

ToM & BIC

Intentional Behavioral Communication as Based on Theory of Mind

Cristiano Castelfranchi
Institute for Cognitive Sciences and Technologies – CNR - Roma
c.castelfranchi@istc.cnr.it

Abstract

In this paper we will analyze unconventional (unspecialized) behavioral implicit communication (BIC) and its relation with ToM, because our claim is that BIC is the most *basic* form of communication from the analytical point of view, and also the most *primitive* (both in evolutionary and in developmental sense). BIC plays an irreplaceable and underestimated role in human interaction and coordination, social order, cultural transmission, and we do expect an important role of BIC in social Agents, robot-robot coordination, and in H-Agent and H-Robot interaction. We will first define BIC making clear the fundamental distinction between *signification* and *communication* and also explaining why its is false that all behaviors in social contexts are communication; then it is explained why BIC has nothing to do with gestures and expressive movements (the so called Non-Verbal-Communication); then we will characterize the ‘transition’ steps from non-communicative behavior to intentional BIC; eventually a few examples of how crucial BIC is in human coordination and interaction will be provided. In doing so it will clearly emerge why BIC is bilaterally based on ToM, and how human capacity for mind-reading has been a cognitive prerequisite for *intentional communication*.

1 Introduction

Intentional Communication is definitely based on *Theory of Mind* (ToM) in both the sender and the addressee’ perspectives. In fact, the sender X while *intending to communicate intends that* the other ‘understands’, i.e. captures the ‘meaning’ of the message. Moreover, she has a representation of the mind of the addressee Y

- as ignoring the content of the message or not already having the intention that X wants to promote/elicit in him;¹
- as able to infer the intended meaning; and possibly even
- as able to recognize the sender’s intention to communicate (i.e. having a ToM of the sender).

Intentional Communication is the intentional modification of the *mental states* of the addressee (beliefs and possibly goals).

On the other side, Y - in conventional communication (for example in linguistic communication)- recognizes X’s intention to communicate and tries to capture the ‘intended’ meaning (on this ‘cooperative’ goal see §2.3). In non-conventional communication Y ‘reads’ X’s behavior also in intentional terms in order to understand what she is actually doing and why, and what to expect for anticipating X’s behavior.

In this paper we will precisely analyze unconventional (unspecialized) behavioral implicit communication (BIC) and its relation with ToM, because our claim is that BIC is the most *basic* form of communication from the analytical point of view,

and also the most *primitive* (both in evolutionary and in developmental sense) (Castelfranchi, 2004a). BIC plays an irreplaceable and underestimated role in human interaction and coordination, social order, cultural transmission, and we do expect an important role of BIC in social Agents and robot-robot coordination (Omicini et al., 2004) and in H-Agent and H-Robot interaction (Giardini et al., 2004). We will first define BIC making clear the fundamental distinction between *signification* and *communication* and also explaining why its is false that all behaviors in social contexts are communication; then it is explained why BIC has nothing to do with gestures and expressive movements (the so called Non-Verbal-Communication); then we will characterize the ‘transition’ steps from non-communicative behavior to intentional BIC and to its overcoming in ‘simulation’ and ‘ritualization’; eventually a few examples of how crucial BIC is in human coordination and interaction will be provided.

In doing so it will clearly emerge why BIC is bilaterally based on ToM, and how human capacity for mind-reading has been a cognitive prerequisite for *intentional communication* (and one of the outcome of the evolutionary pressure for intentional communication), although intentional communication should not be identified with ‘the communication of the intention to communicate’ (see for ex. the important discussion on <http://www.interdisciplines.org/coevolution>). Agents lacking this capacity (mind-reading, intention/plan recognition, beliefs and intentions about the other’s mental states) will never be capable of this fundamental form of communication and will be confined to more primitive BIC forms just based on evolutionary selection or reinforcement learning and reactive behaviors. However, BIC is much more than

¹ More precisely, A does not assume that B already knows or already intends the object of her message.

mere plan recognition (§ 6).

