Online Consciousness Conference 2013

Self-Consciousness, Interaction, and Understanding Others

Katja Crone (University of Mannheim)

Abstract:
The paper explores the basic conceptual relationship between social cognition, intersubjectivity and self-consciousness. A much debated approach to social cognition is the view that the ability to perceive, understand and interpret the behavior of others relies on interaction (Gallagher 2005, 2008; Ratcliffe 2007; Gallagher & Zahavi 2008). Even though the main argument hinges heavily on the concept of interaction, it is not clear what this exactly means. In the paper it will be argued that interaction in this context should be understood as a stance persons adopt towards each other, which in turn presupposes that they acknowledge each other as self-conscious agents. This view is inspired by an argument originally introduced by Fichte.

There seems to be widespread agreement that self-consciousness, broadly understood as the ability to conceive of oneself *as* oneself, and intersubjectivity, in the forms of an intersubjective environment and intersubjective encounters, are closely related (e.g., Baker 2012; Newen & Vogeley 2007; Musholt 2012). Moreover, findings in developmental psychology suggest that self-consciousness in this sense goes hand in hand with the ability to ascribe mental states to other persons and to understand and interpret their behavior (e.g., Gopnik 1993; Lang & Perner 2002). I will take this sketchy observation as a starting point and try to explore in more detail the conceptual relationship between self-consciousness, social cognition and intersubjectivity.

A much-debated approach to social cognition takes interaction, in the sense of embodied intersubjective engagement, to be a requirement for the ability to understand others (e.g., Gallagher 2005, 2008; Ratcliffe 2007; Gallagher & Zahavi 2008; De Jaegher 2009; De Jaegher & Di Paolo & Gallagher 2010). Most versions of this positions argue that only because we interact with each other that we are able to interpret bodily movements like facial expressions or gestures of others as expressing mental states. And this is why, according to these approaches, the ability to understand others cannot be reducible to cognitive abilities of a single subject. However, although the main argument hinges heavily on the concept of interaction, it is not clear what this exactly means. The unclarity also extends to the question whether interaction must be taken as a mutual real-time engagement taking place at the very situation when someone is trying to make sense of someone else's behavior.

My argument will have two parts: (1) First, I will turn to the question whether interaction can reasonably be understood as a direct or real-time engagement between the

interpreting subject and the person whose behavior is interpreted. I will argue that not all cases of social cognition require this sort of direct interaction. Therefore, "interaction" has to be given a different meaning provided one wants to hold on to the overall claim (as I do). (2) I will claim that "interaction" should be understood as a stance persons adopt towards each other, which in turn presupposes that they acknowledge each other as self-conscious agents. This view is inspired by an argument originally introduced by Johann Gottlieb Fichte in his book *Foundations of Natural Right* (1796/97). The argument will not only clarify the interaction view of social cognition but will also help reveal the basic conceptual relationship between self-consciousness, consciousness of others, and intersubjectivity.

1. The interaction theory (IT) of social cognition and the interaction requirement

The question of what actually enables persons to perceive, understand and interpret the behavior of others is a much-discussed problem in philosophy of mind and cognitive science. An important strand of the debate challenges the assumptions of the two standard views to social cognition, namely the theory theory (TT) and the simulation theory (ST). Briefly, TT holds that when we try to make sense of others' behavior we apply a "theory" consisting of law-like generalizations and theoretical concepts (Gopnik & Wellman 1994, Gopnik & Metzloff 1997); we infer from what we see - for instance a person moving towards the fridge and opening it - to 'hidden' mental states like "being hungry" and "looking for food". ST on the other hand holds that we put ourselves in the shoes of others and use our own experiences as a model to understand other persons (e.g., Goldman 1984; 2006; Currie & Ravenscroft 2002); because I imagine the mental state of being hungry I can make sense of how the observed person behaves. However, objections have been raised towards these views. Shaun Gallagher (2008) and Matthew Ratcliffe (2007), for instance, argue that ST and TT do not adequately describe the process of understanding others. They hold, on their part, that interpersonal understanding in everyday life is, in most cases, far less demanding and complicated than ST and TT assume. Normally, we do not have to make long-winded inferences from observed behavior to mental states that we take to be causally responsible what we see, as TT holds. Nor do we have to precisely imagine what the other person might be thinking or perceiving, as ST assumes. Rather, the argument goes, we usually *directly* perceive what the other person thinks and feels by merely looking at her face, her gestures and movements, or by hearing the tone of her voice. According to proponents of the interaction theory, the question of *access* (to others' mental states) is therefore misconceived. Understanding others means interacting with them, which is why social cognition is to be understood in terms of interaction (Gallagher 2008, 536; especially 540; Ratcliffe 2007,

