals.

However, the close association between component movements and action goal does not mean that there is a one-to-one mapping of movements into goals, as very similar movements can be meant to achieve entirely different goals, and the same goal can be achieved by very dissimilar, or even entirely opposite, movements: as an example of the first case, consider that recognizing the type of pre-shaping of a hand that is moving to grasp a knife can be sufficient to guess whether the intention of the agent is to place the knife away or use it to stab someone (Becchio et al. 2008); as an example of the second, consider how different movements of a hand, a mouth, or a plier can be performed to grasp the same object; and that normal or reverse pliers require perfectly opposite movements of the hand in order to accomplish the grasp; and that nonetheless the goal-specific areas of the premotor cortex recruited to perform such morphologically dissimilar, but functionally and teleologically equivalent, movements are exactly the same (Umiltà et al. 2008). The motoric strategies and the effectors that can be used to achieve the same goal are too many, and with details that are possibly too fine-grained to allow an general system to evaluate a priori what movements are the most efficient ones, without previous acquaintance with them: the best way to reduce this complexity in the task of visual recognition is to be capable to perform first-hand those movements, re-using for purposes of recognition and categorization the motoric information that is primarily required to perform certain goal-oriented actions (Prevete et al. 2010).

That is why the motoric familiarity with an observed action, even if not necessarily an a priori precondition of goal recognition, is de facto playing a crucial role in the usual understanding of the others' action goals in real life scenarios: firsthand capability of execution (i.e. motoric familiarity), facilitates the visual assessment of an observed action's efficiency by enormously restricting the number of variables potentially involved (Metta et al. 2006). This has to do with the fact that the notion of efficiency is intrinsically embodied, because it depends on the bio-mechanical and material specifications of the medium, and also because it depends on the contextual way in which an embodied agent and its environment correspond to one another during their reciprocal interactions, continuously disclosing or obliterating the opportunities of their mutual modification: an action can be deemed more or less efficient only in relation to the range of motoric opportunities that the agent has contextually available, with their situational and embodied constraints, not the totality of the action possibilities that would be available to the agent just in principle or that are not available at all. For example if I could fly with a jetpack, or extend my legs five meters, then – in comparison – riding the elevator, taking the stairs, or climbing the wall would all appear inefficient means to get to my office at the first floor of the Philosophy Department building.

But motoric familiarity is not only indispensable to exclude an infinite number of actions that could possibly be efficient, if they were actually available in the current context; it is also important to distinguish between truly rational actions and randomly or incidentally generated movements. For example, all heavy objects fall towards the ground following a straight line, hence realizing a movement that consistently appears equifinal and economical in its structure, but we would never attribute purposefulness or animacy to this kind of natural movement. Were the observer unaware of the different movements potentially available to the agent, he would probably find it hard to tell whether she is moving rationally or just randomly; moreover, were the agent clearly forced by physical, irremovable, and insurmountable constraints to act in one specific way, and in that way only, we would hardly be inclined to attribute to him his action's goal, as the execution of that action would not appear relatable to a rational decision and hence not motivated by a goal. Why? This is because the distinction between natural and voluntary movements is not generated by the child's naïve theory of physics (Gergely et al. 2005, p. 173, building on Gopnik and Meltzoff 1997), but deeply ingrained in the phenomenology of our bodily experience. The concrete mastering of this distinction in real-life circumstances is realized at the sub-personal level by a combination of feed-forward and feedback systems that automatically compare motor commands with the perceptual feedback that follows action execution: since we are familiar with the experience of our body unintentionally falling, our system automatically recognizes that it is not an action generated by a goal-oriented motor program. When the same familiar pattern of movement engendered by gravity is recognized in another agent's body, our system realizes that there is neither agency nor purpose in it.

It is reasonable to suppose that, if we intuitively "know" (in the sense of mastering an operational know-how) what motor opportunities are realistic and actually available to us or to the agent we are observing, this is possible not because we were first mentally inspecting and then excluding an infinite number of improbable or impossible movements, but because certain habitual responses are pre-reflectively embedded in our motor system and immediately triggered by the environmental solicitations, automatically predisposing our conduct, adjusting our evaluative and practical attitudes toward the contingencies. This adjustment is on-line, based on real time perceptual stimulation, and relies on stored motor predispositions that prove suitable for the task at hand according to the subject's experience. This is exactly one of the ideas that characterize EFH, and in particular one that is defining for the direct-matching hypothesis in social cognition. Note that motor skills would be a requirement for action understanding regardless that the observer is able or not to attribute mental contents to the agent, as the recognition of constraints of this kind is more likely to be based on a mere perceptual registration of the physical context than on explicit knowledge of the other's intentions.