2 Behavioral Implicit Communication Theory

Usual, practical, even non-social behaviors can *contextually be used as messages* for communicating. Behavior can be communication without any modification or any additional signal or mark. We will call this form of communication *without specialized symbols: Behavioral - Implicit Communication* (BIC).

“Behavioral” because it is just simple non-codified behavior.

“Implicit” because – not being specialized and codified – its communicative character is unmarked, undisclosed, not manifest, and thus deniable. Normally communication actions are on the contrary special and specialized behaviors (like speech acts, gestures, signals, ...).

BIC is a very important notion, never clearly focused, and very frequently mixed up with other forms of communication (typically the so called “non-verbal” or “expressive” or “extra-linguistic” or “visual” communication). It has been source of a number of misunderstandings and bad definitions. This ill-treated notion is crucial for the whole theory of social behavior: coordination, control, social order creation, norms keeping, identity and membership recognition, social conventions building, cultural transmission, deception, etc. A lot of social control and collaboration monitoring and coordination, are in fact based on this form of communication and not on special and explicit messages (communication protocols).

2.1 Against Watclawicz: *Are we damned to communicate?*

A famous thesis of Palo Alto psychotherapy school was that: *"It is impossible do not communicate", ... "any behavior is communication"* in social domain (Watzlawich et al., 1967).

In this view, a non-communicative behavior is nonsense.

This claim is too strong. It gives us a notion of communication that is useless because is non-discriminative. Is simple understanding already communication? Is it possible to clarify when behavior is communication and when is not?

In order to have communication having a "recipient" who attributes some meaning to a certain *sign* is a non-sufficient condition.

We cannot consider as communication any information/*sign* arriving from X to Y, unless it is *aimed at* informing Y. A *teleological* (intentional or functional) "sending" action by the source is needed. The source has to perform a given behavior "**in order**" the other agent interprets it in a certain way, receives the “message” and its meaning.

Is, for example, an escaping prey “communicating” to its predator/enemy its position and move?

Watzlawich’s overgeneralization cannot avoid considering *communication* to the enemy the fact that a predator can observe the movement of the prey. Although this information is certainly very relevant and informative for the enemy or predator, it is not communication. Receiving the information is *functional (adaptive) for the predator* and for that species which have developed such ability, but it is not functional at all, is not adaptive for the prey. Thus “sending” that sign is not a functional (evolutionary) goal of the prey, that is what matters for having communication.

Analogously, is a pilferer informing or communicating to the guard about his presence and moves? The pilferer does not notice that there is a working TV camera surveillance system and thus *he does not know* that there is a guard that is following him on a screen! Or when a pilferer while escaping from the police is leaving on the ground prints and traces of his direction, are those signs (very meaningful for the police) messages to it?

We should not mix up mere “Signification” with “Communication”. Following Eco (1973) prints on the ground are *signs* for the hunter of the passage of a deer; smoke is the *sign* of a fire; some spots can mean "it is raining" (they are for Y *signs* of the fact that it is raining). We have here simple processes of *signification*.

Notice that *meanings are not conventional* but simply based upon natural perceptual experience and inference. Notice also that *the signal, the vehicle has not been manufactured on purpose for conveying this meaning*, it doesn’t need to be “encoded” and “decoded” via some conventional artificial rule.

The **definition of BIC at the intentional level** (in this paper we will just analyze *intentional* BIC) is as follows:

in BIC the agent (source) is performing a usual practical action α , but he also knows and lets or makes the other agent (addressee) to observe and understand such a behavior α , i.e. to capture some meaning μ from that “message”, because this is part of his (motivating or non motivating) goals in performing α

In sum, BIC is a practical action primarily aimed to reach a practical goal, which is *also* aimed at achieving a communicative goal, without any predetermined (conventional or innate) specialized meaning.

