Vol 3 | Issue 1 | Pages 86-117

Annals of Cognitive Science
ISSN: 2642-4290

Research Article

DOI: 10.36959/447/345

"Mind Reading": How and for what?

*Cristiano Castelfranchi**

GOAL lab ISTC-CNR Roma, Rome, Italy

Abstract

This paper tries to explain: How we tend to automatically ascribe mental representations to social actors on the basis of scripts, roles, categories and prejudices, norms, and several heuristics; or by default; how scripts and roles should be filled in with the actors' mental attitudes; how social interaction systematically requires assumptions about the other's mind; how those mental attitudes can be the non-intended or non-understood function of our behavior/role. What really matters is that we assume that those beliefs and goals are there, and we act "as if" it were so. A further claim of this work is that this mechanism of mind ascription while reading another's behavior or its traces (stigmergy), is the basis for a fundamental form of communication: Behavioral Implicit Communication (or BIC), which works without words or special protocols. An efficient coordination-in humans but also in artificial agents-exploits (or should exploit) not only the mere 'observation' of such agents, but more precisely this form of silent communication (which should not be confused with non-verbal or expressive communication). The message-sending paradigm dominating CSCW, MAS, HCI, and H-Robot-I, is here criticized; necessity and advantages of BIC for coordination and cooperation are presented.

Keywords

Mindreading, Mind ascription, Mind writing, Scripts, Behavioral communication, Stigmergy, coordination, "As-if" minds, Turing' test, AI anthropomorphism

Premise

What will be presented in this work is not an already implemented and experimented AI system, but only some foundational (and possibly inspiring) ideas and theoretical premises for a sociality grounded on "intentional stance" towards and from artificial intelligences. Understanding others' behavior implies having a "theory of mind"; our central question is how to model the mind of other agents. Three claims will be made and argued for, derived from crucial mechanisms of our mindreading-based sociality. It is emphasized that we have to endow AI agents with the same human capabilities, if they have to socially interact with us and among themselves[a] [1].

"What our mindreading or mentalizing abilities in fact consist in has been a matter of fierce dispute. But the dispute exists against the background of a consensus over the fact that folk psychology, mindreading or mentalizing is the backbone of the social world..... Although still being something near to consensus, the idea that social interaction hinges mainly on folk-psychological mindreading is no longer universally accepted. When it comes to explaining social interaction, the variety of competing theories purporting to account for folk-psychology and mindreading *are no longer the only options available".* (for example, more embodied approaches like empathy or simulation theory) [2] p.1[b]. In this contribution we will illustrated another -not enough considered -option, by stressing the role of *automatic ascription* of minds based on scripts and on Behavioral Communication.

[*c]**Thesis 1:** In order to interact with us, autonomous agents must 'communicate' about their minds with their actions, and 'read' our minds from our actions: And they should use human-like heuristics that have been proven to be economic and effective.

[a]For a comprehensive survey of this problem in AI and Agent theory see [1], which starts so: "Much research in artificial intelligence is concerned with the development of autonomous agents that can interact effectively with other agents. An important aspect of such agents is the ability to reason about the behaviours of other agents, by constructing models which make predictions about various properties of interest (such as *actions, goals, beliefs*) of the modelled agents."

[b]For a good overview of the Theories of Mind debate, see Marc Slors and Cynthia Macdonald [2].

[c]Along this contribution we will mark with a * all the local remarks about the possible interest or application of given models and human social mechanisms in AI Agents, (Ro)Bots, HCI and HRI. We will recall them in § 10.

***Corresponding author:** Cristiano Castelfranchi, GOAL lab ISTC-CNR Roma, Rome, Italy

Accepted: August 14, 2019

Published online: August 16, 2019

Citation: Castelfranchi C (2019) "Mind Reading": How and for what?. Ann Cogn Sci 3(1):100-117

Thesis 2: Mind reading is not only functional to predicting others' behavior and adjusting our own behavior to theirs (which is the classic view) but for "mind writing", that is, for changing others' mental contents, in order to influence them, and make them do what we want. This also applies to autonomous agents (in other words, we will manipulate each other).

***Thesis 3:** A non-linguistic communication - behavioral and stigmergic -is essential for autonomous agents (including robots). What we read is in fact some behavior and its traces, not a mind [3].

The paper is divided into two sections[d]: The first about basic heuristics for ascribing minds from behavior observation; taking behavior as a "sign" and reading it, or better *interpreting* it (from §4 to §8) [4]. The second (§9) is about the transition from mere behavioral "signs" to behavioral "messages", from mere "signification" to "communication": How we intend to communicate about our mind [5] (or have the functional goal to communicate) while performing a practical action (non specialized for communication, different from non-verbal behavior) or by leaving traces (stigmergy). This form of communication is fundamental for human coordination but also for artificial agents in smart environments and for HR interaction.

Some misunderstandings about mind and its reading

The cognitive approach to social phenomena is often read with the glasses of the philosophical debate on mind reading, TheoryTheory, etc. vs. mirroring, simulation, empathy, alignment; and of the "Language of Thought".

"In general the tradition of mental states ascription [6,7] is presently subject to reductive interpretations and objections, which are based on a misunderstanding, a wrong Fodorian conception of "mental representation" as necessarily language-like, or propositional, logically constrained, and "subjective". Postulating an explicitly represented belief or goal in X's mind is equated to postulating the agent's awareness of such mental representations. However an explicit mental representation (belief, goal, etc.) is not a synonym for "conscious representation". Mind/cognition is not equal to consciousness; this holds since Chomsky's anti-behaviorist revolution". [4].

Moreover mental representations are not necessarily in a propositional/conceptual format; our *goals* (which notion comes from cybernetic notion of "set-point" that has to be matched with perception!) and our *beliefs* and *expectations* can be mental imageries, sensory-motor representations.

We will put aside these mind ascription mechanisms (TheoryTheory, etc. vs. mirroring, simulation, empathy, alignment, etc.), which are definitely important but not exclusive, to focus on other humans tricks not less important and more relevant for artificial sociality.

[d]This paper is a development of previous works. The first part elaborates on ideas presented in [4], the second part develops a talk, "Silent Agents: From Observation to Tacit Communication", given at *Iberamia-SBIA* 2006.

In reference to Thesis 2, it should be remarked that the monumental literature on the functions of mindreading and "theory-of-mind" has neglected the role of mindreading in social influence, with few exceptions ([8-12]. Instead, scholars prefer to emphasize other adaptive functions: Interpretation, prediction, and social coordination (see for example [13-17]. However, the function of "action" is to change and exploit the world/environment, so a crucial reason and function of "social action" is to change/exploit others' behavior by changing their minds.

According to a widespread view, "mind reading" - or better the reading of Y's behavior for abducing Y's mind (mental representation behind the behavior) - is not true because it would be too computationally costly or intractable, too reasoning-based, too risky; and it does not correspond to our phenomenal experience, that is rather intuitive, immediate, and not reasoning-based.

Actually, in order to read another's behavior in terms of mental contents, that is, in order to ascribe goals and beliefs to him or her, it is not necessary a lot of reasoning, specific individual information, and so on.[e] Aims of this paper are to explain:

- How we just ascribe or presuppose others' mental representations on the basis of scripts, roles, role-signs, categories and prejudices, the use of certain tools, and several heuristics; or by default;

- How we are able to ascribe minds even without projection, empathy, explicit communication, and reasoning, although these are important too [12,18-20].[f] Mind reading (interpreting behaviors in mental terms) is made fast and simple by applying those contextual constraints and a-priori general schemes and knowledge. It is rather automatic: Due to non-ordered, incomplete, overlapping and redundant heuristic rules, quite effective but not rarely fallacious [4].

We will illustrate:

- Script and role-based ascription;
- Norm- and identity-based ascription;
- Ascription by default and by projection;
- Some specific heuristics.

[e]This view-strong in philosophy, cognitive sciences, and social psychology - is not dominating in AI research, especially that exploiting machine learning: "The process of constructing models of other agents, … often involves some form of learning since the model may be based on information observed from the current interaction and possibly data collected in past interactions. For example, an agent may model another agent's decision making as a deterministic finite automaton and learn the parameters of the automaton (e.g. nodes, edges, labels) during the interaction. Similarly, an agent may attempt to classify the strategy of another agent by using classifiers which were trained with statistical machine learning on data collected from recorded interactions" [1].

[f]The "dual processing" tradition ([12,18-20] is now opposing the conscious, slow, and costly reasoning process to the intuitive, emotional, empathic one. In our view the latter system is fast, automatic, associative, implicit ... but not necessarily affective and socially empathic.

In the same vein, a merely functional coordination and cooperation does exist, which can do without intentional "goal delegation" and "goal adoption" [21,22], tacit agreements, and so on. Very fundamental forms of human interaction and cooperation, even rather simple and automatic, require the mutual postulation of specific mental states and their signaling, without complex reasoning. A lot of cooperative social phenomena are not grounded on cooperative minds but just on functionally orchestrated behaviors.

Why should "social models" (scripts) be opposed to mind reading (see Maibom [23])? "Social models" are precisely one of the main tools for ascribing mental attitudes. They make such an ascription rather automatic (and prescriptive) (see § 4). It is simply wrong that in order to form representations, expectations, beliefs or assumptions about the mind of the other one should either abduct, by a complex and risky reasoning, her/his mental attitudes from her/his behavior, or simulate, imagine her/his mind by identification. Alternative is not simply between mirroring/intuition and reasoning/abducing; we have to move beyond this alternative.

Others' minds (not only their behaviors) are inscribed in, and prescribed by, scripts, roles, games, norms, tools, etc. When one is assigned a particular role, (s)he is officially - that is, for the public - assigned a particular mind. We do not simply put on a mask ("persona") for playing that social representation. We rather impersonate and represent a particular character, with given motives and beliefs. Thus, by simply looking at another's mask, we know his or her mind; we are entitled and even requested to assume that (s) he has certain beliefs and goals. And on such a basis we are supposed to interact with him or her, and to play our own role/character.

"This is what we learn since the time of "symbolic play", when we start to pretend to have certain goals and beliefs and that the other has certain goals and beliefs. Roles and characters are not simply behavioral; they are mental and emotional. Scripts (or "Social models")[g] [23] are not an alternative to mind ascription/reading; they are one of the alternatives to complex inferences and abductive reasoning. But this was precisely the cognitive motivation of scripts: For reading others' behavior, for expectation and plan-recognition, and also for behavioral decisions" [4].