Conclusions: Inferring efficiency *vs.* perceiving efficaciousness

The experiments performed by Csibra and colleagues do not capture the minimal preconditions of action-understanding because they look for the wrong kind of minimalism: in fact, by focusing on generic visual stimuli (i.e. stimuli that are deemed to be generally significant for any disembodied agent, regardless of their contextual situation and background) they methodologically exclude the real phenomenology of the biological agents, overlooking the importance of their embodied dimension. The stimuli chosen by the experimenters are minimally specified precisely with the purpose to highlight an abstractly general structure of goal perception: they are intentionally vague, if compared to the specific richness of the bio-mechanical details that characterize functional movements of actual biological agents, because these bio-mechanical details are deemed redundant and accidental. However, I believe I have raised a sufficient number of doubts against the conjecture that the alleged generality of the inferential mechanism postulated by TSH could ever be deducted from the genericity of the visual stimulus that trigger attention in the infants, pointing out that different action goals imply entirely different criteria of efficiency, and typically the relevance of these criteria is correlated to a complex background of embodied constraints that concretely determine how the agent's opportunities of perception and motor execution are coupled with the complexity of real-life contexts. Consequently, vague visual action cues cannot suggest a general typology of goal and are not even sufficient to indicate agency, but (at best) can suggest an underspecified and vague sense of parsimony in generalized (not necessarily agential) movements, such as translations and rotations in an abstract metrical space.

To sum up, the first problem of TSH is that general de-contextualized visual cues are not sufficient to infer the general presence of rationality in the observed agent. Its second problem is that, even if some general kind of rationality could correctly be attributed to an observed agent by means of such cues, visual acquaintance with the movements she performs might be neither a sufficient nor a necessary condition to understand if and how her actions are efficient in pursuing their goals, therefore it would not be enough to guess their specific goals. Guessing the reason behind the movements of some professional football players might be extremely difficult to infer, if the observer has never been exposed before to any experience of team sports, while understanding the player's movements might be a lot easier if he is more familiar with playing such sports. The capability to evaluate the efficiency of this kind of actions is clearly more correlated to performative expertise (a requirement has nothing to do with rationality, in accordance with EFH) than to disembodied rational powers of mental calculus (as TSH implicitly suggests). And this is why recognizing the intention behind a complex motoric routine has to do more with the embodied skills of the athletes than with capabilities of abstract calculus like those cultivated by mathematicians and accountants for their jobs.

I should add that the parsimony of the observed movements, besides not being a sufficient condition for goal understanding, is neither a necessary condition for it: it is possible to attribute goals to one agent's action even when her chosen conduct appears strikingly inefficient and antieconomical, and efficient movements in turn do not always suggest agency or even animacy. On the contrary, simply acknowledging the agent's complex patterns of coordinated movements, based on the observer's familiarity with that action, seems both necessary and sufficient to understand the agent's action motor goal, whether this is pursued with an economical strategy or not. Therefore, even if it is true that goal attribution results from recognition of *appropriate* means, goal recognition and calculus of the most *parsimonious* means can well be achieved independently, without implicating one another.

Recognizing the appropriate and harmonious organization of the component movements of an action, rather than their parsimony, seems key to goal understanding. TSH, by reducing goals' understanding to a calculus of the most parsimonious means, loses sight of the fundamental distinction (Gallese et al. 2009) between *efficiency of the movements* and *efficaciousness of an action*: the former is reaching a certain target with the least topological transformations, the latter is fulfilling the expectation to produce certain functional modifications of the agent-environment relationship. TSH assumes the indiscernibility of these two different terms, as its main claim – that attribution of rationality is based on efficiency of movement – is true only for actions of which it is possible to equate efficiency and efficaciousness. Such coincidence only occurs in a narrow set of situations in which two conditions hold: first, the target (the destination of the movement) and the goal (practical scope) of the action can be conflated (this works well only for navigation tasks, i.e. actions whose goal specifically is to reach a certain distal point); second, metric relationships between points of the Euclidean space are the only variables that matter for the objective assessment of the most preferable (which are, coincidentally, the most parsimonious) movements (this only happens in those scenarios in which all the other variables, such as the tiredness of the agent, his willingness to avoid to be seen, etc. are irrelevant). These conditions typically do not hold for actions aiming at fulfilling the practical project of an embodied agent in real life circumstances, i.e. in all those scenarios in which space is not just a neutral topological parameter but the medium of situated engagement with the world.