23). The approaches stress the particular I-you relationship, conceived of as embodied intersubjective engagement, to be essential to social cognition. The ability to understand others rests upon a two-directional mutual relationship, which is why it is explanatorily relevant. ST and TT are therefore mistaken in focussing only on the cognitive abilities (and enabling conditions) on the part of the interpreting subject and thus making social cognition to be only 'one-directional', that is, a cognitive procedure from a third-person perspective. Instead of theorizing and simulating, as the interaction theory holds, embodied subjects bear a "direct" experiential relation to others' mental states due to the particular I-you-relationship that is typical for interpersonal understanding. This claim is supported in part by Gibsonian psychology as well as by phenomenological insights borrowed from Sartre, Merleau-Ponty and others.

For the most part, I also think that the structure of intersubjective relationships must be taken into account in attempts to explain social cognition. However, existing interaction theories to social cognition have their own problems for they suffer from being in part too vague. For instance, objections have been raised as to what "direct perception" in this context precisely means (e.g., De Jaegher 2009); although this is indeed an important issue I will not enter into it here. Rather, my point is that the key concept of interaction is not sufficiently clear. Shaun Gallagher, for instance, seems to take interaction for granted when he concentrates on what enables the direct perception of others' mental states. Being able to directly perceive what others think and feel is, as Gallagher holds, "informed by my own interaction with them and others" (Gallagher 2008, 540); and he goes on to stress that "in ordinary instances of interaction with others, I am not in the observer position" (ibd.). Even though we do not get any further information as to what interaction precisely means, the scope of the underlying general assumption seems obvious: social cognition is in many cases nothing but a certain form of interaction between an interpreting person and a person whose behavior is being interpreted. Matthew Ratcliffe argues in a similar vein: "[dialogical] interaction, I argue, is itself constitutive of our ability to understand each other. Interpersonal understanding is not usually a matter of deploying internal abilities in observational contexts. Instead, the abilities partly reside in the interaction" (Ratcliffe 2007, 23).

These claims suggest a very strong understanding of the relation between interpersonal interpretation and interaction that IT approaches make use of. Moreover, they seem to make a universal statement, according to which *all* instances of interpersonal understanding require interpersonal engagement (De Jaegher & Di Paolo & Gallagher 2010, 442, 443). And this in turn means that interaction must be at least a necessary if not also a sufficient condition of interpersonal understanding. How could this be better understood? Roughly, at least two quite different versions are possible: (1) either interaction has to be understood as a particular event taking place at the very moment

when someone interprets the behavior of someone else (real-time interaction); (2) or interaction is rather a feature that characterizes the particular stance persons adopt towards each other.[1]

