PREPRINT: Gallagher, S. 2012. Social cognition, the Chinese room and the robot replies. In Zdravko Radman (ed.), *Knowing without Thinking: Mind, Action, Cognition and the Phenomenon of the Background.* London: Palgrave-Macmillan. (Please cite original publication).

Social cognition, the Chinese room, and the robot replies

Shaun Gallagher
Department of Philosophy
University of Memphis (USA)
School of Humanities
University of Hertfordshire (UK)

In philosophy of mind and related disciplines the standard conceptions of mind have been formulated in terms of a problem space that excludes certain solutions to problems defined in that space. I'll argue that this is the case in much of the recent discussion of social cognition, but also in earlier discussions of artificial intelligence. I'll try to show this by looking at versions of the frame problem – a problem that seems to fall into this solution-resistant space. To be precise, it is not that the frame problem itself has not been properly formulated, but rather that the ways various theorists think of the mind prevent certain solutions from coming into place. Even when a solution is on the horizon, it is often blocked from counting as a solution because our general conception of the mind has not been properly formulated.

I'll consider three problems that, I'll argue, have the same solution, namely an appeal to the concept of background. There are clear indications in the discussion of these problems that point to this solution; but things remain unresolved because the way these problems are laid out, namely, along internalist lines, prevents a proper appeal to the notion of background.

The starting problem

I'll begin with what I refer to as the 'starting problem', which is a version of the frame problem found in discussions of social cognition. I begin with this problem because its solution points in a clear way to solutions to two other problems: the frame problem in AI, and an unresolved problem with Searle's Chinese Room thought experiment.

The two standard approaches to the problem of social cognition are theory theory (TT) and simulation theory (ST). Both theories define the problem of social cognition as a problem of other minds. That is, they define it in terms of the lack of access that we have to another person's mind. The mind of the other, like my own mind, is internal to that person – a private collection of beliefs and desires in her head, which ultimately explain her behavior. With respect to the other person, I am

an external observer who is unable to have any direct access to what is going on in her mind – to her beliefs and desires. For this reason, to explain or predict her behavior, I require something more than perception or intuition; I require a way to infer what her mental states may be, or a way to project a model of thought into her mind. Thus, TT argues that we rely on a theory, folk psychology, to infer the other's beliefs and desires. In contrast, ST suggests that no theory is necessary since we have mechanisms that allow us to generate a model, a simulation, of the other person's beliefs and desires, which we then project to her mind. Accordingly, for both TT and ST, we mindread or mentalize the other person; we take her behavior as evidence and we make sense out of it by applying folk psychology, or by running a simulation routine in which we put ourselves in her shoes and draw up pretend beliefs and desires that we then project to her mind.

The starting problem refers to the question of precisely how we get either of these processes off the ground. Consider, for example, the description of a simulation routine provided by Nichols and Stich (2003):

> The basic idea of what we call the 'off-line simulation theory' is that in predicting and explaining people's behavior we take our own decision making system 'off-line', supply it with 'pretend' inputs that have the same content as the beliefs and desires of the person whose behavior we're concerned with, and let it make a decision on what to do' (pp. 39-40).

Now ST claims that the way we understand the other person's beliefs and desires is by employing this kind of simulation routine. We put ourselves in their shoes by drawing up pretend beliefs and desires "that have the same content" as their beliefs and desires. The problem should be quite apparent: we seemingly have to know the content of their beliefs and desires in order to run the simulation that will tell us what their beliefs and desires are. That this kind of solution runs in circles is also apparent in Alvin Goldman's description of simulation. He outlines three steps, to the simulation routine, but it's the first one that's a little tricky. 'First, the attributor creates in herself pretend states intended to match those of the target. In other words, the attributor attempts to put herself in the target's 'mental shoes'' (Goldman 2005, 80). It's not at all clear how we can know which beliefs will match those in the other person unless we already understood the other person. In that case we would have no need of simulation. The starting problem here is just the problem of how we take that first step, precisely, a step into the other person's shoes.