2.2 Why BIC is *not* “non-verbal”, “extra-linguistic” communication

BIC is not the same and has not very much to do with the so-called *non-verbal or extra-linguistic communication* (NVC) although NVC is through some behavior or behavioral features, and BIC is for sure non-verbal and extra-linguistic. The few of BIC that has been identified has been actually mixed up with the never well-defined notion of “Non Verbal Behavior” (ex. Porter, 1969).

Non-verbal and extra-linguistic communication

refers to specific and specialized communication systems and *codes* based on facial expressions and postures, specific gestures, super-segmental features of voice (intonation, pitch, etc.), etc. that communicate specific meanings by *specialized, recognizable signals* (either conventional ex. policeman regulating traffic; or universal ex. emotional signals). BIC on the contrary is not a “language”. Any (verbal or non-verbal) “language” has some sort of “lexicon” i.e. a list of (learned or inborn) perceptual patterns *specialized* as “signs” (Givens, 2003): where “specialized” means either conventional and learned as sign, or built in, designed just for such a purpose (function) by natural selection, or engineering. BIC does not require a specific learning or training, or transmission; it simply exploits perceptual patterns of usual behavior and their recognition. BIC is an observation-based, non-special-message-based, unconventional communication, exploiting simple side effects of acts and the natural disposition of agents to observe and interpret the behavior of the interfering others. BIC gestures are just gestures, they are not symbolic but practical: to drink, to walk, to scratch oneself, to chew. They represent and mean themselves and what is *unconventionally* inferable from them (like the agent’s intentions and beliefs).

2.3 Intentional Behavioral Communication step by step

There are several steps in the evolution from mere practical behavior to BIC and to a conventional sign. Let’s examine this transition.

i) Just behavior: An agent X is acting in a desert word; no other agent or intelligent creature is there, nobody observes, understands or ascribes any meaning to this behavior α .²

Neither “signification” nor -a fortiori - “communication” are there.

ii) Signification: An agent X is acting by its own in a word but there is another agent Y observing it which ascribes some ‘meaning’ μ to this behavior α . There is in this case “signification” (X’s behavior has some meaning for Y, informs Y “that p”), but there is no necessarily “communication”.

By “signification” we mean that the behavior of X is a *sign* of something, means something else for Y. For example: p can be = to “X is moving”, “X is eating”, “X is going there”.

As we know to have communication the signification effect must be on purpose; but this presupposes that X is aware of it. Thus in (ii) we have two possible circumstances:

ii_a) X does not know

Consider the pilferer example where he is not aware of being monitored.

ii_b) X’s awareness: “weak BIC”

Consider now that X knows about being monitored

by a guard, but that he does not care at all of it, because he knows that the guard cannot do anything at all.

Y’s understanding is here among *the known but unintended effects* of X’s behavior. Although perhaps being an ‘anticipated result’ of the action it is not intended by the agent. Not only indifferent or negative expected results can be non-motivating, non-intended, but also positive (goal-realizing) expected results can be non-intended in the sense of “non motivating the action”, neither sufficient nor necessary for the action. In our example the pilferer might be happy and laughing about the guard being alerted and powerless and angry.

iii) True or strong BIC

The fact that Y knows that p is “co-motivating” the action of X.

The behavior is both a practical action for pragmatic ends (breaking the door and entering, etc.) and a “message”.

We call this “strong or true behavioral communication”, the pragmatic behavior which maintains its motivation and functionality acquires an additional purpose: *to let/make the other know/understand that p.*

The important point for fully understanding BIC (and the difference with the following *meta-BIC*) is that: we have here a fully *intentional* communication act, but without the aim (intention) that the other understands that X *intends to communicate* (by this act). ‘Intention of communicating’ and ‘communicating (this) intention’ are not one and the same thing. Given the well-consolidated (and fundamental) Grice-inspired view of *linguistic* communication - that frequently is generalized to the notion of ‘communication’ itself - these two different things are usually mixed up, and it is difficult to disentangle them; but they are clearly different both at the logical and at the practical level.