Let's first consider option (1), according to which interaction must take place at the very moment of interpersonal understanding. In other words: whenever one interprets another person's behavior, there must be some interpersonal engagement actually taking place. This entails that both the person interpreting as well as the person whose behavior is being interpreted *act* on each other: the person interpreting takes an active part in perceiving what the other thinks while the other actively changes his or her facial expression, bodily position gesture etc. But is it really likely that such a face-to-face mutual interaction has to take place in order to understand each other's behavior? I doubt that this is the case. It seems that many counterexamples can be thought of that call the above assumption into question. Think of a private detective who is commissioned by a customer to tail a person. Let's assume that he is very skillful in performing this task, he follows her wherever she goes without ever being noticed by her, he observes her behavior from far away with the help of spy glasses and is thus able to reliably "read" her facial expressions. Wouldn't we say that the private detective is very well capable of understanding his observed object's behavior - even though there is clearly only a one-way action and no literal (mutual) *inter*action? Of course. Since this is so, it seems that the strong case of interaction is mistaken. This, however, doesn't mean that, conversely, real-time interaction never plays any role in social cognition. It is likely that people who have known each other for a very long time very frequently make use of real time interaction to understand each other only by looking at each other's faces and by acting and reacting in a given situation with certain gestures. However, the claim that social cognition *always* requires real-time interaction or embodied intersubjective engagement, is too strong. So if we want to hold on to the general assumption that interaction is a necessary condition of social cognition then "interaction" has to be understood in a different way.

This leads to option (2), which is, at first sight, somewhat less catchy: if "interaction" cannot mean real-time and face-to-face engagement in all cases, it might rather be understood as a defining feature of a stance persons adopt to each other. I will try to clarify this idea with the help of Fichte's approach to intersubjectivity and self-consciousness.

2. Interaction as a necessary condition of an interpersonal stance

[1] Notice that I am am not considering enabling conditions (in the developmental sense) of social cognition here. See for a further discussion De Jaegher & Di Paolo & Gallagher 2010.

What could be the sense of "interaction" if not taken as real-time encounter? Do we have to conceive of it as *potential* interaction (as Ratcliffe 2007 seems to think) - an interaction that may be absent in situations of social cognition but which could *in principle* be present? But what explanatory information does this idea carry? It is important to show how interaction can still have the function assigned to social cognition if not taken as real-time engagement. This is why I will now turn to the notion of a stance and to the question of how interaction may be functionally related to it and thereby to social cognition. In general, a stance is an attitude adopted in the face of an object or types of objects. The specific quality of the stance defines the ways in which one relates to the object. I take up a particular stance when I wish to evaluate the condition of an apple by its appearance (texture, color etc.), and I adopt a quite different one when I look at my brother's face to check whether he is still running a fever. Ultimately, I will claim that persons adopt a particular stance towards each other that is very different from a stance a person has when dealing with, say, inanimate objects - at least normally. I believe that once we get a better grip on what is specific to interpersonal stances, conceived as a necessary (but not sufficient) condition of social cognition, we are in a better position to capture the advantages of IT.

In the remainder of this paper I will try to further characterize what it means to adopt an interpersonal stance and in what sense this might be related both to interaction and social cognition. I will refer to an argument by Fichte because it offers an often overlooked account of the very fundamental relationship between self-consciousness and intersubjectivity.[2]

Fichte's general claim is that intersubjectivity (in a sense to be specified) is a condition of self-consciousness. Before I jump into the argument itself, I would like to make one important preliminary remark: Fichte's notion of self-consciousness has a complex structure. In order to do justice to the general claim one has to follow Fichte's distinction between (what I would call) pre-reflective self-awareness on the one hand and reflective self-consciousness (being aware of one's own mental states, having a concept of oneself as oneself etc.) on the other hand.[3] It is precisely the latter, i.e. reflective self-consciousness (and not pre-reflective self-awareness) which is at stake in the argument at issue. It should nevertheless be noted that, according to Fichte, pre-reflective self-

[2] Note that I will bracket aspects of Fichte's theory that would require a broader consideration of his transcendental system.