This problem is at least part of what motivates the recent shift to hybrid theories that combine TT and ST. That is, one can respond to the starting problem in ST by appealing to folk psychology and by suggesting that we gain traction on the other person's mental states by theoretical inference. Once we do that, then we are off and running the simulation routine and making our understanding more precise. But this hybrid solution ignores that fact that TT also has a starting problem. If we ask, what aspect, or piece, or rule of folk psychology we should appeal to in order to form our inference about the other person's beliefs and desires, we seemingly have to know with some degree of precision

what their situation is and what they would believe or desire in that situation. More generally, to use folk psychology in this way we have to know what situations are the appropriate ones in which to apply this piece of folk psychology rather than that piece, and part of what tells us that, because it is part of the situation, is what the other person is thinking. Of course we can continue on the circle if, as a theory theorist we want to go hybrid. That is, we can try our hand at simulation to figure out what the other person is thinking, and then be able to bring the relevant piece of folk psychology to bear on the situation. Obviously, however, going around in a tight circle like this does not solve the starting problem for either TT or ST.

If we want to stay with this conception of the problem of social cognition – that is, with the idea that we need to infer or simulate because we have no direct access – then we are surely tempted to point to an answer to the starting problem that we may take as obvious. The answer is to appeal to the idea that we get the process of inference or simulation off the ground by employing background knowledge. We know what beliefs or desires to infer or simulate because we have a broad knowledge of the sorts of situations we and others encounter and what to expect in the way of beliefs and desires in such situations. The theory theorists might in fact claim that this just is what they mean by folk psychology. But that doesn't seem right if we are to think of folk psychology as a theory. As a theory, folk psychology can only be generalized knowledge – that is, a set of generalizations or rules or abstract platitudes. The trick is to know when to apply the rules and/or platitudes, and that takes practical knowledge which is much more particular and situated.

Having the requisite background knowledge is much more like the situation that Aristotle describes when he explains how the *phronemos*, the person with practical wisdom, knows when, and how, and in what situation, and with what people, he should engage in action. This is knowledge about the *particularities* of situations; and it is knowledge that may vary case by case; it's the kind of knowledge that cannot be summarized in a set of rules or platitudes. Aristotle also tells us how we get such knowledge. Moreover, his solution is exactly the same solution that we need in order to explain how we get the background knowledge that would solve the starting problem for social cognition. Unfortunately, it goes directly against the conception of the mind that TT and ST start with and, significantly, it suggests that we already have a more basic way of understanding others that in most circumstances is just as adequate, or perhaps more adequate (since it does not involve a starting problem) than theoretical inference or simulation.

Aristotle's answer is that we get the kind of background knowledge essential for practical wisdom by being brought up in the right way, and by hanging around with the right people, by seeing and understanding and imitating their actions. Setting aside the moral context in which he answers this question (although normativity is clearly part of what we would need to consider for a full answer), recent developmental studies help to explain precisely how we come to understand the intentions and actions in a way that is more primary than theoretical inference or simulation.

Here, without going into great detail, I will refer to interaction theory (IT) – that is, to an embodied and enactive approach to social cognition that emphasizes the role of our interaction with others from the very beginning of life (Gallagher 2001; 2004; 2005; Gallagher and Zahavi 2008; Hobson 2002; Reddy 200X; Rochat 2010; Trevarthen 1979; Trevarthen and Hubley 1978). According to IT, we do not have to mindread or try to infer or simulate the mental states of others to understand them. Rather, because from the very beginning we have been interacting with others, imitating them, engaging them in joint attention and joint actions, hanging around with them in a vast variety of situations, communicating with them, often in an emotional key, and sharing both personal and cultural narratives (Gallagher and Hutto 2007; Hutto 2008), and because within such rich contexts we are able to enactively perceive (that is, to perceive in terms of our possibilities of responding to social affordances offered by others) their intentions and emotional expressions, and the meaning of their actions, we have little if any need for mindreading or concerning ourselves with what Rochat (2010, 1) calls the 'cold calculations and logical inferences' of theory of mind.