With a BIC message X *intends* that the other recognizes her action, and perhaps that recognizes and understands her *practical intention* motivating the action (eating; having the door closed; knowing what time is it; etc.), but X has not necessarily (at this communicative stage) the *intention* that the other realizes her higher-*intention* that Y understand this, that is her *intention to communicate* something to Y through that practical action: I want that Y understands that I intend to go, but not that I intend that he understands that I intend to go.

It is now clear how intentional BIC is bilaterally – that is on both sides – based on ToM:

First, it presupposes Y’s ability of ‘reading’ X’s behavior; the most primitive level is the mere recognition of the movement, a more advanced level is the recognition of the ‘goal’ of the action. Mirror neurons seem able to provide this faculty to primates (Rizzolati et al., 1996; 2001; Arbib, 2003). More advanced forms entail the recognition of the higher intentions, motives, and beliefs of the agent. In other words BIC presupposes that Y has a representation of X’s mind.

² Although sometimes we use BIC and stigmergic messages with ourselves.

Second, BIC presupposes that X realizes Y's understanding of her goals or intentions and beliefs; that is that Y's has a theory of X's mind; and this implies X's representation of Y's mind (additional considerations on § 5).

iv) Meta-BIC

In meta-BIC, there is a meta-communication, typical of higher forms of communication like language. BIC meta-message is as follows: "*this is communication, this is a message not just behavior; it is aimed at informing you*".

Frequently BIC has such a high level (Grice's way) nature. For example the act of *giving* or *handing* is not only a practical one, but is a meta-communicative act where X intends that Y understands that she is putting something closer to Y *in order* Y (understanding that she intends so) takes it.

v) Beyond BIC: actions for communication only

The behavior α is intended and performed by X *only* for its meaning μ , only for making Y believe that p. There are no longer practical purposes. The act is usually performed either out of its practical context or in an incomplete and ineffective way.

v_a) Simulation

Notice that in the pilferer's scenario, that fact that the \square has *only a communicative goal* means that it is a **fake action!** In fact, if α has no other goals apart from communicating to Y, Y will be deceived, and the information he will derive from observing α will be false (and α is precisely aimed at this result). It is just a bluff.

v_b) Ritualization

The practical effect becomes irrelevant: the behavior is ready for **ritualization**, especially if is not for deception but for explicit communication. Ritualization means that α can lose all its features that are no longer useful (while were pertinent for its pragmatic function) while preserving or emphasizing those features that are pertinent for its perception, recognition and signification. After *Ritualization* the behavior will obviously be a *specialized* communicative act, a specialized and artificial signal (generated by learning and conventions, or even selection). This is the ontogenetic and the evolutionary origin of several 'gestures' and 'expressive movements'.

3 Ubiquitous BIC

We are so used to BIC and it is such an implicit form of communication 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 - even risking to be a bit anecdotal -, first of all just for understanding the phenomenon, second, because several of these uses can be exploited in HCI, in computer mediated H collaboration, in Agent-Agent interaction.

BIC acts can convey quite different meanings and messages. Let's examine some on the most important

of them for human social life (also applicable to Agents)

3.1 "I'm able" or "I'm willing"

The most frequent message sent by a normal behavior is very obvious (inferentially very simple, given an intentional stance in the addressee) but incredibly relevant:

(as you can see) I'm able to do, and/or I'm willing to do; since I actually did it (I'm doing it) and on purpose.

There are several different uses of this crucial BIC message.

Skills demonstration in learning, examines, and tests

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

More in general, doing the same action α of a model, imitating, is the base for a possible tacit BIC message of X, for the possible use of the action α as a message to Y: "I'm doing the same". But for this specific additional conditions are needed:

i) X performs α (imitates Y)

ii) Y observes and recognizes (i), and forms the meaning μ "X is doing α /like me"

iii) X knows that (ii)

iv) X intends that (ii)

v) X performs α also in order (ii) (that is because of (iv) & (iii))

In this case α is a real (successful) *message* to Y.