[3] Basically, what I call "pre-reflective self-awareness" refers to what Fichte calls "intellectual intuition" (*intellektuelle Anschauung*). In my understanding, pre-reflective self-awareness roughly means the same as having an experiential first-person perspective.

awareness is yet a further condition for reflective self-consciousness - an argument put forward by Fichte that I will not discuss here.[4]

In the *Foundations of Natural Right* Fichte claims that the awareness of being an individual subject requires that one be part of a realm of intersubjective mutual recognition.[5] First, it is especially important to note that, according to Fichte, self-consciousness includes not only the ability to conceive of oneself as oneself but most notably the awareness of oneself as a rational agent capable of setting one's own goals for actions; Fichte's concept of self-consciousness is practically embedded. This is also expressed by the basic assumption that self-consciousness is essentially a (mental) activity. Formally, all instances of self-consciousness are an "activity that reverts into itself" (Fichte 2000, § 1) - figuratively speaking. However, Fichte is puzzled with the question of which condition must be fulfilled for self-consciousness to be contentful or "real", as Fichte puts it; the subject's cognitive requirements alone are not sufficient for this. The reason is that the awareness of being an individual subject and free agent is necessarily bound to the awareness of an outside world - furthermore, of a world in which free action makes sense after all, of a world consisting of other free agents. And this is why the awareness of one's agency requires ascribing the same capacity to other rational beings - and vice versa. I believe that this idea of an intersubjective realm must be at the very basis of the concept interpersonal understanding that can be drawn on in approaches to social cognition.

But where exactly does the interaction part come into play, and how can its meaning be appropriately integrated into the accounts at issue? The structure of the above mentioned basic intersubjective realm can shed light on this. It is important to note that although Fichte uses terms that seemingly refer to actions, Fichte is offering a *structural* description that subjects are not necessarily consciously aware of. Intersubjectivity, as Fichte argues, implies both "demand" (*Aufforderung*) and "recognition" (*Anerkennung*) (Fichte 2000, § 3, iv): in order to be aware of oneself one has to be addressed *as* a self-conscious and free subject by another subject. Being faced by another person means being implicitly demanded to be aware and to make use of one's free agency. And to be 'addressed' in this very way means to (again: implicitly) *accept* the demand and thereby to *recognize* the other as an equally self-conscious and free agent by the same token. Note that this recognition is not just a theoretical acknowledgement of someone's status as a free agent; it also requires that he be treated as such by others.[6] Being part of such an intersubjective

[4] See especially Fichte's *Wissenschaftslehre nova methodo* (1796/99) (English title: *Foundations of Transcendental Philosophy*).

[5] Fichte's argument is at length unfolded in §§ 1-3 of the *Foundations of Natural Right*. Another but shorter and more scattered version can be found in the *Wissenschaftslehre nova methodo* §§ 18-19.

[6] This idea is captured in the following quote: "The human being (like all finite beings in general) becomes a human being only among human beings" (Fichte 2000, § 3, corollary 1).

realm means adopting this particular stance towards others even in the absence of direct encounter.

This basic structure of intersubjectivity is precisely what constitutes the interpersonal stance, a stance persons adopt towards each other as distinguished from inanimate objects. Interaction here is not real-time engagement but rather what characterizes an attitude that entails that they mutually acknowledge and treat each other as self-conscious and free agents. Due to this attitude they make sense of each other's behavior, of their facial expressions, their gestures, and their movements in a person-specific way. The private detective observes his object with his spy glasses from a distance, he interprets her facial expression without any mutual engagement taking place, but he views her from a typical personal stance - conceiving of her as a self-conscious subject, and this shapes the way he interprets her bodily expressions. Without necessarily having to take place in the very moment of interpreting someone else's behavior, interaction in the above sense bears a functional relation to social cognition and its fundamental patterns of interpersonal understanding. To be sure, the fact that this is just a very basic requirement calls for further conditions that must be met. I believe that the most promising direction to take would here be an approach that not only takes phenomenological considerations but also conditions of theorizing and simulating into account - depending on the particular context in which social cognition takes place.