IT appeals to the developmental concepts of 'primary' and 'secondary intersubjectivity' (Trevarthen 1979; Trevarthen and Hubley 1978). Primary intersubjectivity consists of embodied sensory-motor capacities that infants develop in close interactions with others, starting with their care-givers. These are face-to-face interactions that include early imitation (Meltzoff and Moore 1977), the ability to follow gaze, to recognize when someone is directing attention to me, to be able to parse intentional actions into their meaningful units (Baldwin 1993; Baldwin et al. 2001), and so on. Secondary intersubjectivity begins at 9-12 months with joint attention which lays the basis for joint action (see Fiebich and Gallagher, submitted). In further interactive processes we learn what others mean, and what the world means, by seeing and engaging with others in very concrete contextualized actions. We also start to notice that in certain contexts others can take on socially defined roles.

What comes with this embodied and engaged interaction with others is just the practical knowledge, the background for understanding others, which we put into play in our continuing interactions and communicative practices. When we already share a broad culture and a set of specific social norms and practices with others, and when that background knowledge is there at our disposal, then the minds of others are not closed books that we have to read by inference or simulation. The actions and the interactions that we want to understand are not reducible to beliefs and desires hidden away in the other's head. The other person, in most of our everyday interactions (and excluding certain psychopathologies and puzzling cases), is in-the-world, engaged with us, in the same situation, or an understandable variation of the situation that we are in, or, as we acquire language and episodic memory, in a situation that we are familiar with through communicative and narrative practices (Gallagher and Hutto 2008).

For IT there is no starting problem because it is a question of bootstrapping development. Before we have a chance to wonder how we will ever know what the other person is thinking, we're caught up in shared emotions, exchanging smiles and vocalizations, sharing attention, being told stories, playing with others, being told what they want and what they don't want, being told how to behave, and so on. In this developmental story, one's social interactions are not initiated on one's own; others are already there providing us with a background that we begin to share in earliest development. If we were in the situation described by theory theorists or simulation theorists, and were confronted with a hidden realm of unknown mental states that we had to puzzle out, the only way we could get our mindreading processes off the ground would be by enacting just such background knowledge, or what, following Bruner and Kalmar (1998; see Gallagher, in press), I've called the 'massive hermeneutical background' that comes along with the intersubjective interaction I've just described. But to the extent that we are already situated and engaged in this kind of interaction, and already have this background knowledge, which includes a pragmatic understanding of others, then the minds of others are not so hidden away, and we do not have to theorize or simulate.

The Chinese Room

John Searle's (1980) famous thought experiment involves a non-Chinese-speaking person sitting in a room. The room has a table, a large book containing a set of rules, and paper on which to write. There are two slots in the walls – an input and an output slot. Through the input slot pieces of paper containing Chinese characters come into the room. Each time this happens the person has the task of writing Chinese characters on blank sheets of paper, using the book of elaborate rules which tell him which characters to write when he sees a specific combination of characters on the paper that comes in through the slot. He then pushes what he written through the output slot. This person doesn't know that the Chinese characters he receives from outside of the room are questions composed by Chinese speakers. If he follows the set of rules perfectly, the Chinese characters that he writes and outputs are answers to precisely those questions. From the outside, observers infer that the person in the room understands Chinese. The person in the room, however, does not understand Chinese, and doesn't even know that he is processing questions or composing answers. He is performing a set of syntactical operations, following the instructions (the syntax) contained in the book. Thus, Searle concludes, there is no understanding of Chinese, no Chinese semantics or intentionality involved.

The target of Searle's Chinese Room (CR) argument is what he calls 'strong AI'. In contrast to the latter, which would explain the mind purely in functionalist terms of computational syntax, the CR argument demonstrates that semantics cannot be reduced to computational syntax -- or that syntax by itself can never give you semantics (intentionality, meaning). Searle lays the problem out in a vocabulary shared with strong AI, one which make the questions of intentionality and the mind a matter of physics, syntax, and/or semantics. The use of this vocabulary, however,

seems more than rhetorical since in the end Searle will frame his own view in these terms.

If we accept Searle's point that syntax does not add up to semantics, then the question becomes what does give us semantics? The CR may not have been designed to give a positive answer to this question; its design was specifically framed in terms of defeating strong AI using the categories that AI was using at the time. I have argued (Gallagher 2009) that the design of the CR argument, although perfectly adequate for purposes of critiquing AI, nonetheless frames the problem of semantics in a way that oversimplifies the cognitive system, and leads Searle to one particular answer (where the physics or physical system is equated with the brain) that excludes a more adequate answer that he himself points to. The various "replies" that were made to the CR argument are also locked in the same oversimplified framework.