*Ascribing mental representations (feelings, attitudes, goals, beliefs) is mainly an automatic process in human beings; either based on basically innate devices (mirroring, empathy,...) or on cultural well-defined memory structures. That's why we need "acculturated" AI agents [24].

Minds to Wear

Scripts should not be reduced to behavioral and behavioristic structures. This would be a superficial interpretation: A script specifies not only what I have to do but also what I have to (or I'm supposed to) believe and want.

A script [25,26] is a specific kind of conceptual network. It represents a sequence of nodes corresponding to sets of events or actions, expressing different 'permissible alternatives' or 'functional equivalents', or 'functional complements' ('terms') that may be substituted into the nodes. The sequence may be temporal, logical or causal.

A typical Script representation just as a *sequence of events/actions* is the restaurant script:[h]

e1 [enter, nsubj, {customer, John})] e2 [enter, dobj, {restaurant}]

e3 [order, nsubj, {customer}] e4 [order, dobj, {food}]

e5 [eat, nsubj, {customer}] e6 [eat, dobj, {food}]

e7 [pay, nsubj, {customer}] e8 [pay, dobj, {bill}]

e9 [leave, nsubj, {customer}] e10 [leave, dobj, {restaurant}]

Temporal Ordering = e1 < e2 < e3 < e4 e4 <e5 <e6 <e7 <e8 <e9 <e10

Scripts are schemas for both recognizing and understanding behavior and for planning and performing context-adapted behavior. Notice that this is a very interesting, modern, and topical view, connecting - in a pragmatist perspective - the structures for action recognition and those for action control and execution.

Scripts as a form of generalized episodic knowledge (a chain of events in common activities), as a formalism for representing causal interactions between events in given contexts, demonstrate that a simple rule-based method makes it possible to infer causal relationships between connected events.

The inclusion of mental representations in scripts does not make them lose their (computational and cognitive) advantages. Mental representations are just *inscribed* in the scripts, in their roles and typical actions and tools; and their recognition remains automatic, not reasoning-based. As an example of how traditional scripts presuppose a lot of mental stuff, consider the restaurant script: We not only find the observable events, the expected behaviors of a customer or a waiter (to enter, to order, to eat, to pay,...), but also the *reasons* for such events: The fact that the customer has the *goal* of eating, and this is why he enters in a restaurant, orders, and so on. Tools (forks, menu, bill,....) also play an important role in behavior prescription and mind ascription: In fact their use implies a goal in the user: The actualization of their function.

For example, after we accept to sit at a table, the waiter (W)-without asking or saying anything-may just give us the menu: For doing so, W is assuming that we might have the goal to read the menu and that we believe we will find there the list of foods we can order (today); and that we have the

[g]Maibom [23] does not mention the term "scripts". He introduces a less specified but very close notion of "social models" for explaining the same kind of phenomena. However, "social models" also include specifications of the expected (prescribed) mental attitudes of the actors.

[h]Adapted from Niels Kasch & Tim Oates Mining Script-Like Structures from the Web http://www.aclweb.org/anthology/W/W10/W10-0905.pdf

goals to choose some food. Or is W just performing a prescribed behavior in the script sequence and expecting our next move? If this were the case, why is it possible that W adds, "Would you like to know what's the special food of the day? It is not on the menu, but I can tell you"; or "Sorry, but we are out of ravioli". Why "sorry"? If not in relation to a possible goal/desire of us and to its consequent frustration? And why is it possible that we refuse to read the menu, and say, "No, thank you, I already know what I want". When C (the customer) takes the menu and/or looks at it, this "means" (and communicates): "C intends to read the menu, and choose some food from the list"; when C puts on his glasses this *means* (and communicates): "C intends to read the menu". Menu is for reading it, for knowing what kind of food is here (today), and - on such a knowledge basis - for ordering. It is just a tool inserted in a goal-chain. When C starts to read it W can ascribe him such a goal-chain.

We can and must even ascribe second and third layer beliefs and goals to the actors: W's belief that C believes....; W's belief that C wants....; W's goal that C believes/wants....

A reductive, merely behavioristic, view of scripts cannot be seriously supported. It is true that the behaviors included in a script are routine and mostly automatic moves, but part of these automatic moves are the *mental assumptions and attitudes* to be adopted step by step [27].

When adopting a 'role' or entering a 'script' or a 'game', X automatically adopts the goals, assumptions, tools, required by that role, necessary for playing that game; while people automatically ascribe such a mental repertoire to X, and take it as a common ground. This is precisely one of the advantages/reasons for creating institutional roles: They are functional to endow individuals with a publicly known and prescribed mind, in terms of goals (motivations, values, mission) and of assumptions and skills (knowledge) [4]. Those mental assumptions are part of the "rules of the game": If you want to correctly play it you have to assume so and so, and have such and such goals. Your mind is strictly part of the game.

Minds Ascribed because Prescribed

As said, we have not only to read others' minds for coordinating our behaviors and, more importantly, for influencing others and changing their conduct by changing their goals; we also have to ascribe mental attitudes to others in accordance with the roles they play in scripts, conventions, tools, norms. Moreover, we do not simply *ascribe* (or predict) specific mental states and decisions to them, but we *prescribe* those minds. When interacting (or about to interact) with us, people take into account not only norms, but also their *expectations* (Exps) about our minds. But Exps [28] are not simple predictive beliefs; they also imply goals. In 'positive' Exps we both foresee (as certain or probable) and want an event p to occur. In 'negative' Exps we both foresee and don't want that p occurs. In other words, we not only have beliefs about another's mental attitudes and behaviors; we also have goals about their goals. And these goals are not only private hopes and wishes--they are usually imposed on our targets; they are real *prescriptions* [29], something required as due: A deontic conformity of their minds (and conduct) to their roles, scripts, norms, conventions: They *should* have those mental attitudes or at least should behave "as if" they had them. For example, if John plays the part of the waiter in a restaurant script, I will not only believe that he has certain beliefs and goals, but he *should* have them; it is prescribed by his role and by me (as client). And in fact scripts (roles, etc.) are not only necessary for reducing predictive uncertainty [30] and coordination complexity, and for choosing whether to trust them or not [31]; scripts also serve to *shape* our fellow beings, for prescribing particular goals and behaviors to them. Wearing a uniform, suit, smock etc. in a given role means wearing a ready-made mind.

The same applies to (social, moral, legal) norms (and in fact scripts have a normative character). Norms usually prescribe visible things, such as behaviors; however, they also prescribe invisible things, such as mental attitudes. In fact, to start with, they prescribe and shape behaviors *through* mental devices; secondly, mental states as such are often culturally prescribed, expected, or prohibited. Our mind can be prescribed or prohibited. Some of the main cases of norms about minds are:

a) Meta-norms for the implementation of behavioral norms (norm adoption);

b) Norms about good or bad minds per se.

Meta-norms: You have to 'obey' not just 'do'

The aim of a norm is not just our behavior; in fact, the norm is not really satisfied by an accidental conformity. The norm wants to be followed for specific reasons [32]: The agent *should* be intrinsically motivated to 'adhere' to it (regardless of external rewards or punishments), for a non-instrumental goal of respecting the authority and its norms. This is ideal "obedience".[i] More precisely, any norm implies meta-norms:

-A meta-norm of obeying that norm (and the authority's will);

-A meta-norm of not violating norms in general.

This is the real intentional effect of a norm: This is what one *should* care about.

And there even is another crucial component of these meta-Ns: Not only you have the N/commandment to obey the norm as norm, but (as we said) to do that for specific motives, thus to really (mentally) "obey". A three layered structure:

N1: Obligation/Duty/You Have to: Do (in) action A

N2: Obligation/Duty/You Have to: Obey, not violate N1

N3: Obligation/Duty/You Have to: Follow/Adhere to N2 for the specific motive of respecting Authority, doing one's own duty.

[i]We define 'authority' as a normative notion and institution; it is characterized by, and endowed with, a special kind of 'power over' others: the 'normative power', that is the power of influencing people's (intentional) behavior through the expression of its impersonal "will", regardless of any other possible higher goal, such as doing a favor, bullying, threatening, promising, exchanging, explaining and persuading, …

Let us focus on N3: The implications of our motivation to "obey". Consider the order of an army general: It should not be "obeyed" because of courtesy, sympathy, friendship, pity, fear, money, ... but just because it is an order of the right person; this is its "ideal" working, its aim [32]. So, the norm wants:

(i) my behavior,

but due to

(ii) my goal (mind), due to

(iii) my adhesion (mind),

(iv) motivated by (higher-goal) internalized non-instrumental values.

Ideally we 'have to' adhere to the norm

- not instrumentally, i.e., not for obtaining external rewards or avoiding punishment;

- not because of external explanations of the norm's utility;

but

- for and as recognition of the authority; and thus signaling such a recognition and subjection.

This is the mind *expected and prescribed* to us as norms subjects, and thus ascribed to us.

Why is deontic 'autonomization' and 'appropriation' of law/order (which is more than simple 'internalization': Goal adoption, decision) so important?

Not only because in such a way no external monitoring and sanctions will be strictly necessary (cost reduction, higher probability) but also because no external control can compete with the eye of consciousness on the self [33].

> One of the main advantages of the 'internalization' of the prescription, of such final, intrinsic motivation, is that while we *discipline ourselves* (both through a Super-ego imposition and control, and through habituation - routines, scripts, rules,..), we no longer perceive the reduction of our autonomy/freedom. This self-normation is part of the "spontaneous" subjection of people to power;[j] which is not so "spontaneous" and just exploited, but in part is constructed by social powers.

> Moreover, an advantage is that such kind of subjection is an example of N and authority recognition, and it spreads around not only that specific behavior but also those values: Building the authority and norms confirmation.

> Last but not least, this normative use of our self-regulation, this social-actions towards ourselves (persuasion, imposition, control, punishment, promises, ...) [3] give us the feeling and power to "choose" in a ruled way, rationally and canonically: We influence and govern ourselves and our behavior. The internal discipline is not just for society power, but also for feeling an individual auto-nomos (giving Ns to ourselves).

[j]Etienne De La Boètie "Discours de la servitude volontaire", 1595

We feel that not only about our mind but also about the others' mind.

The relation between norm violation and autonomy is not only semantically but psychologically intrinsic. What gives us autonomy and makes us feel autonomous is precisely the fact that we can violate norms; and thus that we self-impose the prescription and the obedience to it.