When and why should X inform Y about imitating him? Especially when Y has the goal that (i).

Also the behavior of the teacher is a BIC; its message is: "look, this is how you should do". Usually this is also joined with expressive faces and gestures (and with words) but this is not the message we are focusing on.

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

Warnings without words

This is a peculiar use of exhibition of power that deserves special attention.

Mafia's "warning", monition. The act (say: burning, biting, destroying, killing) is a true act and the harm is a very true harm, but the real aim of this behavior (burning, killing, etc.) is communicative. It is aimed at intimidating, terrifying via a specific meaning or threat: "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 - the "promise" implicit in the practical act - is what really matter and what induces the addressee (that not necessarily is already the victim) to give up. The practical act is a

show down of power and intentions; a “message” to be “understood”.

The message is “if you do not learn, if you will do this again I will do even worst”.

The same do nations: consider for example the repeated reactions of Sharon after terrorist attacks in Israel; it is not only a revenge, it is a message: “do this again and I will do this (bombing) again”; the same holds for terrorist bombs. Perhaps it would be better communicating via words and diplomacy. This is a horrible way of communicating.

3.2 “I did it”, “I’m doing it”

This is another obvious possible meaning of any action, and it is used for many social messages based on the others’ expectations about our behavior.

For example, to finish your food can be a message your guest: “I finished it, I liked it”, as the guest wish and expects.

The satisfaction of social prescriptions

Consider for example a psychiatric patient that *shows* to the nurse that he is drinking his drug as prescribed. (See later on social order). In the next section we will spend some more word on BIC and Social Order, let us focus here on the message “I did it- I’m doing it” for tacit reciprocal coordination.

3.3 BIC for Coordination

In coordination it is not so important the fact that I intend to do (and keep my personal or social commitments – which is crucial in cooperation) or the fact that I’m able and skilled, it is more relevant communicating (informing) about *when, how, where I’m doing* my act/part in a shared environment where we interfere with each other, so that you can coordinate with my behavior while knowing time, location, shape, etc. (Castelfranchi, 1998; Castelfranchi 2004b).

Clearly in order to coordinated with a given event or act Ev X should perceive it or foresee it thanks to some perceptual hints, 'index' or sign. In other word usually it is an intrinsic necessity of Coordination activity that of observing and interpreting the word in which X is acting pursuing its goals, and in particular observing Ev.

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 - 'signification'.

A large part of Coordination activity (and social interaction) is not simply base on Observation and Signification but is BIC-based.

For example, clearly enough in **mutual coordination** not just Signification is needed but true BIC. 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. But let’s more systematically examine this.

In **unilateral Coordination**:

Non-BIC-based Unilateral: Y coordinates (adapts) his own behavior to the interfering behavior of X, who does not perceive at all or does not care at all of those (reciprocal) interferences. In this case X’s behavior is highly significant for Y (signification) but is not communication since X does not know or does not care of the fact that Y is observing her and interpreting her behavior.

BIC-based Unilateral: only Y coordinates (adapts) his own behavior to the interfering behavior of X, but X knows and intends this, although she does not want to coordinate her own action with the other. X’s behavior is communicative.

In **bilateral (symmetric-unilateral) Coordination**: both Y and X coordinate their own action on the actions of the others but they ignore or do not intend that the other do the same. Again there can be no communication at all, but if one of the agent acts also in order the other perceives and understands what s/he is doing, there is BIC.

In **mutual Coordination**: both X and Y wants the other to coordinates with his/her own behavior and understands that s/he intends to coordinate with the other's behavior. As we said, mutual coordination, based on symmetric intentions and mutual awareness (shared beliefs) entails and requires BIC: each coordination act (adaptation of the behavior) is a message to the other.

Let us draw some conclusions on this point.