The "systems reply," for example, claims that it is not the syntax alone, but the whole system – the syntax and the physics (the person, but also the room, the Chinese characters, the syntactic rules, etc.) – that generates the semantics. The systems reply, however, doesn't go beyond the elements that Searle and strong AI agree are contenders for explaining the mind. The "robot reply" argues that the system has to be embodied in some way, and exposed to the world outside of the CR. Some thinkers (Rey 1986; Harnad 1989; 2002; Dennett 1991; Crane 1996) follow this line of reasoning back towards an enhanced and strengthened computational model of the mind. The robot reply, however, suggests an alternative route, which, as we'll see, represents a continuing challenge to robotics and contemporary AI. Moreover, the resources needed to map out this alternative route are to be found in Searle's own work, although Searle misses this because of the way that he has defined the problem space.

Searle's solution is to grant life to the physics. For Searle, semantics/intentionality is an emergent property of the brain, not because of its high degree of complexity (although Searle does not deny this kind of complexity), but because of its biological nature. "Whatever else intentionality is, it is a biological phenomenon and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena" (1981, p. 305). Or, as he puts it in his later work: "There are brute, blind neurophysiological processes and there is consciousness, but there is nothing else" [1992, p. 228]. Of course, there is already plenty of neurobiology in the CR – the individual in the CR does have a brain. Indeed, all of the identified elements, expressed in the circumscribed vocabulary of physics and syntax, seem to be present in the CR, so why doesn't the individual develop the semantics – that is, why doesn't he gain an understanding of Chinese?

Searle's response to the systems reply is that if we internalize all the elements of the system, i.e., *memorize* the rules and symbols and let the person compute these things in his head, the person will still not understand Chinese. Searle suggests, reflecting a suggestion made by the robot reply, that "we can even get rid of the room and suppose he works outdoors." That is, we could let the system walk around in the world. Even in that case, Searle contends, there is still no understanding of Chinese. I'm not so sure. Once we let the syntactical processor out

of the room, and into a social world of Chinese speakers, and especially if the processor is neurobiologically embodied, a number of other elements – including social interaction and the massive hermeneutical background – start to play an essential role.

It is odd that Searle arrives at a narrowly and neurobiologically-based internalist position with respect to the mind, since the concept of the "Background" of intentionality (1983; 1992) plays an important role in his thinking. The Background, as he conceives it, contains "certain fundamental ways of doing things and certain sorts of know-how about the way things work ..." (1983, p. 20). Indeed, Searle makes the Background a prerequisite for intentionality. "Without the Background there could be no perception, action, memory, i.e., there could be no such Intentional states .... [T]he Background provides necessary but not sufficient conditions for understanding, believing, desiring, intending, etc., and in that sense it is enabling and not determining" (1983, 151-152, 158).

Life in the Chinese Room, which is a small and non-Chinese space, excludes the relevant Chinese Background. Moreover, the occupant's capacities for action and interaction, including linguistic activity, with Chinese speakers, are non-existent. Locked in the Chinese Room one is in an artificially impoverished environment that excludes the kind of social interactions through which one could make sense out of the Chinese language and gain the relevant background for understanding it. Fodor is right to remark that "Searle gives no clue as to why he thinks the biochemistry is important for intentionality and prima facie, the idea that what counts is how the organism is connected to the world seems far more plausible" (1991, p. 521). Once we liberate the syntactical processor from the narrow confines of the Chinese Room, and allow the system to engage in the external complexities of the physical and social environment, cultural traditions, and the intersubjective interaction that can only be realized in embodied practices, contextualized speech acts, and developing narratives that provide the necessary background, it would be difficult to prevent the person from gaining the kind of semantics that Searle seeks.

Searle will have none of this, however. For him, all of these extra-syntactical elements that make up the Background enter into the system by way of neurophysiology. Thus, “when we describe a man as having an unconscious belief, we are describing an occurrent neurophysiology. ... The occurrent ontology of those parts of the Network that are unconscious is that of a neuro-physiological capacity, but the Background consists entirely in such capacities” (1992, p. 188). Indeed, no sooner does he liberate us from the Chinese Room than he locks us up in a vat.