Forbidden minds and deontic mind ascription

Unsurprisingly, religious and moral norms do not prescribe or proscribe only (in-)action. They typically prescribe or proscribe mental contents: (Not) to think or desire or feel something can be by itself a norm violation; it can be "bad" or "good" independent of its practical consequences. For example, in our culture, a mere pedophile desire or fantasy is immoral and sordid. Although is not punishable by law, it is socially reprehensible and punishable, as well as a possible source of self-blame and self-punishment, of shame and guilty feelings, because of an interiorized moral value. The same holds for racism, homophobia, and so on. A "bad" agent is not just an agent who *does* something bad; it is already bad if (s)he has negative or antisocial desires and attitudes.

Also this device is important for mind-reading/ascription: Except we have clues, reasons for presuming that the mind of this guy is "bad" (norm violating), like something he did or said, or prejudices about his group/culture etc., by-default we ascribe him a correct mind, conform to cultural norms. There is a complex inferential reasoning, intention recognition, only if I see that you are not conforming to the norm (and this does not correspond to a clear role, group, etc. of you, like anarchist, delinquent, mad, ..., or to a conflicting norm). In this case I wonder why, I try to understand what you have in mind, since my automatic and usual expectation has been violated and I need a special explanation of your strange/deviating mind (§ 3).

Social norms and conventions are a basis for the habitual, automatic, by default ascription of certain *beliefs* and *goals* to others' minds; and they work thanks to such presupposition.

No doubt, respecting certain norms can be fully routinized, and become a merely automatic unconscious reflex. The subject is no longer aware of the norm, that is, of his or her normative belief ("It is forbidden....") and normative decision. However, on the one side, it is always possible to evocate and explicit consider such normative beliefs and goals. For example, we automatically stop at the red light, but if an ambulance with a loud car speaker arrives behind us and asks us to move, in that very moment we might explicitly consider the norm to stop at the red light, and that presently it is better to violate it.

On the other side, unconscious automatic obedience is not possible for several norms, and we have to consider the existence of the prescription and make real decisions. For example, stopping at a policeman's signal is not fully automatic; paying a fine or tax requires some reflection and a conscious deontic reasoning; and so on. Moreover, what matters for social order, is that it is "as if" the agent were remembering and considering that there is a norm (see below).

Mind reading for presuming deception from the other

To believe or suspect that somebody lies is intrinsically a form of "mindreading"-our view about of his or her mind. In fact "lying" is different from "making false statements": It is (a) Saying something different from what we believe to be true, (b) In order to deceive someone else. Therefore, viewing another as a liar implies ascribing to him or her both a belief (different from what he says) and a goal: To deceive.

By default we assume that the other knows the truth. We 'expect', not just 'predict', this.

*This dimension of sociality (very relevant in competition, commerce, politics, etc.) will be quite important also in human agents or agent-mediated human interaction, also considering that deception can even be a tutelary attitude [34].

In H-C interaction, computer-supported cooperation and organization, computer- mediated commerce, intelligent databases, teams of robots, etc. there will be purposively deceiving computers. In particular, within the agent-based paradigm there will be deceiving agents. Several kinds of deception will be present in interactions between artificial agents and users, or among people via computer, or among artificial agents-not only for malicious reasons (war, commerce, fraud, etc.) but also in good faith and in our own interest. Social control, trust [22], and moral aspects in artificial societies will be the focus of theoretical work as well as of implementation [34].

*In this regard, there is a nice statement by Asimov:[k]

The inhumanity of the computer is in the fact that once programmed and put at work, it behaves in a perfectly honest way.

However the statement is (or will be) wrong: Socialized computers, able to interact in a hybrid society, implying competition, enemies, protection, etc. will necessary become more and more human-like; that is, able to deceive, to violate (moral, social and legal) norms. Especially in commerce, in politics (fake news), but also in helping us. Thus:

"in hybrid situations where artificial agents interact with human agents it is important that those artificial agents can reason about the trustworthiness and deceptive actions of the human counterpart" [35].

In the same vein, artificial agents will be able to deceive.

"…we have the following possibilities:

a) The agent deceives for its principal: i.e. the mandatary deceives through its agent

b) The agent deceives autonomously;

c) The agent deceives its own principal or user" [35].

And, as we know, to intentionally deceive somebody, the agent has to read the mind of its victim, in order to transmit the "false". Thus AI agents interacting with humans have to read their minds.

[k]"Change! Seventy-One Glimpses of the Future" 1981, Chapter 6: Page 17, Houghton Mifflin Company, Boston, Massachusetts.

Other Tricks

Let's give a look to some other rules, heuristics, and tricks for fast and costless ascription of mental stuff to others.

Common ground and presuppositions

Any social interaction among adults - not only "conversation" [36] - requires and presupposes a "common ground", which consists not only of common knowledge about the world, shared perception, etc., but also of beliefs about each other and our mental attitudes: Knowledge, beliefs, and goals/motives; in particular, social beliefs and motives.

If X is about to perform an action (for example, opening a door), Y can assume that X assumes/presupposes that: (a) The door can be opened (it is not locked); (b) X is able to do this; (c) It is a true door connecting rooms and one can pass through it; and so on.

Mind ascription from shared values and membership

If I see signs of your belonging to a given group of people, I automatically ascribe the values of that group (and its habits, uses, etc.) to you. And if I recognize signs of your membership to my group, I immediately assume that we share the same values; and I would feel betrayed if this were not the case, because you - by spreading those signals of belongingness (§ 8) - are deceiving people and their spontaneous trust in you as a member of the same group and as one who shares the same values.

One of the most fundamental functions of mind-reading, or better of "mind-ascription", is the possibility to share mental states: In particular beliefs and goals. In the human species, this is also an end in itself; it can even motivate social contagion and conformist attitudes: "I'm like them, I'm one of them, they accept and like me,... also because (they know) I have the same values, beliefs, motives" [4].

Mind ascription by default and by projection

A very basic and simple mechanism of attribution of mental states is based on the projection (in a specific common context) of our own beliefs and goals to others. This is a simple default rule:

Unless you have specific reasons (signs) for ascribing different beliefs and different goals to others, assume that he has your mind, that both of you share the context-related beliefs and goals [37].

This basic default attribution of our own mental states to others is simpler than imagination and identification, which implies the ability to put myself in the other's shoes. Moreover, notice that such "simulation" eventually seems to imply a sort of projection that is the ascription of what I have in mind to the other's mind.

Other heuristics: Some examples

Simple rules triggering beliefs.

> Looking at:

If X is looking at object O/ event E, and doesn't seem "absentminded" ==> "He sees /knows O/E" I see that you see ==> "You know". "Shared attention" also implies (implicit or explicit) assumptions on what the other knows; not only shared representations of the world.

> Memory ascription:

If we know that somebody was present at a particular event and was looking at it or reacting to it, we are likely to assume that (s)he will remember that event. Rule: Since he was there and noticed it (and since I remember that too) ==> he has to remember that.

Notice that there is no real "reasoning"; it is just a production rule, an automatic attribution of memory: "How is it possible that you do not remember that? You were there and you said....".

> Action recognition:

"Actions" are defined in terms of goals; they are "for" something, and this qualifies them. Actions are memorized and stored in a plan library relative to their outcomes; and are retrieved on the basis of their outcomes, that is, their goals. In a sense, actions are tools, like material instruments: They are "behaviorally-built" artifacts to be used again; to "slice" is a tool, like a knife is. To recognize an action means to recognize its goal [17]; thus, it is quite understandable that one automatically ascribes the goal of the action to (the mind of) its agent.

> Tool use:

Independent of complex scripts (which include actions with tools), the simple use of a tool with its perceived "affordance" and function induces the ascription to the agent of an intention actualizing that use and function: For instance, if one takes a knife, this means that (s)he has to cut something.

Automatic inferential rule for goal-ascription: *If X is doing action A or using tool T, and A/T has the goal/ function of G, than X intends that G.*

> Speech acts:

Speech acts -as well as other intentional actions- also induce the recognition/attribution of non explicitly declared mental attitudes in the speaker.

If X says, "I want that you go out!" we will obviously believe that X wants so; but this is also true for the sentence: "Go out!" We immediately know that aim of X's, without any reasoning, because we understand the meaning of the imperative. This is why a perfect and immediate answer might be: "Why do you want that I go out?".

In sum, all the above mentioned automatic tricks are as important as mirroring, empathy, intuition, identification, and also affective expressivity, non-verbal communication, and so on, for reading and signaling our mental attitudes. The opposition in the human mind is not between reasoning and rationality, on the one hand, and emotion and affective activation, on the other hand; automatic and routinely cognitive processes are not less significant.

Pretended Minds in Society "Symbolic Game"

Let us now consider more radical cases, where we conventionally 'presume' goals from another's behavior and deal with them "as if" they were mentally represented.

Ascribing role functions as intentions

For playing a social function (e.g., the role of father or that of leader) it is not necessary for the agent to fully understand it, to have a mental representation of its "goal", and intentionally pursue it; that is, it is sometimes unnecessary to "adopt" that goal [38] and to pursue it as an internally represented intention. "Intention" is the last step of mental goal processing. Goals are of different kinds (like "desires", "needs", "obligations"....); an intention is a goal that the subject has evaluated as "possible" (achievable), not already realized or self-realizing, preferable to other active goals, and thus "chosen"; and when "planned" and situated, and skills, resources, and conditions are ready to be used, an intention will be translated into "actions", that is, "executed" [3,39].

The goals/functions of a role do not necessarily coincide with the internal goals of an agent, even when his or her behavior is not just a routine. Social functions can impinge upon agents and be played through personal intentional actions without being intended [40,41]. However, we may sometimes consider an agent's role and consequent behavior in this mentalized, internalized, perspective; especially when the role is institutional and formal - for instance, not the role of "father" or "leader", but that of "policeman" or of "professor". In those cases we feel allowed to ascribe at least a partial understanding, introjection, adoption of the role's functions; hence, of the mental representation of the goal-structure specific of that role (but not necessarily of the more high-level goals: Social order, social cooperation, ..). [4].

It is not so important if we ascribe that goal to the agents mind, or else to the character he is playing, or to the social institution. We adjust ourselves to that "pursued" goal. This is what matters. Goals are even detached from persons, who actually "pursue" them, "alienated", in social and institutional life.

In sum, "pursuing" goals is not necessarily a "mental" operation; consequently, also "ascribing" goal pursuit to an agent is not necessarily a real "mental" ascription. It may be "as if" the agent were intending the goals he is pursuing/ realizing.