Coordination is possible without any communication both in human and artificial societies (Castelfranchi, 1998; see also Franklin, <http://www.mscl.memphis.edu/~franklin/coord.html>³). This is an important statement against common sense. However, usually coordination exploits communication.

Since BIC is i) a very economic (parasitic), ii) a very spontaneous, iii) a very practice and rather effective form of communication just exploiting side effects of acts, traces, and the natural disposition of agents to observe and interpret the behavior of the interfering others, a rather important **prediction** follows.

One can expect that *agents acting and perceiving in a common world will use a lot of BIC and will spontaneously develop it.*

Actually a very large part of communication for coordination in situated and embodied agent exploits 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 behaviors that have been ritualized.*

This kind of observation-based, non-special-message-based communication should be much more exploited in CSCW and computer/net mediated

³ However, Franklin seems to miss the difference between ‘no communication’ and ‘tacit/behavioral communication’.

interaction, in Multi-robot coordination, in Human-robot coordination, in MA systems (see § 6).

4 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.

4.1 Fulfilling Social Commitments and Obeying Norms as BIC

This is another kind of demonstrative act, not basically aimed at showing power and abilities, or good disposition, but primarily intended to show that one have done the *expected* action. Thus the performance of the act is also aimed at informing that it has been performed! This is especially important when the expectation of X's act is based on *obligations* impinging on X, and Y is monitoring X's non-violation of his duty. Either X is respecting a prohibition, or executing an order, or keeping a promise.

A second order meaning of the act can also be: "I'm a respectful guy; I'm obedient; I'm trustworthy", but this inferential meaning is reached through the first meaning "I'm respecting, obeying, keeping promises".

A **Social-Commitment** of X to Y of doing the act, in order to be really (socially) fulfilled, requires not only that agent X performs the promised action α , but also that the agent Y knows this (Castelfranchi, 1995).

Thus, when X is performing the act in order to keep his promise and fulfill his commitment to Y, he also intends that Y knows this.

(If there are no explicit and specific messages) *any act of S-Commitment fulfillment is also an implicit communication act about that fulfillment.*

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.

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 obeying behaviors is that of informing the others about norm obedience. At least at the functional level X's behavior is *implicit behavioral communication*.

Frequently, X either is aware of this function and collaborates on this (thus he intends to inform the 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*. In both cases,

his conform behavior is also an *intentional* implicit/implicit communication to the others.

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

At the collective level, when I respect a norm I pay some costs for the commons and immediately I move from my mental attitude of norm addressee (which recognized and acknowledge the norm and its authority, and decided to conform to it) while adopting the mental set of the norm issuer and controller (Conte et al., 1995):

I want the others to respect the norm, pay their own costs and contribution to the commons.

While doing so I'm **reissuing** the norm, **prescribing** a behavior to the others and **checking** their behavior (expectation). Thus the meaning of my act is twofold: "I obey, you have not to sanction me"; "Do as I do, norms must be respected".

This kind of routine and tacit maintenance of social order is relevant also for MAS and HCI: doing what I promised or just passing the product of my activity to the other is a *message*; sending additional explicit messages is not necessary and usually is disturbing.

5 Reciprocal ToM between BIC-sender and receiver

Let us now focus on the relationships between *intentional* BIC and ToM as emerged from this analysis.

i_a. X's goal in sending the BIC message is that Y believes that X is doing action α ; but action α frequently enough is conceptually defined/characterized in an intentional way, that is by its purposive result (for example 'water' is not just dropping water on plants); moreover, X frequently intends that Y understands what X has in mind while doing α : her beliefs or goals.

ii_a. X assumes that Y does not already knows/believes the content of the message, and if the message is an 'imperative' does not already intend to do that action.

iii_a. in Meta-BIC X also plans that Y realizes that X *intends* to communicate and that Y understands the message.

Thus X has (and bases her message on) a rather complex ToM of Y, even a recursive one: "X wants/believes that Y believes that X wants/believes...."