> Even if I am a brain in a vat--that is, even if all of my perceptions and actions in the world are hallucinations, and the conditions of satisfaction of all my externally referring Intentional states are, in fact, unsatisfied--nonetheless, I do have the Intentional content that I have, and thus I necessarily have exactly the same Background that I would have if I were not a brain in a vat and had that particular Intentional content. *That* I have a certain set of Intentional states and *that* I have a Background do not logically require that I be in fact in certain relations to the world around me …” (1983, p. 154).

Searle's internalist position keeps him locked into a problem space that rules out just the solution he needs. The brain takes the place of the Chinese room.

> The brain is all we have for the purpose of representing the world to ourselves and everything we can use must be inside the brain …. Each of our beliefs must be possible for a being who is a brain in a vat because each of us is precisely a brain in a vat; the vat is a skull and the 'messages' coming in are coming in by way of impacts on the nervous system (1983, p. 230).[1]

Even as he confines us to a CR-like brain, Searle points to a solution that requires some connection to the world around us: "I could not, as a matter of empirical fact, have the Background that I do have without a specific biological history and a specific set of social relations to other people and physical relations to natural objects and artifacts" (Ibid.). Yet he can't have that solution because he defined the problem in terms that already exclude the social world; social relations are impossible if we think of the mind purely in the limited terms that define the problem space: physics, syntax, semantics.

The robot replies and the frame problem[2]

The *original* robot reply suggests that we allow a robot to wander around the world outside of the CR in order to causally interact with worldly objects to which the Chinese words refer.[3] Since knowing the meaning of a word requires that a speaker know to what it refers, only a robot that explores the physical world could learn what the Chinese characters mean. There is a second robot reply, however, that comes closer to getting it right. I'll call this the *social robot reply*. Responding to Searle's proposal to memorize the syntactical rules and to allow the CR occupant to venture out into the world, Tim Crane comes closest to stating the social robot reply: "…if Searle had not just memorized the rules and the data, but also started acting in the world of Chinese people, then it is plausible that he would before too

---

[1] "My own view (and in this I think I do depart from Wittgenstein) is that ultimately our explanations of these [Background] capacities will be biological. That is to say, the existence of Intentional states is explained by the fact that we are creatures with the certain sort of neurophysiological structure, and certain sorts of biological capacities" (1991, p. 293; see 1992, p. 188).

[2] Research on this section was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

[3] One might wonder why we have shifted to a discussion about robots rather than, say, the human who originally occupied the CR. The CR argument targets strong AI; it's an argument against the computational conception of the mind. The proponents of the robot reply attempt to respond to Searle in terms that defend the possibility of AI – in doing so they remain tied to the conception of the mind that belongs to the original problem space defined by GOFAI (Good Old Fashioned Artificial Intelligence) (Haugeland 1985).

long come to realize what these symbols mean" (1996, 127). The emphasis here should fall on "people," and we should add social interaction and culture. Obviously this is what works for humans, and the point is that neither syntax nor neurobiology is sufficient for semantics. One needs to be immersed in a social world. Neither a human nor a robot can simply wander about the world alone and expect to put words to things. We know that humans learn such things from other humans through processes of imitation and other forms of interaction. To learn Chinese – to get the semantics – one has to interact with Chinese speakers in physical and social contexts.

This argument is put in terms of the robot reply because Searle's CR argument is directed against strong AI or GOFAI. The question is whether we can engineer an artificial system that would be able to navigate and negotiate itself in a human social world. The social robot reply doesn't answer the question of whether a robot could obtain an understanding of things, and the words (Chinese or otherwise) that signify things, in a way that is any different from the socially-grounded way that humans learn to understand things and signifiers. The challenge is, whether by engineering or by being socially grounded, the robot, like the human, would be able to avoid running into the frame problem. Robots and humans are on opposite sides of the frame problem. Humans, because of the way that they become immersed in meaning via social contexts, have a tremendous amount of background general knowledge which breaks down only in very specialized areas that require expertise. On the one hand, for example, I have no idea of what's relevant or not relevant for solving a problem in chemical engineering because I don't have the specialized background in that area. Nonetheless, I do relatively well in regard to everyday life. On the other hand, robots and intelligent systems are usually engineered to have specialized knowledge in a circumscribed area. Within that frame, they are proficient. But even within the circumscribed frame, near the edges perhaps, what counts as relevant starts to exceed the frame and the robotic systems begin to fail in performance. Robots do relatively poorly in regard to everyday life and any area that is outside of their circumscribed operational space.