*No doubt, we need goal-oriented, teleological agents. But those agents can be either "intentional" AI/agents or simply "function-oriented"; they can pursue goals without understanding and representing them as goals.

"Implicit" mental stuff

As we said, beliefs, goals, explicit mental representations are not necessarily "conscious" (Sect.1): "The unconscious is not identifiably less flexible, complex, controlling, delib-

erative, or action oriented than its counterpart"[l] [42]. The ascription of a belief or a goal to another's mind is a belief, which can be either conscious or unconscious. Analogously, the ascribed belief may actually be unconscious in the other's mind; that is, (s)he has the ascribed belief without being aware of it and of its role (for instance, the activation of some goal). Moreover, ascribed mental attitudes can be just "implicit": Either non-activated in a particular moment, absent from the agent's mental processing; or "potential". "Implicit" beliefs are part of the game, and part of the agent's mind - especially of his or her "social" mind, that is, the mind, which people deal with. In fact, people do not dwell on considering whether a particular belief is already explicit and activated in an agent's mind or it is just implicit or silent).

> Non-activated beliefs remain in the background of the process. They are activated only in case of unexpected, surprising conditions. For example, while walking towards the door of our office to open it, we believe that the floor will sustain us, and that the door can be opened. These beliefs are not necessarily "active" and taken into account. Since this action/plan is part of a routine, they remain silent. Only in problematic situations (for example, a strange noise from the floor) we will activate them and work on them to understand what happens and solve the "problem".

> By "potential" belief (a relevant kind of "implicit" beliefs) we mean here a belief that is not 'written', that is not contained in any 'database' (short term, working, or long term memory) but is only potentially known by the subject because it can be derived from actual beliefs. For example, while my knowledge that Buenos Aires is the capital of Argentina is an explicit belief that I have in some memory and I have just to retrieve it, my knowledge that Buenos Aires is not the capital of Greece/Italy/India, and so on, is not contained in any memory, but can be derived (when needed) from what I explicitly know. As long as it remains a merely potential belief, it has no effect on my mind. Suppose for example that I believe p as well as q, but p implies Not q; if I have never derived Not q from p, I cannot perceive possible contradictions between my beliefs.

For example, if you are a seller and I have to pay, when I give you the money I not only assume that you "know" that that is "money", and this belief is active in your mind and used for recognition of money; but also that you believe that this money has an exchange value, that one can circulate it, buy and pay by it,....

In a sense, ascribed mental attitudes[m] are "institutional" objects; or better, scripts, roles, social games, all are "institutions" [43]; they are socially constructed and ascribed minds[n]. One of the reasons why the "as if" works is that our behavior (in a given context) signals particular beliefs or goals even when they are just implicit (see § 8). It doesn't matter whether the agent has them or not in his or her mind: In any case, one can ascribe such mental attitudes to the agent, and interact with him or her "as if" (s)he had them. It works.

Procedural "tacit" knowledge

As said by Nooteboom [44] "Scripts can be retained, in the sense of being present in knowledge in a procedural or declarative sense, in a weak or a strong form. In a weak form, only the set of nodes is retained, but in strong form also the 'architecture', in terms of logical, causal or temporal connections between the nodes, in the script. Note that one can procedurally retain a script in strong form (there is an architecture), without knowing it in a declarative sense. One may also retain a script procedurally only in weak form. In that case one is in a stage of trying to cope with a novel modus operandi without a determinate architecture; the beginning of a new script, which has not yet consolidated into a prototype".

*AI agents and robots too have to "play" their roles, with their "functions" (practical and social) like be "consumers" or "game partner" without necessarily an explicit representation of these functions, of the goals they are pursuing as character. Our hybrid society with them necessarily will ground on such "as if" ascription and conventionalization of behavior meanings and finalities.

Not Just Reading but Writing Minds

Sociality doesn't just require modeling the other's mind in term of imagining, recognizing, ascribing, abducing; that is, understanding and predicting what she has in mind. "Modeling" the other's mind means also "shaping" it. This is the function of culture, education, norms, values, etc., which make high predictable interaction. On this we agree with Tadeusz Zawidzki[o] [45,46]. But also personally shaping - in a given context for a given purpose - the mind of the others: Direct *influence* and *manipulation*. To change Y's behavior we have to change his mind driving the behavior, thus we have to have some map of Y's mind. As we said one of the main function of mind-reading is influencing: Changing the mind of the others.

This means that we not only have 'beliefs' about the other's mind (about her beliefs, her goals, her emotions) but

[l]And this holds also for complex social interaction. See section on "Social Behavior as Unconsciously Guided by the Current Context" in [42].

[m]Or better 'assumptions', which in our vocabulary means that I act on the ground of such implicit presupposition: The belief has a pragmatic function and is supported and verified not by reasoning and arguments but by action/behavior.

[n]In this perspective, Dennett's reluctance to take a position on the reality of the states postulated by the "intentional stance" (Do beliefs, desires, etc, really exist? Or are they just external constructions and descriptions?) acquires a different sense. It is not just a matter of epistemology and truth theory. The point is that it doesn't matter whether those beliefs and desires really exist or not in the other's mind (see below).

[o]Recently also Peters Uwe [45] strongly converges on this: "Why do we think about and ascribe propositional attitudes (beliefs, desires, intentions etc.) to people? On the standard view, folk psychology is primarily for mindreading, for detecting mental states and explaining and/or predicting people's behaviour in terms of them. In contrast, ... Zawidzki maintains that folk psychology is not primarily for mindreading but for mindshaping, that is, for moulding people's behaviour and minds (e.g., via the imposition of social norms) so that coordination becomes easier." p.1.

that we formulate 'goals' about the other's mental states: We have *goals* about her beliefs, about her goals (meta-goals), about her emotions. Also our goals about her behavior must be translated into goals about her goals and decisions, since we assume that her behavior is controlled by her goals and her preferences due to her beliefs. Thus "modeling" is not just representing but it is designing and building.

How do we *give* goals to the other? By various mechanisms we already mentioned:

> By *imitation*: We make the other to understand which is our goal, why and how we are doing a given action, in order she formulates the same goal, imitates us (since we know that we are her model, or teacher, or leader, etc.).

> By *adoption*: Our personal, selfish goal becomes her goal, is pursued by her, in order we achieve it; for us. Either for altruistic help or because she has an advantage by that outcome.

> By *adhesion*: I make the other know my goal about her goal/action in order she adopts it [21,22].

- Either I have power over her: She has to "obey" (§ 5.1) to my request/order;
- Or for "exchange" (I depend on her, she depends on me: I pursue and realize her goal IFF she realizes my goal);
- Or for "cooperation" in strict sense: We have the same objective but individually we lack such power: We need a co-power: I do my share and she does her share.

We orient the choice of the other and change her preferences by providing arguments: (True or deceptive) *beliefs* about advantages or costs and risks of possible alternatives. But we can also orient her choices by evocating goals she forgot or was not considering: Evocating values, norms, identitarian features ("This wouldn't be worthy of you!").

In order to "give" to the other a goal not necessarily we have to make manifest our goal over her mind, our goal of influencing her. We can hide our intention, and just "manipulate" the other's mind without she is aware of our influencing.

Thus, obviously, another way for ascribing (reading) mind to Y is just by "writing" it by us. We can write and change the others' minds since we read them; and we can read them also because we write them.

Behavior as Communication about Minds

Our behavior is not just (arbitrarily) "read" in term of mental background ("intentional stance"), but it socially works thanks to this, it is aimed at being interpreted/read in such a way.

Usually, communication is based on specialized signals linked to specific meanings. In order to use these signals, we have either to learn them or refer to some sort of innate knowledge. The former is the case of spoken language that we learn during our childhood; the latter is the case of expressive communication, which is in some way innate, so that we know what a smile means without previous learning.

Besides these two kinds of communication, there is a third one that we call *Behavioral Implicit Communication* (BIC), where there isn't any specialized signal, but the message is implied in the practical behavior itself. This form of communication shouldn't be mixed up (as often happens) with the "non-verbal communication" or "expressive behavior", i.e. gestures, conventionalized signals, facial expressions, and so forth. All this sort of *specialized* messages (either cultural or inborn) are different from simple practical actions like walking or sitting. The difference will be clear in our definition (below); see also [47].

BIC is very useful in a coordination context, where we send a message to our interaction partner(s) by simply performing an action. This message can even be intentional, i.e. the sender wants that the receiver knows that the sender is performing that action and why. However, this is not necessarily so. The message presupposes a more primitive and basic substrate: The unilateral capability of the agent to observe the other's behavior and to "read" it, to understand what she is doing, what she intends and plans to do (her goals), or at least (in rule-based systems or neural net systems) to predict and expect her next position or action [21]. In other words, communication is based on and exploits 'signification'- the semiotic ability of cognitive agents; for example the ability to take 'smoke' as a sign of 'fire', or to ascribe thirst to a drinking agent.

BIC evolution from observation

Behavioral Implicit Communication is a wide-ranging phenomenon, which acquires more and more complexity as the communicative intention becomes more important.

In the weakest form of Behavioral Communication, agent A is not acting in order to let agent B understand what A is doing (this isn't one of A's motivating intentions); A is simply aware of this possible result of her behavior and lets it happen. This is a very weak form, because there is no real intention: The communicative result is just a known (but not motivating) side-effect of the action.

True BIC requires that the communicative effect should be an intention of the sender, or at least a side-intention-that is, the sender anticipates that his/her action will also have a communicative meaning (additional result) and likes this, but his/her action is not *aimed* at achieving this communicative goal; is not motivated by it. What is relevant here is that:

> The expected side-effect that one lets happen is good but neither necessary nor sufficient for one's action.

> The sender knows and likes that the receiver sees and understands the behavior in question.

This is a side or passive intention.

A third step occurs when the sender chooses and performs the practical action also with the purpose to make the receiver see and understand it. The communicative effect is

necessary (although not sufficient) for the sender's action. This is full Behavioral Implicit Communication since communicating is part of the sender's aims and motivates the action.

The last step is when the behavior is performed only for communication and loses its practical purpose (or even the practical effect). In this case, the act is either

> Faked (simulation, bluff) (which is very important in conflict coordination); or

> A ritual, in that the action has fully become nonverbal communication or a conventional symbolic "gesture" with a practical origin.

What we have just described here is the 'intentional' [5] path in BIC evolution, when communication is intentional. But there are more primitive forms of BIC for sub-cognitive agents without true intentions, based on reinforcement learning or on selection (see § 7.1.2).