3. Conclusion

The aim of this paper was to explore the basic conceptual relationship between self-consciousness, intersubjectivity and interaction viewed from the angle of the problem of social cognition. I started from the main idea of the interaction theory of social cognition, namely that understanding and interpreting others' behavior is necessarily related to interaction. Since not nearly all cases of social cognition require real-time interaction, I argued that "interaction" should be understood as a feature characterizing the particular stance persons adopt towards each other when they try to grasp what the other is thinking and feeling. I referred to an argument by Fichte in order to show how an interpersonal stance is necessarily related to self-consciousness. Fichte argues that self-consciousness, the ability to conceive of oneself as oneself and as a free agent, requires being part of an intersubjective setting. And this setting is created by persons who adopt an interpersonal stance towards each other. I argued that a relevant sense of "interaction" is captured by the structural features of this stance. This view, although relying on the concept of interaction, is compatible with the fact that we sometimes, if not very often, make sense of others' behavior without effectively being part of a real-time mutual engagement. Since the approach only spells out a fundamental requirement for social cognition, further conditions

will have to be specified in order to get a fuller picture. This may comprise not only further phenomenological descriptions but also conditions of theorizing and simulating, depending on the context.

References:

Baker, Lynne R. (2012) "From Consciousness to Self-Consciousness", in: *Grazer Philosophische Studien* (84) (ed. by K. Crone & K. Musholt & A. Strasser), 19-38.

Currie, Greg & Ravenscroft, Ian (2002) *Recreative Minds: Imagination in Philosophy and Psychology*, Oxford: Oxford University Press.

De Jaegher, Hanne (2009) "Social Understanding Through Direct Perception? Yes, by Interacting" (Commentary), in: *Consciousness and Cognition* 18, 535-542.

De Jaegher, Hanne & Di Paolo, Ezequiel & Gallagher, Shaun (2010) "Can Social Interaction Can Constitute Social Cognition?", in: *Trends in Cognitive Science* 14 (10), 441-447.

Fichte, Johann Gottlieb (2000) *Foundations of Natural Right. According to the Principles of the Wissenschaftslehre* (ed. by F. Neuhouser; trans. by M. Baur), Cambridge: Cambridge University Press.

Fichte, Johann Gottlieb (1992) *Foundations of Transcendental Philosophy* (ed. and trans. by D. Braezeale), Ithaka: Cronell University Press.

Gallagher, Shaun (2008) "Direct Perception in the Intersubjective Context", in: *Consciousness and Cognition* (17), 535–543.

Gallagher, Shaun & Zahavi, Dan (2008) *The Phenomenological Mind*, London: Routledge.

Goldman, Alvin I. (1989) "Interpretation Psychologized", in: *Mind and Language* 4, 161-185.

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

Gopnik, Alison (1993) "How We Know Our Minds: the Illusion of First-Person Knowledge of Intentionality", in: *Behavioral Brain Sciences* 16,1-14.

Gopnik, Alison & Metzloff, Andrew (1997) *Words, Thoughts, and Theories*, Cambridge, Mass.: Bradford MIT Press.

Gopnik, Alison & Wellman, Henry M. (1994) "The 'Theory-Theory'", in: L. Hirschfeld & S. Gelman (eds.) *Domain Specificity in Culture and Cognition*, NewYork: Cambridge University Press.

Lang, Birgit & Perner, Josef (2002) "Understanding of Intention and False Belief and the Development of Self-Control", in: *British Journal of Developmental Psychology* 20 (1), 67-76.

Musholt, Kristina (2012) "Self-Consciousness and Intersubjectivity", in: *Grazer Philosophische Studien* (84) (ed. by K. Crone & K. Musholt & A. Strasser), 63-89.

Newen, Albert & Vogeley, Kai (2007) "Menschliches Selbstbewusstsein und fie Fähigkeit zur Zuschreibung von Einstellungen", in: H. Förstl (ed.) *Theory of Mind*, Heidelberg: Springer, 99-116.

Ratcliffe, Matthew (2007) *Rethinking Commonsense Psychology. A Critique of Folk Psychology, Theory of Mind and Simulation*, New York: Palgrave Macmillan.