We may throw some light on these issues if we consider some of the problems involved in designing robots that can communicate with humans. With respect to sending the robot out into the social world, the issue is not simply speech recognition; we would also want the robot to reply appropriately, and this involves a number of problems (see, e.g., Kollar et al. 2010). To have a robot capable of replying across a wide variety of circumstances would require a design that would allow not just task-related domain knowledge, but the capacity for transference of knowledge across domains. Moving around the world we are frequently confronted by circumstances that are not predictable, consistent or familiar. Moreover, the meaning of words and gestures can change from one context to another. Knowing what word or gesture to use in a specific circumstance requires that a robot recognize that circumstance for what it is. This requires background knowledge. In this respect we are seemingly caught in a circle: the only way to gain sufficient background knowledge to avoid the frame problem is by interacting (communicating) with others; but successful interacting (communicating) with others requires background knowledge. The solution here is not to think in

computational engineering terms, but in the developmental terms outlined in the first section. Robots require the embodied capacities involved in primary and secondary intersubjectivity – pre-requisites in the human for being able to learn language and to develop episodic memory, and for acquiring the massive hermeneutical background necessary to solve starting and frame problems. This is not GOFAI; and it's more than what Rodney Brooks (1991) suggested as a non-representational way (using the world as its own model) to get around limiting design to "specialized subproblems." The solution is not simply an enactive and dynamic linking of perception and action; it requires *interaction* with others who already have the background (Gallagher 2007).

So far, robots, even when they are designed to interact with humans, remain autistic. They have a difficult time recognizing connections where there are no literal connections; that is, they have a difficult time with metaphorical association, something that non-autistic humans have a difficult time avoiding. The human memory system – especially with respect episodic memory – is unlike computer memory. It's 'leaky memory' (Gallagher 2009). It leaks because it is constantly and imperfectly interconnected with a full intentional and affective system. Imperfect because it cannot neatly isolate semantic elements along strict logical boundaries or quarantine them in neat ontological categories. For example, if the English language speaker locked in the Chinese Room sees the Chinese character 人 (which unbeknownst to him means 'man' or 'human') often enough, it could easily call forth a memory of a tent or of drawing a stick-man. For a less transparent reason, the character 閉 might serve to remind him of his own situation as the occupant of the Chinese room. Without knowing the Chinese meaning of the characters one might still discern similarities in shape between 人 and a component of 閉, which looks a bit like a stick-man pushed into a small room, and which, in Chinese, actually signifies 'confinement' (see Wieger 1965). A character may have such aesthetic appeal that it starts to manifest itself in his sketches or doodles. It's also possible that a syntactic rule designed to function in the CR may invade his concentration when he attempts to solve a mathematical problem. The point is not whether he gets it right (seeing 人 as 'man' rather than 'tent'), but that humans are inclined to make these associations – because episodic memory leaks into semantic memory and vice versa, and our memory systems invade our everyday tasks. This involves what semioticians refer to as blending (Brandt and Brandt 2005; Fauconnier and Turner 2002); it's the basis for metaphorical thought and creative solutions, as well as for silly fantasies. This sort of leaky and metaphorical blending happens not only in our heads. It is often institutionalized in cultural expressions (think of the sexual innuendos that we are bombarded with in commercial advertisements), and such expresssions often shape our social interactions and communicative practices.

Whether we are locked in the Chinese Room, or allowed to interact with others in the everyday world, to internalize syntactic rules and Chinese characters is not simply to commit them to memory; it is rather to introduce a

potentially infinite linguistic system into a general and leaky system of intentional experience that tends to see meaning wherever it can find it. The background takes shape and comes into our cognitive experience just in these kinds of processes. Without this extraordinarily productive imperfection, robots remain autistic.