In fact, this is the primitive spatial dimension of our original embodied phenomenology, one that is more fundamental than the intellectually sophisticated experience of measuring and comparing objective distances in an abstract Euclidean space. The primacy of the practical spatial know-how over a detached form of spatial cognition is not only backed by traditional philosophical ideas whose legacy is influential in embodied cognition (e.g., Poincaré's relationist view on space /1908/, Heidegger's pragmatist doctrine of the "ready-to-hand" /1927/, and Merleau-Ponty's motor intentionality and spatiality of situation /1945/). It is also supported by scientific models of adaptive decision-making and bounded rationality (Berthoz 2003; Gigerentzer 2008): they confirm that embodied agents reduce the computational complexity of spatial tasks, such as catching a ball on the fly, not approaching the problem in an analytic and detached way, but engaging in sensorimotor interaction with the world. This means that the details of their embodied situation not only constrains the range of their rational choices of the agents, but actively scaffold the development of their intelligent ways to negotiate the environment.

All moving animals, including human cognizers, evolved brains to cope with situated spatial tasks by exploiting a constitutive coupling between their sensorimotor skills and the opportunities of action offered by the environment, and they did this long before developing capabilities of objective quantitative analysis and neutral comparison of metric relations. The evolutionary and developmental background of action understanding is deeply rooted in the sensitivity to spatial contingencies whose meaning is practical, ego-centered, and intrinsically mapped into opportunities of bodily intervention, as opposed to context-independent, universal, and decoupled from bodily habits. Different experimental results (for example those on canonical neurons, or the transformations in far/near space due to tool use, Costantini et al. 2014), confirm that one subject's goal-specific motor skills are selectively and flexibly recruited by his brain to make sense of the physical surroundings and highlight the opportunities of bodily interaction that are most appropriate to the spatial context. As I mentioned before, this picture does not rule out TSH, which is a valid model to explain at least one out of the many possible cognitive strategies that we use to attribute goals without relying on demanding meta-representations; but it undermines its presumption that the fundamental mechanisms of social cognition are inferential in nature and, therefore, it subtracts credibility to the claim that TSH is more fundamental and primitive than EFH.

References

Anscombe, E. (1957). *Intention*. Oxford: Blackwell.

Astington, J.W. & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology*, *9*, 7–31.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mind reading. *Curr. Psychol. Cogn.*, *13*, 513–552.

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention. *Consciousness and Cognition*, *17*, 557–564.

Bertenthal, B.I., Proffitt, D.R., Spetner, N.B., & Thomas, M.A. (1985). The development of infant sensitivity to biomechanical motions. *Child Development*, *56*, 531–543.

Berthoz, A. (2003). *La Décision.* Paris: Odile Jacob.

Bratman, M. (2007). *Structures of Agency: Essays*. Oxford UK: Oxford University Press.

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

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*, 187–192.

Cappuccio, M. (2012). Conoscenza nella mani. Come comprendiamo le azioni degli altri. In I. Gamelli (ed.), *Ma di che corpo parliamo? I saperi incorporati nell'educazione e nella cura*. Roma: Franco Angeli.

Cappuccio, M. & Wheeler, M. (2012). Ground-level Intelligence: Inter-context frame problem and dynamics of the background. In Z. Radman (ed.), *Knowing without Thinking. Mind, Action, Cognition and the Phenomenon of the Background*. Basingstoke: Palgrave Macmillan.

Costantini, M., Frassinetti, F., Maini, M., Ambrosini, E., Gallese, & V., Sinigaglia, C. (2014). When a laser pen becomes a stick: Remapping of space by tool-using observation in hemispatial neglect. *Exp. Brain Res.*, *232* (10), 3233–3241.