On the side of the addressee, we have:

i_b. Y (even *before* BIC and as one of the conditions for its evolution) interprets X's behavior in mental terms: as due to given beliefs and goals. He reacts to these goals, intentions, and beliefs of X more than to X's actual behavior, especially for anticipatory coordination.

ii_b. Y is able to contextually interpret X's behavior as a *message*, i.e. as intentionally aimed at changing his own mental states ("X believes that I believe..... X intends that I believe.....").

Not only in and for BIC communication we have ToM on both sides, but we also have goals about the mind of the other and we arrive to cooperation on such goals. We may consider that in BIC there are two goals/functions meeting each other:

a) the communicator's goal: X's behavior has the goal or function that Y "understands", recognizes, and comes to believe that p (and this holds from step (iii) § 2.3)

b) the interpreter's goal: Y has the goal/function of interpreting X's behavior in order to give it a meaning (and this holds from step (ii) § 2.3)

However, those goals in the initial forms of BIC are simply *independent* from one the other. "Cooperation" is just accidental. X and Y do not really have a "common goal".

Since, in step (ii - *Signification*) X does not know that Y wants to understand her behavior; while in step (iii) Y does not know that X is communicating to him through it behavior α . Thus Y has not the goal of: "understanding what X means by α "; that is the real *common goal* of higher form of communication (like *linguistic communication*) on which usually X and Y cooperate for a successful communication.

In *meta-BIC* on the contrary Y knows that X is communicating. Therefore he has a special form of goal (**b**), the goal of catching what X is *communicating*:

b') goal of Y to understand what X's intends to communicate, to understand which is the meaning in X's mind.

The agents in such away arrive to *cooperate* in strict sense (like in linguistic exchange), and the two goals (a) and (b) become complementary, convergent and functional to each other; that is X and Y have the same goal and they *know* the goal of each other.

6 Are there BIC-Agents in our future?

Will Agents be able to read the other's (user or Agent) behaviors as 'action' i.e. in intentional terms (the aimed results)? Will they be able to recognize the intention and the plan (higher-goals) of the user or of the other agents?

Without this, how might they be able to *anticipate* the other behavior for an appropriate coordination; or to take the initiative in *helping*, or over-helping (Falcone et al., 2001) the other going beyond the literal request?

This basically exploits 'intention and plan recognition'. And there is a long tradition in AI and Agents studies on this. However, first plan-recognition is not enough for truly collaborative Agents; second, BIC is much more than 'plan-recognition'.

As for the latter issue, BIC is not only 'intention' or 'plan' recognition. It can imply the recognition of beliefs of X, of motives of X, of X's social status, etc.:

whatever (mental or non-mental) feature of X

gleams through or can be reasonably inferred by her act α can be signified by a BIC message.

Second, we are not speaking of simple plan *recognition* ('signification') but of plan *communication*, which is much more.

Let us now explain why plan-recognition is not enough and BIC goes further: If Agents will be able to anticipate and understand our intentions in doing \square , and if we realize this, clearly next time we will do \square *expecting* that the Agent recognizes what we intend to do and reacts appropriately. In other terms, our action will become a BIC message to the Agent, and later even just a 'gesture', just a ritualized gesture, a hint for 'ordering' to it to do something (like in child evolution of the 'pointing' gesture (Castelfranchi, 2004a)). Moreover we (or some other Agent) will know that the Agents (monitoring us) will recognize whether we are following or not norms and rules or keeping our commitments, and while doing so we will in fact send a message to them: "I'm doing as due!". So also social order will strongly be based on BIC messages among Agents (or Agents and users). And so on, as it is in human interaction.