*Also these forms are applicable to MAS: The first one is more suitable for BDI-like agents, while the other two fit well with reactive or ALife agents. We claim that in Human-Human, H-Agent, H-Robot [48], Ag-Ag, Robot-Robot interaction [49] the possibility of communicating through an action could be the solution for several coordination problems, and, more than this, the interaction becomes natural and intuitive.

Conditions for establishing BIC from observation via evolution, learning, or inference: Let us now briefly consider the cognitive prerequisites of BIC. Our aim is to show how primitive are its components and how easy is an interaction based on BIC, either at the intentional level or at the functional one. But first we want to better define what we mean by "learning": What is learned is the communicative use of an action, not the action or the gesture itself. In other words, agents do not learn the action but the latter, performed in order to reach a pragmatic goal, acquires a possible use as a message.

The first preliminary condition for BIC emergence through learning, evolution or reasoning is that:

X's movements or positions or X's traces or footprints are perceivable by Y. X's behavior leads to an environment modification which we will call m/t (movements or traces).

A second condition is as follows:

Given a behavior of X's, Y should be able to recognize, understand, ascribe meanings to m/t.

This could happen at the level of pattern matching (key stimuli, activating a reflex behavior), or even better at the level of categorization. Here, what Y perceives should be meaningful to him, either as the perception of a current behavior or as its detached sign. In both cases there is a sort of "sign" and a sort of "signification"; thus the second condition is some capability of Y to "understand" something. At the higher cognitive level the meaning should be "X did/is doing this action", and even "X intends to do so", "X has such a goal", "X believes that…". At the level of merely reactive agents, the recognized stimulus should activate a specific response, which is the left part of a production rule "IF…THEN…". Even at this simplest level we can say that X's behavior "means" something in Y's mind.

The third condition is:

X perceives/understands that when (s)he performs behavior Bx, Y either reacts with the expected behavior (By) and/or perceives m/t and understands it (second condition).

Put together, these three conditions imply that:

X performs Bx also in order to make Y perceive (and read) m/t.

The famous assertion of the Palo Alto school (Bateson, Watclawicz) that "any behavior is *communication*", and that in social situations "it is impossible do not *communicate*" is wrong precisely for the lack of the crucial distinction between "signification" and "communication". They reduce "communication" to simple "signification". Whereas it is impossible not to "signify" something to others with our behavior, we are not necessarily "communicating" what the other understands. What makes a practical action a "message" (not simply a "sign") is the goal (intention or function) of the "source" ("sender") about the other's reading and understanding.

Thus, we can summarize the steps for establishing BIC via evolution, learning or reasoning:

X modifies the environment (she leaves m/t) => Y perceives that this modification has a sort of "meaning" for him (we have signification and not yet communication) => X perceives that her action leads to a behavioral response by Y => X performs that specific action also in order to make Y perceive m/t.

This sequence, although very basic, can account for different kinds of phenomena, with different degrees of complexity. It describes what happens either at the intentional level (X intends that Y sees his traces and believes that…) or at the functional one, where X's behavior has been selected by evolution or reinforced through learning.

In "functional" BIC the practical action -beyond the possible intention of the "source"-acquires (by evolution or design) a systemic function of informing the addressee. This is, for example, the case of Stigmergy in insects [50,51] where communication is not an intention but just a function of the physical trace left in the environment [52,53].

*****BIC can be successfully generalized to the domain of the interaction among artificial agents, where robots or agents can observe the traces left by other agents in the physical or virtual environment, and learn a correspondence between these and an appropriate action. Communication protocols, or natural language and expressive gestures and "faces" (NVC) will not be enough. Indeed, they are not enough for human-human interaction and coordination!

The stigmergic over-generalization: The notion of "stigmergy" comes from biological studies on social insects; more precisely, the term has been introduced to characterize how termites (unintentionally) coordinate themselves in the reconstruction of their nest, without sending direct messages to each other. Stigmergy is essentially the production of a certain behavior in agents as a consequence of the effects produced in the local environment by previous behavior [51].

This characterization of stigmergy cannot discriminate between simple signification and true communication[p] [51], and between prosocial and antisocial behavior. It would for example include prey-predator coordination as well as a thief (unintentionally) leaving footprints, which are very precious for the police. In order to have "communication" proper, it is not enough that an agent coordinates its behavior with the (traces left by the) behavior of another agent. Also this is an overgeneralization. A definition of stigmergy as "indirect communication through the environment" is in our view rather weak and unprincipled.

Any kind of communication exploits some environmental 'channel' and some physical outcome of the act; any communication is "through the environment"! This cannot be the right distinction. The real difference is that in stigmergic communication we do not have specialized communicative actions, specialized messages (that unambiguously would be "direct" messages because would be just messages); we just have practical behaviors (like nest building actions) and objects, which are also endowed with communicative functions. In this sense communication is not "direct" (special communicative acts or objects) and is "via the environment" (i.e. via actions aimed at a physical and practical transformation of the environment). The message is also not necessarily addressed to a specific addressee, but this holds for the other forms of communication too.

From our perspective, stigmergy is communication via post-behavioral [43] physical practical outcomes. Moreover, in insects (and some simple artificial agents) stigmergy is a functional form of behavioral communication, where the communicative goal cannot be represented in the agent's mind (intention) but it is a functional effect selected by evolution or built in by a designer. This means that stigmergy is a sort of "innate" and trace-based Behavioral Communication (see later) and, from our point of view, very effective especially for sub-cognitive agents.

Stigmergy is only a sub-case of BIC, because in fact any BIC is based on the perception of an action, which necessarily means the perception of some "trace" of that action in the environment (for example air vibrations). We restrict stigmergy to a special form of BIC where the addressee does not perceive the behavior (during its performance). The only difference is that when we refer to communicating via "traces" we have in mind traces that persist after the practical action: The "receiver" observes these traces while could not observe those who performed the action. Moreover, for a trace-based communication to be a form of stigmergy it is necessary that both the perceived "object" be a practical one and the originating action be performed for practical purposes (like nest building).

Stigmergy concerns not only insects, birds, or non-cognitive agents. Many examples are also present in human behavior. Consider a sergeant that - while crossing a mined ground - says to his soldiers: "Walk on my footsteps!". From that very moment any footstep is a mere consequence of a step, plus a stigmergic message (descriptive "here I put my foot" and prescriptive "put your foot here!") addressed to the followers. Also consider the twofold function of guard-rails: On the one side, they physically prevent cars from invading the other lane; on the other side they in fact communicate that "it is forbidden to go there" and also normatively prevent that behavior. This is what law theorists call "materalization" of a norm: The norm its "hardwired" because there is no possible decision to infringe it.

What is 'coordination' and why it needs 'observation' and 'signification'

'Coordination' is that additional part or aspect of the activity of an Agent specifically devoted to deal and cope with a dynamic environment interferences, either positive or negative, i.e. with opportunities and dangers/obstacles.

A 'common word' in fact means 'interference' (the action of an Agent Y could affect the goal of another Agent X) [21]. Thus X has to perceive (or infer) those 'interferences' in order to avoid or exploit them.

Coordination is not necessarily "social" (one can coordinate himself with a rolling stone); also when social, it is not necessarily mutual or cooperative as is usually assumed to be, but anyway it must first be among minds in order to be among behaviors.

The basic forms of coordination are:

Unilateral: X just coordinates her own behavior with Y's, ignoring Y's coordination or non-coordination activity.

Bilateral: X coordinates his behavior with Y's observed behavior; and Y does the same. Bilateral but independent: X coordinates his behavior with Y's observed behavior; and Y does the same, but in an independent way.

Reciprocal: X coordinates his behavior with Y's behavior by taking into account the fact that Y too is coordinating her behavior with X's behavior.

Mutual: it is based on symmetric and interdependent intentions and mutual awareness (shared beliefs). Both X and Y wants the other to coordinate with his/her own behavior and understand that the other intends to coordinate with her/his own behavior.

Coordination can also be distinguished into:

Reactive: Is the coordination response to an already existing and perceive obstacle, for example after a first collision.

Anticipatory: Is the coordination in relation of a not yet perceived event, announced by some sign and foreseen by the agent.

As we said, even when bilateral and reciprocal, coordination is not necessarily cooperative. Also in conflict and war there is coordination, and clearly is not cooperative and is not for cooperation. Prey and predator for example (consider a

[p]For example "…Coordination of the agents' movements is achieved through stigmergy. The principle, initially developed for the description of termite building behaviour, allows indirect communication between agents through sensing and modification of the local environment which determines the agents' behaviour" [8] pag. 181).

leopard following a gazelle) coordinate with each other: The leopard curves left and right and accelerates or decelerates on the basis of the observed path and moves of its escaping prey; but at the same time the gazelle jumps left or right and accelerates or not in order to avoid the leopard and on the basis of the observed moves of it. Their coordination moves mainly are 'reactive' (just in response to the previous move of the enemy), and for sure non-cooperative, not mutual. Moreover, they obviously are not communicating to each other their own moves (although they are very informative and meaningful for the other). This is an observation based but not a communication/message based (BIC) reciprocal coordination.

Observation for coordination: One of the main functions of observation in agents living in a world inhabited by other agents is coordination. In order to coordinate with a given event or act X should perceive it or foresee it thanks to some perceptual hint, 'index' or sign. In other words, observing and interpreting the world in which other agents are acting and pursuing their goals is usually an intrinsic necessity of coordination. In social coordination, X must observe the other agents' behaviors or traces for understanding what they are doing or intend to do. In sum coordination is based on observation and -more precisely - on 'signification'[q].

BIC for coordination: A large part of coordination activity (and social interaction) is not simply based on observation but on BIC. For example, in mutual coordination mere signification is not enough: True BIC is needed. Actually, since X wants that Y coordinates his behaviors observing and understanding what she is doing, she is performing her action also with the goal that Y reads it, i.e. she is communicating to Y -through her action -what she is doing or intends to do[r] [54].

In coordination the most important message conveyed by BIC is not the fact that one intends to do action *a* or why one intends to do *a*, or the fact that one is able to do *a*, etc. (see section 5). The most important message concerns when, how, and where one is doing one's own *a* so as to permit coordination with other agents who share the same environment.

As already pointed out, coordination is possible without any communication both in human and artificial societies [21] (see also Franklin[s]). This is an important statement against common sense. However, coordination usually exploits communication.