Csibra, G. & Gergely, G. (1998). The teleological origins of mentalistic action explanations: a developmental hypothesis. *Dev. Sci.*, *1*, 255–259.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*, 237–267.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cogn. Sci.*, *27*, 111–133.

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philos. Trans. R Soc. B Biol. Sci.,* 29, 447–458.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Does social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14* (10), 441–447.

Dennett, D.C. (1987). *The Intentional Stance.* Cambridge, MA: MIT Press.

Di Pellegrino, G. di, Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research / Experimentelle Hirnforschung / Experimentation cerebrale*, *91*, 176–180.

Dreyfus, H.L. (1992). *What Computers Still Can't Do*. Cambridge, MA: MIT Press.

Dreyfus, H.L. (2002). Intelligence without representation: Merleau-Ponty's critique of mental representation the relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, *1*, 367–383.

Froese, T. & Gallagher, S. (2012). Getting interaction theory (IT) together. Integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction Studies*, *13 (3)*, 436–468.

Gallagher, S. (2001). The practice of mind: Theory, simulation, or primary interaction? *Journal of Consciousness Studies*, *8 (5–7)*, 83–107.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17 (2)*, 535–543.

Gallagher, S. & Hutto, D. (2008). Understanding others through Primary Interaction and Narrative Practice. In T. Zlatev, T. Racine, C. Sinha, & E. Itkonen (eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins, 17–38.

Gallagher, S. & Povinelli, D.J. (2012). Enactive and behavioral abstract accounts of social under- standing in chimpanzees, infants, and adults. *Rev. Phil. Psych.*, *3*,145–169.

Gallagher, S. & Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*, 23–48.

Gallese, V. (2010). Embodied simulation and its role in intersubjectivity. In T. Fuchs, H.C. Sattel, & P. Henningsen (eds.), *The Embodied Self: Dimensions, Coherence and Disorders*. Stuttgart: Schattauer, 78–92.

Gallese, V., Rochat, M., Sinigaglia, C. & Cossu, G. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology*, *45* (1), 103–113.

Gergely, G. (2002). The development of understanding self and agency. In U. Goswami (ed.), *Blackwell's Handbook of Childhood Cognitive Development*. Malden, MA – Oxford – Carlton – Berlin: Blackwell, 26–46.

Gergely, G. & Csibra, G. (1994). On the ascription of intentional content. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, *13* (5), 584–589.

Gergely, G. & Csibra, G. (1997). Teleological reasoning in infancy: the infant's naive theory of rational action. A reply to Premack and Premack. *Cognition*, *63*, 227–233.

Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7* (7), 287–292.

Gergely, G. et al. (2002). Rational imitation in preverbal infants. *Nature*, 415, 755.

Gergely., G, Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gibson, E.J., Owsley, C.J., & Johnston, J. (1978). Perception of invariants by five-month-old infants: Differentiation of two types of motion. *Developmental Psychology*, *14*, 407–415.

Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. New York: Oxford University Press.

Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16* (1), 1–14.

Gopnik, A. & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hutto, D.D. (2013). Action understanding: How low can you go? *Consciousness and Cognition*, *22 (3)*, 1142–1151.

Mandler, J.M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*, 587–604.

Merleau-Ponty M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Metta, G., Sandini, G., Natale, L., Craighero, L., & Fadiga, L. (2006). Understanding mirror neurons: A bio-robotic approach. *Interaction Studies*, *7*, 197–232.

Morganti F., Carassa A., & Riva G. (eds.) (2008). *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*, Amsterdam: IOS Press.

Poincaré, J.-H. (1908/1996). *Science and Method*. London: Routledge.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*, 1–16.

Prevete, R., Tessitore, G., Catanzariti, E., & Tamburrini, G. (2010). Perceiving affordances: a computational investigation of grasping affordances. *Cognitive Systems Research*, *12* (2), 122–133.

Rizzolatti, G., Camarda, R., Fogassi, M., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Exp. Brain Res.*, *71*, 491–507.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review Neuroscience*, *2*, 661–670.