What we learn when we learn Chinese or any language, when that language is our first language, is not simply word-thing correspondence. We learn, as Wittgenstein might say, a form of life, and meaning is tied to that form of life. That is, part of what we acquire in learning language, and more generally in communicative and narrative practices, is the massive hermeneutical background necessary to make sense of the world and others in the world. Indeed, we can see that this fails dramatically if we simply program word-thing correspondence into computers. Once the non-leaky, rigid rule-based parameters of use are broken, the computer fails to respond appropriately.

Add another very basic issue concerning communicative attunement, which forms part of the pragmatic skill-background required to enter into successful communication. This attunement involves the embodied dynamics of interacting with others – something that we learn very early in development, and that is essential to the kind of interaction that characterizes intersubjective communication (see, e.g., De Jaegher et al. 2010).[4] The timing involved in this kind of embodied interaction among humans involves turn taking, for example, but is not something that always happens according to an orderly and static statement-response-statement-response schema. In real dialogues one speaker does not always wait for the other to respond or to finish their response, although this practice does not always interrupt the ongoing dynamics and possibly even defines that dynamics. Yet this kind of disorderly dynamics can lead to break downs when one of the interlocutors is a robot designed for orderly conversation, or is simply slow in responding (see Green & Eklundh 2003).

Conclusion

The problems encountered in designing social robots and in human-robotic interaction, send us back to the issues of social cognition discussed in the first section. In each case, whether we are attempting to explain how we understand one another, or how semantics (intentionality) comes to be, or how we can design intelligent robots, we have seen that the way is blocked when we conceive of mindedness or the mental system too narrowly in terms of internal processes – whether they be strictly "in the head" mental states, representational, computational or syntactic operations, or neurobiologial activities. I'm suggesting that we can move forward on these problems only when we look more widely, not only to embodied action in the physical environment, but to intersubjective processes in the social world and to the massive hermeneutical background of cultural knowledge and practical know-how that is not only the continually

[4] Emotional attunement and recognition of another person's emotional state is another important aspect that is neither knowledge nor skill, but involves an embodied perception that is capable of recognizing emotional expression in face, gesture, movement, vocal intonation etc. (Hashimoto et al. 2009).

constituted product of such processes but also the scaffolding that allows such processes to get off the ground.

References

Baldwin, D. A. 1993. "Infants' ability to consult the speaker for clues to word reference." *Journal of Child Language* 20: 395-418.

Baldwin, D. A., Baird, J. A., Saylor, M. M. and Clark, M. A. 2001. "Infants parse dynamic action." *Child Development* 72 (3): 708-17.

Brandt, L. and Brandt, P. A. 2005. Making sense of a blend: A cognitive-semiotic approach to metaphor. *Annual Review of Cognitive Linguistics*, 3, pp. 216-249.

Brooks, R. 1991. Intelligence without representation. *Artificial Intelligence,* 47, 139-59.

Bruner, J. and Kalmar, D. A. 1998. Narrative and metanarrative in the construction of self. In M. Ferrari and R. J. Sternberg (eds.), *Self-Awareness: Its Nature and Development* (pp. 308-331). New York: Guilford Press.

Crane, T. 1996, *The Mechanical Mind*: *A Philosophical Introduction to Minds, Machines and Mental Representation*, London: Penguin.

De Jaegher, H. Di Paulo, E. and Gallagher, S. 2010 Can social interaction constitute social cognition? *Trends in Cognitive Sciences*. Published online 30 July 2010. 10.1016/j.tics.2010.06.009

Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown and Company.

Fiebich, A. and Gallagher, S. (submitted). Joint Attention: from Interaction to Joint Action. *The Review of Philosophy and Psychology*

Fauconnier, G. and Turner, M. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* Basic Books.

Fodor, J. A. 1991. Searle on what only brains can do, in D. M. Rosenthal (ed.), *The Nature of Mind*, Oxford: Oxford University Press.

Gallagher, S. In press: Narrative competency and the massive hermeneutical background. In P. Fairfield. *Education, Dialogue, and Hermeneutics*. New York: Continuum.

Gallagher, S. 2009. The key to the Chinese Room. In. Karl Leidlmair (ed). *After Cognitivism* (87-96). Dordrect: Springer.