Since BIC is a) a very economic, b) a very spontaneous, c) a very practical and rather effective form of communication that just exploits the side effects and traces of actions, and the natural disposition of agents to observe and interpret other agents' behavior, a rather important prediction follows:

> One can expect that agents perceiving and acting in a common world will use a lot of BIC and will spontaneously develop it. A very large part of communication for coordination in situated and embodied agents exploits the reciprocal perception of behavior or of its traces and products; i.e. it is just BIC. Even more, (second prediction).

> Both in natural and in social systems a lot of specialized (conventional or evolutionary) signs derive from BIC behavior that has been ritualized.

*This kind of observation-based, non-special-message-based communication should be much more exploited in CSCW[t] [48] and computer/net mediated interaction, in Multi-robot coordination, in Human-robot coordination, in MA systems [55].

Behavioral communication for coordination (in particular stigmergy) has some nice properties and advantages that deserve to be stressed. BIC is naturally and intrinsically 'situated', merged in concrete location, time, and objects. Thus it transmits this kind of information in a perceptual, immediate way without any special, abstract, arbitrary codification. More explicit and symbolic forms of communication induce the receivers to reinterpret such signs and to re-apply them to the actual context, via deictic reference, whereas by using BIC the addressee extracts the information directly from the environment (while/when 'observing' it). This is also a possible advantage in terms of memory charge. Consider the difference between explaining a movement through verbal or figural instructions: Go right, stop at x, pick up, ….; and explaining a movement via its 'perception'. This information has also the very nice property of not being discrete, digitalized-which might be crucial. Suppose that you have to move a heavy table together with somebody else: You will use the table itself as a coordination device, and the physical sensations as messages to him: He will 'feel' the direction and the speed that you want to move the table at; he will feel the evolving equilibrium. Imagine on the contrary that you have to give verbal instructions to him/her. The messages might be not enough precise, fast, etc. This doesn't exclude that sometimes also movements might be ambiguous or vague (and we have to add some word or picture).

"Scripts" as tacit communication devices

Let now apply this to scripts. For example, by just entering a shop or going around looking at the merchandise, X is *communicating*: "I'm looking for a possible purchase", "I'm a potential client", etc. In fact the shop assistant is likely to

[q]There is a possible (but very difficult, rare, and risky) form of coordination that can do without any observation (perception), because it is based on 'preestablished armonies', i.e. perfectly predesigned, programmed, and syncronized movements of X and Y.

[r]A nice example in [54] where Buskens and Royakers present a general use of coordinating and cooperating through signaling intentions by acting (they even define "commitment" as an act that signal an intention): "…. starting to cross the road is the commitment signaling the intention of the pedestrian to go first without the consent of the car driver". In my view, the message is not only declarative (and commissive) but also a request or command to have the precedence; the pedestrian relies on the driver's tacit assent to this request, not simply on an automatic breaking the car.

[s]http://www.msci.memphis.edu/~franklin/coord.html

However, Franklin seems to miss the difference between 'no communication' and 'tacit/behavioral communication'.

[t]See our criticisms to CSCW systems in [48] especially section 4 "From saying to doing in action workflow".

say, "May I help you?" (Giving for granted "for what"). He presupposes that X is a client and that he is looking around for buying something. The shop assistant doesn't ask, "Who are you?", "Why did you come here?", "Why are you looking at our things?". This is a common ground (X also knows that Y assumes so), the basis for pragmatically adequate speech acts. In the same vein, the client doesn't ask, "Who are you?", "Why do you talk to me?", and so on. The shop assistant's location, behavior (and perhaps dressing) is a message suited for the client: "Yes, you are in a shop, and I'm the shop assistant". This behavioral communication is (at least partially) intentional.

Other script-based communicative behaviors are not necessarily understood and *intended* by the agent. For example, in a restaurant, by seating, X is signaling that he has accepted to eat there and to sit there; by taking a fork, X is communicating "I'm starting to eat with this fork"; by finishing all the food in the plate and mopping up the plate with bread, X is communicating "I have really enjoyed this food"; whereas leaving food in the plate signals "It was too much for me" or "not so good" (that's why the waiter can ask "Was not good? Didn't you like it?"). Sometimes, we know and intend to send these messages; sometimes we are not aware of them, and we just wait for the next move/step in the script execution. But we always read the other's behavior also in mental terms.

Therefore:

> On the one side, the knowledge of scripts provides meaning to the participants' behaviors, which become behavioral (intentional or functional) messages.

> On the other side, vice versa, the scripts work thanks to those 'messages' confirming or making explicit the role and intentions of the participants; also because scripts can contain alternatives and variants that must be chosen and recognized for an appropriate coordination.

No script performance, no interactive role-playing would be possible without behavioral implicit communication among the partners. To verbalize what we are doing and what we believe and intend would not only be costly, difficult, and redundant, but also frequently imprecise and inadequate.

BIC and some of its functions and meanings in human agents

We are so used to BIC that we do not realize how ubiquitous it is in social life and how many different meanings it can convey. It is useful to give an idea of these uses and meanings, first of all for better understanding the phenomenon; and second, because several of these uses can be exploited in HCI, in computer mediated H collaboration, in agent-agent interaction. BIC actions can convey *different meanings* and messages. Let's examine some of the most important for human social life (*also applicable to AI/agents).

From teaching to terrorism:

A) "I'm able" or "I'm willing"

The most frequent message conveyed by a normal behavior is very obvious (inferentially very simple) but remarkably relevant:

(as you can see) ***I'm able to*** *do action a, and/or* ***I'm willing to do;*** in fact, I intentionally did it (or I'm doing it).

There are several different uses of this crucial BIC message about our mind and skills.

Skills demonstration in learning, exams, and tests

When A is teaching something to B via examples, and observes B's behavior to see whether B has learned it or not, then B's performance can be not only aimed at producing a given practical result but is (also or mainly) aimed at showing the acquired abilities to A.

Also the teacher's behavior is often a BIC; its message is: "Look, this is how you should do that".

In general, if *showing* and *exhibiting* are intentional acts they are communication acts.

Warnings without words

A peculiar form of power display that deserves special attention is Mafia's "warning" behavior. The latter (be it destroying goods, or beating or killing people) is an actual act, and the harm it produces is an actual harm, but the real aim of this behavior is communicative. It is aimed at intimidating, terrifying through a specific threat and its implications: "I can do this again; I could do this to you; I'm powerful and ready to act; I can even do worst than this". This meaning - i.e., the threat implicit in the practical act - is what really matters and what induces the addressee (who is not necessarily the present victim) to give up his or her "rebellious" attitudes. The practical act is a display of power and dangerous intentions - a "message" to be "understood". Nations also behave in a similar way: Consider for example Sharon's repeated reaction to terrorist attacks in Israel, which is not only a revenge, but a message for the future: "Do this again and I will do my bombing again". This is a horrible way of communicating.

B) "I did it"

As already mentioned, finishing our food is often a message for the cook, or waiter, who wishes and expects exactly this: "I liked it".

Satisfying social commitments and obligations

Consider for example a psychiatric patient that shows to the nurse that he is drinking his drug as prescribed (see later on social order).

BIC in teamwork

Teamwork is strongly based on BIC. Consider a soccer game, where many actions and movements of the players are messages for some member of the same team, while their opponents view them only as meaningful signs. For the opponents too there are BIC messages: For instance, feints and counter-feints, i.e. simulated actions. Also the ball and its trajectory or position can be a *message* for the others: Stigmergic Communication through the ball.

C) Silence as communication

As everybody knows, silence can be very 'eloquent'. In general, doing nothing, abstaining from an action, *is* an action. Thus - as any behavior - it can be aimed at communicating via

BIC. The meanings of silence or inaction are innumerable, depending on the context and their underlying motivations; for example, agreement, or else indifference: "I'm not involved, I do not care", "I do not rebel", "I do not know", etc. The most important social use, however, is for 'tacit agreements', that by definition are BIC-based. Tacit agreement or consent ('Qui tacet consentire videtur') is the way social conventions and informal norms emerge [29,56]. It is opportune that we spend some more word on BIC and Social Order.

BIC basement of social order

BIC has a privileged role in social order, in establishing commitments, in negotiating rules, in monitoring correct behaviors, in enforcing laws, in letting spontaneously emerge conventions and rules of behaviors. If there is a 'Social Contract' at the basement of society this Social Contract has been established by BIC and is just tacitly signed and renewed.

A) Imitation-BIC as convention establishment and memetic agreement

Imitation (i.e. repeating another's behavior-the model) has several possible BIC uses-provided that Y (the model) can observe (be informed about) the imitative behavior of X.

We can consider at least the following communicative goals:

a) In learning-teaching via imitation: "I'm trying to do like you; please, check it: Is it correct?"

b) In convention establishment and propagation: "I am showing the same behavior you displayed, I accept (and spread) it as a convention; I conform to it". This is probably the first form of memetic propagation through communication: Y (the model) interprets X's imitation as an agreement on the appropriateness of the act. Consequently, Y expects that X will perform that behavior again and again it, at least in the same context and interaction.

c) In emulation and identification: "I'm trying to do like you because I want to behave and to be like you; you are my model, my ideal".

d) In membership: "I'm trying to do like you because I belong to the same group; I want to be accepted by you; I accept and conform to your uses (see - b)".

B) Fulfilling social commitments as BIC

This is another kind of demonstrative act. Here, rather than showing one's power and abilities, one is primarily interested in showing that(s) he has done an expected action. In other words, the performance of the act is also aimed at informing that that act has been performed! This is especially important when Y expects X's act because there is some obligation impinging on X, and Y is monitoring whether X is respecting a prohibition, or executing an order, or keeping a promise [57].

A further, second-order, meaning of the act can also be: "I'm a respectful person; I'm obedient; I'm trustworthy".

A *social-commitment* to do act *a*, for being really fulfilled, requires not only that the committing agent (X) performs the promised action, but also that the committed-to agent (Y) will know this. Thus, when X performs the act in order fulfill his commitment to Y, X also intends that Y knows this.

Notice that what is important for exchange relationships or for social conformity is not that X really performed the act, but that Y (or the group) believes so.

C) Obeying to norms as BIC

One of the functions of norm obedience is the confirmation of the norm itself, of the normative authority of the group, and of conformity in general. Thus, one of the functions of norm-compliant behaviors is that of informing others about norm obedience.

At least at the functional level, X's behavior is implicit behavioral communication.

others about his respect of norms, or he is worrying about social monitoring and sanctions or seeking for social approval, and he *wants the others see and realize that he is obeying the norms.*

X is either aware of this function and intends to inform others about his norm compliance, or worries about social monitoring and possible sanctions or is in search of social approval. In any case, his norm compliant behavior is also an intentional behavioral/implicit communication to others[u].