Rochat, M., Serra, E., Fadiga, L., & Gallese, V. (2008). The evolution of social cognition: goal familiarity shapes monkeys' action understanding. *Current Biology*, *18*, 227–232.

Rizzolatti, G. & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*, 264–274.

Rosenbaum, D. (2008). Reaching while walking: Reaching distance costs more than walking distance. *Psychonomic Bulletin & Review*, *15* (6), 1100–1104.

Sinigaglia, C. (2008). Enactive understanding and motor intentionality. In F. Morganti, A. Carassa, & G. Riva (eds.), *Enacting Intersubjectivity: A Cognitive and Social Perspective to Study of Interactions*. Amsterdam: IOS Press, 17–32.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*, Harvard University Press

Wheeler, M. (2008). Cognition in context: phenomenology, situated robotics, and the frame problem. *International Journal of Philosophical Studies*, *16* (3), 323–349.

Wellman, H.M. (1991). From desires to beliefs: acquisition of a theory of mind. In A. Whiten (ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*. Oxford: Basil Blackwell, 19–38.

Wellman, H.M. (2002). Understanding the psychological world: developing a theory of mind. In U. Goswami (ed.), *Blackwell's Handbook of Childhood Cognitive Development*. Malden, MA – Oxford – Carlton – Berlin: Blackwell, 167–187.

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

Woodward, A.L. & Sommerville, J.A. (2000) Twelve-month-old infants interpret action in context. *Psychol. Sci.*, *11*, 73–77.

Umiltà, M.A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., & Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 2209–2213.

Massimiliano L. Cappuccio

Zaključivanje ili upoznatost?

Utjelovljeni korijeni socijalne spoznaje

Sažetak

Razmatram dvije različite deflacijske teorije u socijalnoj spoznaji koje nastoje objasniti razumijevanje djelovanja bez potreba za metareprezentacijskim procesima ili procesima čitanja uma. Prva je »hipoteza teleološkog stava« (TSH), koja tvrdi da o namjeravanom cilju određenog opaženog djelovanja zaključujemo samo na temelju opažanja učinaka djelovanja i situacijskih ograničenja. Drugu teoriju sam odlučio nazvati »hipoteza utjelovljene sličnosti« (EFH) kako bih obuhvatio sve teorije koje tvrde da namjeravani cilj određenog djelovanja prepoznajemo na temelju opažajne ili motoričke ekspertize razvijene unutar senzomotoričkih kontingencija povezanih s kontekstom djelovanja. Temeljni je zahtjev TSH da promatrač može pripisati efikasnost, stoga i racionalnost, opaženim pokretima djelatnika, dok je temeljni zahtjev EFH teorije taj da je promatrač na neki način izložen opažajnim ili motoričkim detaljima opaženog djelatnikovog djelovanja. Tvrdim da EFH opisuje primitivniji i temeljniji oblik razumijevanja djelovanja, tj. onaj oblik koji TSH nužno pretpostavlja: ustvari, iako prepoznavanje efikasnosti nije ni nužan niti dovoljan uvjet za detektiranje povezanost i s ciljem, neka vrsta opažajne ili motoričke upoznatosti s detaljima opaženog konteksta djelovanja uvijek je nužna za bilo kakvo pripisivanje efikasnosti, a time i racionalnosti, promatranom djelatniku. Zaključujem da, iako TSH zasigurno može biti učinkovita u opisu određenih racionalnih oblika razumijevanja djelovanja, implicitno zahtijeva da EFH bude istinita, budući da bi takav sustav zaključivanja bio neutemeljen bez pretpostavljene upoznate pozadine utjelovljene ekspertize.

Ključne riječi

socijalna spoznaja, teleološki stav, zrcalni neuroni, racionalnost, intencionalni stav, hipoteza direktnog odgovaranja

Massimiliano L. Cappuccio

Schlussfolgerung oder Vertrautheit?