Gallagher, S. 2007. Social cognition and social robots. *Pragmatics and Cognition* 15 (3): 435-54.

Gallagher, S. 2005: *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S. 2004: Understanding problems in autism: interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, and Psychology*, 11, 199–217.

Gallagher, S. 2001: The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies,* 8(5-7), 83-108.

Gallagher, S. and Hutto, D. 2008: Understanding others through primary interaction and narrative practice. In: J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.), *The Shared Mind: Perspectives on Intersubjectivity* (pp. 17-38). Amsterdam: John Benjamins.

Gallagher, S. and Zahavi, D. 2008: *The Phenomenological Mind.* London: Routledge.

Goldman, A. I. 2005: Imitation, mind reading, and simulation. In: S. Hurley and N. Chater (eds.), *Perspectives on imitation II* (pp. 80-91). Cambridge, MA: MIT Press.

Green, A. and Eklundh, K. S. 2003. Designing for Learnability in Human–Robot Communication. IEEE Transactions on industrial electronics 50 (4): DOI: 10.1109/TIE.2003.814763

Harnad, S. 1989. Minds, machines and Searle, *Journal of Experimental and Theoretical Artificial Intelligence* 1: 5-25

Harnad, S. 2002. Minds, machines, and Searle 2: What's right and wrong about the Chinese Room argument, in J. Preston, and M. Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, New York: Oxford University Press

Hashimoto, M. Misaki, Y. and Tatsuy, U. 2009 Effects of emotional synchronization in human-robot KANSEI communication. The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, Sept. 27-Oct. 2, 2009

Haugeland, J. (1985), *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: MIT Press

Hobson, P. 2002. *The Cradle of Thought*. London: Macmillan.

Hutto D. D. 2008: *Folk Psychological Narratives: The Socio-Cultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Kollar, T., Tellex, S., Roy, D. and Roy, N. 2010. Toward Understanding Natural Language Directions. *Human-Robot Interaction* (HRI), 2010 5th ACM/IEEE International Conference. DOI: 10.1109/HRI.2010.5453186. pp. 259-66.

Meltzoff, A. N. and Moore, M. K. 1977. Imitation of facial and manual gestures by human neonates. *Science* 7 (4312): 75-78.

Nichols, S., & Stich, S. P. 2003. *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Clarendon Press.

Reddy, V. 2008: *How Infants Know Minds*. Cambridge, MA: Harvard University Press

Rey, G., 1986. What's really going on in Searle's "Chinese Room", *Philosophical Studies* 50:169-85.

Rey, G., 2002. Searle's misunderstandings of functionalism and strong AI' in J. Preston, and M. Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, New York: Oxford University Press

Rochat P. 2010. Is social cognition an oxymoron? Comments on Astington and Edward, Miller, Moore and Sommerville. In: Tremblay RE, Barr RG, Peters RDeV, Boivin M, eds. *Encyclopedia on Early Childhood Development* [online]. Montreal, Quebec: Centre of Excellence for Early Childhood Development:1-5. Available at: http://www.child-encyclopedia.com/documents/RochatANGxp.pdf. Accessed 2 November 2010.

Searle, J., 1980. Minds, brains and programs, *Behavioral and Brain Sciences* 3: 417-57.

Searle, J. 1981. Minds, brains, and programs, in J. Haugeland (ed.), *Mind Designs* Montgomery, VT: Bradford Books.

Searle, J. 1983. *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.

Searle, J. 1991. Response: The background of intentionality and action, in E. Lepore and R. Van Gulick (eds.), *John Searle and his Critics,* 289-299. Oxford: Basil Blackwell.

Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Trevarthen, C. B. 1979: Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (ed.), *Before Speech* (pp. 321-48). Cambridge: Cambridge University Press.

Trevarthen, C. and Hubley, P. 1978: Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (ed.), *Action, Gesture and Symbol: The Emergence of Language* (pp. 183-229). London: Academic Press.

Wieger, L. 1965. *Chinese Characters: Their Origin, Etymology, History, Classification and Signification. A Thorough Study from Chinese Documents*, trans. L. Davrout. New York: Dover/Paragon.