At the collective level, when one respects a norm (s)he pays some costs for the commons and immediately his/her mental attitude of norm addressee changes into the attitude of a norm issuer and controller: *X wants others to respect the norm, and pay their own costs to the commons.*

While doing so X is *re-issuing* the norm, and prescribing and expecting the same behavior from others [57]. Thus the meaning of X's act is twofold: On the one hand, "I obey, you have not to sanction me"; on the other hand, "Do as I do, norms must be respected".

*This kind of routinary and tacit maintenance of the social order should be relevant also for MAS and HCI.

*BIC and AI-agents

As we have shown, observation, and more specifically 'signification' (the capability to interpret and ascribe meaning to observed facts) is the basis of a crucial form of communication without words or special protocols. Efficient coordination - in humans but also in artificial agents - exploits or should exploit not only 'observation' but more precisely this form of silent communication, in which agent X relies on the fact that agent Y is observing her in order to let Y

[u]Of course, X can also *simulate* his respect of the norms, while secretly violates them.

In conformity to conventions the behavior is less intentional and more routinary and automatic; they are our habits, they do no require a conscious decision. Thus, although clearly there is an informative aim of this behavior (since the whole convention is based on mutual assumptions and expectation about the others' conformity), this aim is usually a 'function' of the behavior not an intention; it becomes an intention in cases that I want that people notice that I'm following that convention.

understand that p, i.e. for communicating that p to Y.

A BIC theory looks relevant in several IT application domains, especially with autonomous agents, and in relation to several important issues such as:

-The problem of social order and social control in MAS, CSCW, virtual organizations.

It is fairly implausible that social order could be created and maintained mainly by explicit and formal norms, supported by a centralized control, formal monitoring, reporting and surveillance protocols, etc. Social order will mainly be self-organizing, spontaneous and informal, with spontaneous and decentralized forms of control and sanction [58]. In this perspective BIC will play a crucial role.

Punishments like socially excluding the norm-breakers will be BIC messages; the same for the act of leaving the group. Monitoring others' behavior, fulfilling commitments, obeying norms, etc. will all be BIC acts. We should design systems able to behave in this way; for example, capable not only of sending messages but also of observing other agents' behaviors or their results and traces. In this perspective a theory of 'observability' and the appropriate design of the 'environment' and of 'coordination artifacts' become crucial [52,43].

- Friendly and natural HM interaction

Collaboration and initiative (for example the so called 'Over-Help' [21]) is based upon the possibility of observing and understanding what one is trying to do, and in anticipating or correcting it. The same holds in Human-Robot interaction [48,49,59,60]. It is not simply a matter of specialized and explicit messages (words or gestures); this seems rather unnatural. Also expressive NVC signals (faces, emotions - see [61]) are not enough. A robot's coordination with humans in a physical environment requires that it should be provided with the ability to interpret human movements, understand them and react appropriately [62]. At that point a human action performed in presence of the robot will be performed also as a BIC message. Analogously the human agent should be in condition to monitor what the robot is doing and to intervene appropriately by regulating its degree of autonomy. At this point the robot's behaviors might become messages for asking approval, help, coordination, and so on.

In other words we are claiming that the explicit message-sending paradigm dominating CSCW, MAS, HCI, and H-Robot-I is inadequate, unnatural for humans and disturbing: Communication and mutual understanding (mind reading) should be less explicit, less costly, and more spontaneous and self-organizing.

Concluding Remarks

With regard to agents and (Ro)Bots, a nice conclusion of the above is offered by a contemporary author:

"It is not enough that they do what he (the master) orders, but they must understand what he desires, and frequently, to satisfy him, they must even anticipate his thoughts. For them it is not enough to obey him, they have to please him ..... They have to put attention to his words, to his voice and his eyes ... completely outstretched to catch his wants and to guess his thoughts. .....Can you call this 'life'!" (E'tienne De La Boe'tie "Discours de la servitude volontaire", 1595) Poor robots!

Social interaction among humans and between AI agents and humans systematically requires the ascription of mental states to others. There are several different mechanisms for this, which do not necessarily imply costly reasoning processes of inductive or abductive plan/intention recognition. There are a lot of much simpler, cheaper, automatic mechanisms for attributing beliefs, goals, etc. to others. A very important one is the "cognitive" filling in of scripts and roles: They are not only behavioral sequences, but imply/require/prescribe specific mental attitudes (I give you some money because "I want to pay; in order to pay" for some good; I enter and sit down at a restaurant because "I want to eat" and "I believe that this is a restaurant and it is open"). There are - we know - also other rules, heuristics, and mechanisms (including the very well studied empathy, simulation, etc.) but less relevant - for the moment - for artificial creatures.

We also argued that sometimes those mental attitudes are either unconscious or implicit (that is potential or tacit, not activated), or just the non-intended or non-understood *function* of the behavior/role. This fiction works.

Reading others' behavior and ascribing mental attitudes to them is the basis of a fundamental form of communication: Behavioral Implicit Communication. That is, others' practical behavior informs us about their minds, but it is also aimed at (either intentionally or functionally) *communicating* such mental attitudes. Social coordination is deeply grounded on this kind of behavioral communication (nothing to do with expressive behaviors and gestures - see note 15 and § 9.1).

Reflexive sociality: Reading our behavior and ascribe minds to ourselves

Moreover, humans not only try to read the minds of other people, but they also read their own behavior in terms of mental states. Our behavior is a "sign" and even a "message" (BIC) that we send to ourselves, and contains information similar to the messages sent to others: "I'm able", "I have done it", "I'm determined and persistent", "This is my duty", "I did punish myself", etc. Actually, this capability will reasonably be necessary also in intelligent artificial creatures.

The self-ascription of mental states (beliefs, goals, intentions, emotions, feelings,..) is a crucial part of meta-cognition. Meta-cognitive representations are not only means for predicting and explaining our behavior, and maintaining coherence and identity. We also use them for influencing ourselves, for self-imposing goals, for persuading or obliging ourselves to do or not to do something. Willpower is precisely this. But for applying influencing instruments to ourselves we need to represent and intentionally modify our minds.

*A question to be answered is: Should AI intelligent (Ro) bots have just intentions or also "will"? Would (and could) they be able to "consciously" influence themselves? [3].

In the same vein, coordination is also needed *with ourselves*, especially when complex actions and plans are to be developed. For instance, long-term temporal coordination requires that we perform a part of the plan now, and another part next month or year. In this kind of self-coordination, we have to trust and count on our 'future self'. We have to believe (or better 'expect') that we will keep in mind that commitment, and that we will be coherent and willing to do the next part of the plan, without wasting our resources and efforts. Not only we have such expectations on our future mental states but we try to influence our future self and his preferences. For instance, buying a flight ticket, paying the registration to a school, etc. will create pressures to maintain one's own commitments and achieve 'our' objectives. We also use other tricks, like oaths, promises, threats to ourselves etc. Thus mind ascription and writing (influence) is crucial also in coordination with us in long-term plans (typically human).

*Will AI agents have the same necessity? Why should they change their preferences about future commitments? A reason is that they might receive new orders or pressures from new stimuli and data, which would jeopardize their previous commitment. Therefore, they could also have a need for coordination with themselves, so as to maintain coherence and stable preferences in time.

Along our contribution we have signaled with a * all the local remarks about the possible interest or application of given models and human social mechanisms to AI Agents, (Ro) Bots, HCI and HRI. We will not recapitulate all of them here, but - as said - n our view "mind-reading" and "ascription" will be fundamental for socially interacting AI agents, with all its different functions: Prediction, explanation, trust and diffidence, coordination, communication, influencing and manipulating, deceiving. But a clarification is needed.

Anthropomorphism or operationalization of mental stuff?

*What AI is doing and has to do is building an artificial and hybrid society, based on human and artificial sociality, requiring social agents, that is "social minds" in relation with each other. Philosophers frequently claim that what AI and cognitive scientists are doing is to "anthropomorphize" machines that cannot in principle really have "mind", "intelligence", "intentions", etc. but just "simulate" them. (On this debate see e.g., [63]).

On the one side in fact we (have to) anthropomorphize machines, not just in the sense of (naively) perceive them as humans, as thinking and feeling something, but in the sense of really *making* them like us, not only simulating us (pretense).

However, our main disagreement with such view of what is happening with AI is more radical. It is exactly the other way around:

What we are doing is to "de-anthropomorphize" such concepts (intelligence, mind, intention..), making them no longer "anthropocentric" but more general and abstract, and more clear, formalized, and "operationalized". No longer common-sense "words".

These functions will no longer be only of humans or of living creatures.

*AI mission isn't that of uncritically importing concepts and theories from human and social sciences and philosophy for making practical/technological "applications". It gives back a crucial "scientific" contribution, not just "technological", by changing those concepts, models, and theories.

*The theater of sociality with our (ro)bots

The second things AI can do with those mental concepts it is realizing what humans do with them: As conventional presupposition, common background; by "assuming" (more than "believing/knowing" in the strict sense). That is, by acting "as if" those assumptions were true (verified); they will "count as" (§ 7).

A hybrid society cannot renounce this crucial *foundation* of human society: The extension of children's "symbolic game" to adults. Pretend together with others, act "as if" this puppet dog was a real dog or that cooker was really aglow, or this piece of paper was real "coin/gold". It is not important if you believe (deception) or simply you act "as if" you were interacting with a person, not a computer (Turing' test!); it works, let's do.

This grounding of sociality on pretense and shared conventional assumptions is the deep meaning of Turing's test, not a good *imitation*, not deception[v], but the "count as". You may believe that the computer you are interacting with is a woman (or the other way around: That the woman is a computer!): It doesn't matter; it works. "Pretending" is not necessarily deceiving, it also is to "stage": Actors and the public [64].

"I hold the world but as the world, a stage where every man must play a part" (William Shakespeare).