Verkörperte Wurzeln der sozialen Kognition

Zusammenfassung

Ich ergründe zwei unterschiedliche deflationäre Theorien innerhalb der sozialen Kognition, die angestrebt sind, das Verständnis des Handelns zu erläutern – ohne die Bedürfnisse nach metarepräsentativen oder gedankenleserischen Prozessen. Die erste ist die ‚Hypothese der teleologischen Haltung' (TSH), die behauptet, dass wir das beabsichtigte Ziel eines bestimmten beobachteten Handelns auf der Grundlage der bloßen Wahrnehmung dessen Effekte und Situationseinschränkungen erschließen. Ich habe beschlossen, die zweite Theorie ‚Hypothese der verkörperten Vertrautheit' (EFH) zu nennen, um sämtliche Theorien einzuschließen, die darauf bestehen, dass wir das beabsichtigte Ziel eines bestimmten beobachteten Handelns auf der Basis perzeptueller oder motorischer Expertise erkennen, die innerhalb der sensomotorischen, mit dem Kontext dieses Handelns verbundenen Kontingenzen entwickelt wurde. TSHs Hauptanforderung besagt, der Beobachter könne die Effizienz, und damit die Rationalität, den beobachteten Bewegungen des Handelnden zuschreiben, während EFHs Hauptanforderung lautet, die Beobachter seien auf irgendeine Weise perzeptuellen oder motorischen Details des beobachteten Handelns des Agierenden ausgesetzt. Ich vertrete die Ansicht, dass die EFH eine primitivere und fundamentalere Form des Handelnsverständnisses beschreibt, d. h. jene Form, die von der TSH notwendigerweise vorausgesetzt wird: In der Tat, während die Effizienzerkennung weder eine notwendige noch eine hinreichende Bedingung für die Detektion der Verwandtheit mit dem Ziel ist, erweist sich eine Art perzeptuelle oder motorische Vertrautheit mit den Details des beobachteten Handelnskontextes als ständig notwendig für jedwede Zuschreibung von Effizienz, und demgemäß auch von Rationalität, an den beobachteten Handelnden. Ich ziehe die Schlussfolgerung, dass während die TSH in der Schilderung gewisser rationaler Formen des Handelnsverständnisses sicherlich effektiv sein könnte, verlangt sie implizit das Zutreffen der EFH, da ihr Folgerungssystem ohne den angenommenen bekannten Hintergrund der verkörperten Expertise haltlos wäre.

Schlüsselwörter

soziale Kognition, teleologische Haltung, Spiegelneuronen, Rationalität, intentionale Haltung, Hypothese des direkten Zusammenpassens

Massimiliano L. Cappuccio

Inférence ou familiarité ?

Les racines incarnées de la cognition sociale

Résumé

J'examine deux théories de cognition sociale déflationnistes et distinctes, visant à expliquer la compréhension de l'action sans recours à des processus méta-représentatifs ou à ceux de lecture de pensée : la première est « l'hypothèse de la position téléologique » (TSH), affirmant que nous inférons le but visé d'une certaine action observée en nous fondant sur la simple perception des ses effets et de ses contraintes situationnelles ; j'ai décidé de baptiser la seconde « l'hypothèse de la familiarité incarnée » (EFH) afin d'englober toutes les théories qui affirment que nous reconnaissons l'objectif visé d'une certaine action en nous appuyant sur l'expertise perceptuelle ou motrice développée dans le cadre des contingences sensorimotrices associées au contexte de cette action. Le critère principal de TSH est que l'observateur pourrait attribuer l'efficacité, et par conséquent la rationalité, au mouvement de l'agent observé, tandis que le critère principal d'EFH est que l'observateur soit en quelque sorte exposé aux détails perceptifs ou moteurs de l'action de l'agent observé. J'affirme qu'EFH décrit une forme plus primitive et fondamentale de la compréhension de l'action, soit une forme qui est nécessairement présupposée par TSH : en fait, tout en reconnaissant que l'efficacité n'est ni une condition nécessaire ni suffisante pour détecter le rapport avec l'objectif, une sorte de familiarité perceptuelle ou motrice avec les détails du contexte de l'action observée est toujours nécessaire à toute attribution de l'efficacité, et par conséquent de la rationalité, à l'agent observé. Je conclus que, tant que TSH pourrait certainement être efficace pour décrire certaines formes rationnelles de la compréhension de l'action, elle requiert implicitement qu'EFH soit vraie, puisque son système inférentiel serait sans fondement s'il était sans un contexte familier supposé de l'expertise incarnée.

Mots-clés

cognition sociale, position téléologique, neurones miroirs, rationalités, position intentionnelle, hypothèse de correspondance directe