Finally, Agents able to read BIC messages in human interaction might make them explicit and automatic in Computer-Mediated-Human-Interaction like CSCW. Several CSCW original systems were definitely boring and oppressive because they obliged people to unnatural forms of interaction. For example, after tacking a Commitment with Y, after doing what promised – possible on the same computer – I'm obliged to send an explicit message to Y to inform Y that "I did". In human collaborative work usually our action, or its product, or the transmission of the results *per se* is also a message "As you see I did". Agent might relieve the users from unnatural and boring practices like this, recognizing the tacit message and automatically sending an explicit message to Y.⁴

Acknowledgments

I would like to acknowledge the contribution of Francesca Giardini, Isabella Poggi, Luca Tummolini, Emiliano Lorini, Rino Falcone. Thanks for their critical comments, discussions, and common work on these subjects. This research has been developed within the Eurocores Project *OMLL* (Origin of Man, Language, and Languages) & the EC's 6th Framework Programme IST priority; Unit: Cognitive Systems - PROJECT: *Mind RACES* N°511931

References

Michael, A., Arbib, The evolving mirror system: A neural basis for language readiness. In: *Language*

⁴ This idea emerged during a nice discussion with Francis Brazier.

- Evolution* (Ed. by Christensen, M. & Kirby, S.), pp. 182-200. Oxford: Oxford University Press, 2003.
- Cristiano Castelfranchi, "Modelling Social Action for AI Agents", *Artificial Intelligence*, 103, pp. 157-182, 1998.
- Cristiano Castelfranchi, Commitments: From Individual Intentions to Groups and Organization. *ICMAS'95 First International Conference on MultiAgent Systems*, S. Francisco, AAAI, 1995.
- Cristiano Castelfranchi "When doing is saying: implicit communication before and without language and gestures". In "Evolving communication: from action to language An 'implicit vs. explicit' cognitive and pragmatic perspective"- International workshop Siena / ISTC-CNR/ OMLL May 13-14, 2004 Siena (Italy), 2004a.
http://www.unisi.it/ricerca/dip/fil_sc_soc/dot-sc/evcoimes.html
- Cristiano Castelfranchi SILENT AGENTS: From Observation to Tacit Communication. *AAMAS-Workshop on 'Observation'*- N.Y. July 2004b.
- Rosaria Conte, Cristiano Castelfranchi *Cognitive and Social Action*. UCL Press, London; Taylor & Francis, 1995
- Umberto Eco, *A Theory of Semiotics*. London, Indiana University Press, 1976
- Rino Falcone, Cristiano Castelfranchi. 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, Special Issue on "Socially Intelligent Agents - the Human in the Loop, Volume 31, Number 5, September 2001, pp. 406-418. 2001
- Francesca Giardini, Cristiano Castelfranchi, Behavior Implicit Communication for Human-Robot Interaction, *Proceedings of the AAAI Fall Symposium 2004 (The intersection of Cognitive Science and Robotics: From Interfaces to Intelligence)*, pp. 91-96, 2004.
- David, B. Givens "Nonverbal Dictionary of gestures, signs and body language cues". Spokane, Washington: Center for Nonverbal Studies Press, 2003
- Andrea Omicini, Alessandro Ricci, Marco Viroli, Cristiano Castelfranchi, Luca Tummolini (2004) "Coordination Artifacts: Environment-based Coordination for Intelligent Agents" , *AAMAS 2004 Conference Proceedings*.
- David Premack and Guy Woodruff Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences* 4:515-526, 1978
- Oscar, W. Porter, "Non Verbal Communication", *Training and Development Journal*, 23, pp.3-8, 1969.
- Giacomo Rizzolatti, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi, Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3: 131-141. 1996
- Giacomo Rizzolatti, Leonardo Fogassi, Vittorio Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience Reviews*, 2, 661-670. 2001
- Michael Tomasello, & J. Call, *Primate Cognition*. Oxford: Oxford University Press, 1997.
- Paul Watzlavich, Janet Helmick Beavin, Don D. Jackson,, *Pragmatic of human communication: a study of interactional patterns, pathologies, and paradoxes*, W.W. Norton &Co., Inc., New York, 1967.
- http://en.wikipedia.org/wiki/Theory_of_mind
<http://www.interdisciplines.org/coevolution>