Acknowledgments

As for the general theses and arguments of the papers I have to thank our research group (Rino Falcone, Maria Miceli, Fabio Paglieri, Luca Tummolini, Giovanni Pezzulo, Giulia Andrighetto, Isabella Poggi, Emiliano Lorini); only in this critical context it was possible to build our general cognitive modeling of sociality (norms, trust, coordination, power, coop-

[v] "*A False Conclusion: Where Artificial Intelligence Really Lies*. Let us conclude just insinuating that deception is a nice vocation for AI; it is a return to its origin and deep challenges. We just mentioned believable agents, but, after all, isn't "believability" just a form of deception? The more believable the more deceptive? Moreover, in remembering the so called "Turing's test" we could say that deception is the original, foundational challenge and proof for AI. In fact in the "game" imagined by Turing (1950) there are three people, a man (A), a woman (B), and a interrogator (C) - either a man or a woman; C is closed in a room and communicates with A and B by a telewriter, and his/her goal is to determine which of the two is male and which is female. The goal of A in the game is to deceive C and induce him/her to a wrong identification; while the goal of B is to help C find the truth. Both A and B have to answer C's questions. As Turing notices, B might also use such expressions as "Don't believe him! I am the woman", but this will not help very much since A too could say something like this" [34].

eration, lie, surprise, emotions..) and to transfer it in Agent and MAS domain. As for BIC theory I would like to thank in particular Frascesca Giardini for her contribution. I thank Luca Tummolini, Giovanni Pezzulo, and also Andrea Omici, A., Ricci, M., Viroli with whom we work on 'coordination artifacts' for Agents. BIC research was partially funded by the European Project Eurocores -OMLL "The Origin of Man, Language and Languages". I would like also to thank an anonymous reviewer of this journal for his/her precious remarks. Special thanks to Maria Miceli for her generous help in revising this text.

References

1. M Slors, C Macdonald (2008) Rethinking folk-psychology: Alternatives to theories of mind. Philosophical Explorations 11: 153-161.
2. SV Albrecht, P Stone (2018) Autonomous agents modelling other agents: A comprehensive survey and open problems. Artificial Intelligence, 66-95.
3. F Paglieri, C Castelfranchi (2008) Cambiare la mente: Mindreading, azione intenzionale e coscienza". Sistemi Intelligenti 20: 489-520.
4. C Castelfranchi (2011) Ascribing Minds. Cognitive Processing 13: 415-425.
5. C Castelfranchi (2005) ToM & BIC Intentional behavioral communication as based on the theory of mind. Proc. AISB Symposium on Social Virtual Agents 37.
6. DC Dennett (1987) The Intentional Stance. The MIT Press.
7. JA Fodor (1990) A theory of content and other essays. The MIT Press.
8. Krebs J, Dawkins R (1984) Animal signals: Mindreading and manipulation. In: J Krebs, N Davies, Behavioural ecology: An evolutionary approach. (2nd edn), Oxford, Blackwell, 380-402.
9. Whiten, R Byrne (1988) Tactical deception in primates. Behavioral and Brain Sciences 11: 233-273.
10. S Baron-Cohen (1999) The evolution of a theory of mind. In: M Corballis, S Lea, The descent of mind: Psychological perspectives on hominid evolution. (edn), Oxford, Oxford University Press, 261-277.
11. D Sperber (2001) An evolutionary perspective on testimony and argumentation. Philosophical Topics 29: 401-413.
12. R Sun (2002) Duality of the mind: A bottom up approach toward cognition. Lawrence Erlbaum Associates.
13. S Baron-Cohen (1995) Mindblindness: An essay on autism and theory of mind, Cambridge, Mass, MIT Press.
14. M Tomasello (1999) The cultural origins of human cognition. Cambridge, Harvard University Press.
15. D Hutto (2004) The limits of spectatorial folk psychology. Mind and Language 19: 548-573.
16. A Rosas (2004) Mind reading, deception and the evolution of Kantian moral agents. Journal for the Theory of Social Behaviour 34: 127-139.
17. G Csibra, G Gergely (2007) 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol (Amst) 124: 60-78.
18. Amos T, Daniel K (1982) Judgment under uncertainty: Heuristics and biases. Cambridge University Press, New York.
19. T Gilovich, D Griffin, D Kahneman (2002) Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press. New York.
20. RE Petty, JT Caccioppo (1986) Communication and persuasion. Central and peripheral routes to attitude change. Springer, New York.
21. C Castelfranchi (1998) Modelling social action for AI Agents. Artificial Intelligence 103: 157-182.
22. C Castelfranchi, R Falcone (2010) Trust theory. A socio-cognitive and computational model. John Wiley & Sons, Chichester, UK.
23. HL Maibom (2010) Making decisions in a social world. Proceedings of ILCLI International Workshop on Cognitive Science, University of the Basque Country Press, San Sebastian.
24. J Mc Breen, G Di Tosto, F Dignum, et al. (2011) Linking Norms and Culture. Second International Conference on Culture and Computing.
25. RP Abelson (1976) Script processing in attitude formation and decision making. In: JS Carroll, JWPayne, Cognition and social behavior. (edn), Lawrence Erlbaum, Oxford, England, 33-45.
26. RC Schank, RP Abelson (1977) Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Lawrence Erlbaum, Oxford, England.
27. JA Bargh, PM Gollwitzer, A Lee-Chai, et al. (2001) The automated will: Unconscious activation and pursuit of behavioral goals. J Pers Soc Psychol 8: 1014-1027.
28. M Miceli, C Castelfranchi (2014) Expectancy and Emotion. Oxford University Press.
29. C Castelfranchi, F Giardini, E Lorini, et al. (2003) The prescriptive destiny of predictive attitude: From expectations to norms via conventions. Proc of Cognitive Science Conference.
30. H Garfinkel (1963) A conception of, and experiments with, 'trust' as a condition of stable concerted actions. In: OJ Harvey, Motivation and social interaction. (edn), The Ronald Press, New York, 187-238.
31. N Luhman (1976) Trust and Power. Wiley, Chichester.
32. C Castelfranchi (1999) Prescribed mental attitudes in goal-Adoption and norm-adoption. Artificial Intelligence and Law 7: 37-50.
33. RL Trivers (1971)The evolution of reciprocal altruism. The Quarterly Review of Biology 46: 35-57.
34. C Castelfranchi (2000) Artificial liars: Why computers will (necessarily) deceive us and each other. Ethics and Information Technology 2: 113-119.
35. C Castelfranchi, YH Tan (2001) The role of trust and deception in virtual societies. International Journal of Electronic Commerce 6: 55-70.
36. HH Clark, SA Brennan (1991) Grounding in communication. In: LB Resnick, JM Levine, SD Teasley, Perspectives on socially shared cognition. (edn), American Psychological Association, Washington, DC, USA, 127-149.
37. Y Wilks, J Barnden, A Ballim (1991) Belief ascription, metaphor and intensional identity. Cognitive Science 15: 133-171.
38. R Conte, C Castelfranchi (1995) Cognitive and social action. UCL Press, London.
39. C Castelfranchi, F Paglieri (2007) The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. Synthese 155: 237-263.

40. C Castelfranchi (2000) Through the agents' minds: Cognitive mediators of social action. Mind and Society 1: 109-140.

41. C Castelfranchi (2001) The theory of social functions: Challenges for multi-agent-based social simulation and multi-agent learning. Journal of Cognitive Systems Research 2: 5-38.

42. JA Bargh, E Morsella (2008) The unconscious mind. Perspect Psychol Sci 3: 73-79.

43. L Tummolini, C Castelfranchi, A Ricci, et al. (2005) "Exhibitionists" and "voyeurs" do it better: A shared environment approach for flexible coordination with tacit messages. Lecture notes in computer science 3374: 215-231.

44. B Nooteboom (1996) Towards a cognitive theory of the firm: Issues and a logic of change. IDEAS.

45. P Uwe (2019) The complementarity of mindshaping and mindreading. Phenomenology and Cognitive Science 18: 533-549.

46. Zawidzki T (2013) Mindshaping: A new framework for understanding human social cognition. MIT Press, Cambridge.

47. G Pezzulo, F Donnarumma, H Dindo, et al. (2018) The body talks: Sensorimotor communication and its brain and kinematic signatures. Phys Life Rev 28: 1-21.

48. F Giardini, C Castelfranchi (2003) Will Ever Humans and Robots Coordinate Each Other Via Behavioral Communication? EPSRC-Workshop.

49. T Fong, I Nourbakhsh, K Dautenhahn (2003) "A survey of socially interactive robots. Robotics and Autonomous Systems 42: 143-166.

50. H Holland, C Melhuish (1999) Stigmergy, Self-Organization, and Sorting in Collective Robotics. Artif Life 5: 173-202.

51. R Beckers, OE Holland, JL Deneuborg (1996) From local to global tasks: Stigmergy and collective robotics. In: Rodney A. Brooks and Pattie Maes. (edn), Artificial Life IV, 181-189.

52. Omicini A, Ricci A, Viroli M, et al. (2004) Coordination artifacts: Environment-based Coordination for Intelligent Agents.

53. L Tummolini, C Castelfranchi (2007) Trace signals: The meanings of stigmergy. International Workshop on Environments for Multi-Agent Systems, 141-156.

54. VW Buskens, LMM Royakkers (2002) Commitment: A game-theoretical and logical perspective. Cognitive Science Quarterly 2: 448-467.

55. O Boissier, J Padget, V Dignum, et al. (2005) Coordination, organizations, institutions, and norms in multi-agent systems". AAMAS: International Conference on Autonomous Agents and Multiagent Systems, 25-26.

56. C Castelfranchi (2006) When doing is saying. Implicit Behavioral communication and the foundation of collective activity. Invited talk. Collective Intentionality, Helsinki.

57. G Andrighetto, C Castelfranchi (2013) Norm compliance. The prescriptive power of normative actions. Franco Angeli 2: 120-135.

58. R Falcone, C Castelfranchi (2001) The human in the loop of a delegated agent: The theory of adjustable social autonomy. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 31: 406-418.

59. T Kanda, H Ishiguro, T Ono, et al. (2002) Development and evaluation of an interactive humanoid robot "Robovie". IEEE International Conference on Robotics and Automation 1848-1855.

60. F Giardini, C Castelfranchi (2004) Behavioural implicit communication for human-robot interaction. AAAI Fall Symposium Series, 21-24 ottobre, Arlington, Virginia (USA), Technical Report FS-04-05, AAAI Press, 91-96.

61. C Breazeal (1998) A motivational system for regulating human-robot interaction. Proceedings of AAAI98, Madison, WI.

62. C Nehaniv, K Dautenhahn (2000) Imitation in animals and artifacts, MIT Press, Cambridge, Mass.

63. Lfloridi, JW Sanders (2004) On the morality of artificial agents. Minds and Machines 14: 349-379.

64. E Goffman (1956) The presentation of self in everyday life.

DOI: 10.36959/447/345