APhEx 15, 2017 (ed. Vera Tripodi) Redattore: Federica Berdini



INTERVISTE

Conversation with Johanna Seibt

by Laura Racciatti

Johanna Seibt is Professor with special responsibilities at Aarhus Universitet, Denmark. After receiving a Ph.D. in philosophy from the University of Pittsburg, she worked as Assistant Professor at the University of Texas at Austin, and completed her Habilitation (Dr. phil. Habil.) at the University of Konstanz. She is coordinator of two interdisciplinary research groups: the "Research Unit for Robo-philosophy," and the TRANSOR network. Johanna Seibt is best-known for her interpretation of Wilfrid Sellars' philosophy. Her research interests focus on the analytical ontology and history of metaphysics, the philosophy of dialogue, and the robo-philosophy. Over the last twenty-five years she has been working on a new process ontology, called "General Process Ontology," and the application of this theory to other ar-



eas, such as cognitive science, linguistics, conflict resolution, and robophilosophy. Her most recent project, entitled "What Social Robots Can and Should Do—Towards Integrated Social Robotics," is supported by a "Semper Ardens grant" from the Carlsberg Foundation.

1. You studied philosophy at the Ludwig Maximilian Universität in Munich, and then as a Ph.D. student at the University of Pittsburgh under the supervision of Wilfrid Sellars. How did you become fascinated by his philosophy and what role did his ideas play in your own work in philosophy?

JS: I became interested in Wilfrid Sellars' work when I was still in Munich and worked with Lorenz Puntel, a philosopher whom I hold in very high regard and who in my view, is too little known. He is one of the very few philosophers who has both overview over, and in-depth scholarly knowledge of, analytical and continental philosophy. He was in many ways my most important teacher in Munich and beyond. Due to his influence, I had from the beginning a focus on systematic philosophy, and I started writing my Master thesis on the problem of universals. I quickly found out that Wilfrid Sellars was the first, and so far perhaps only, philosopher who managed to develop a consistently nominalist or naturalist system in philosophy. So, I ended up writing a Masters thesis about Sellars rather than on universals. Later, I met Nicholas Rescher in Munich, and decided to become a visiting student in Pittsburgh. But as soon as I arrived in Pittsburgh I was so fascinated with the (non-sexist!) environment that I immediately applied for the PhD program. I became Wilfrid Sellars' last PhD student (after Sellars' death in 1989 I continued under the supervision of Nicholas Rescher). My dissertation was not on Sellars' philosophy, but on substance ontology, in order to lay the foundations for later work on a non-Whiteheadian process ontology.

Throughout my studies I was interested in a kind of a philosophical approach that begins with a conception of being that is in some fashion dynamic or active. Process philosophy, in that sense, was from the very beginning a major fascination. Then I realized that in order to establish a convincing process philosophy, one would need to begin with a presuppositional analysis of the ontological scheme that we have been using for 2000 years. I used a method that I call 'inductive ontology construction:' I looked at the contemporary discussion of standard ontological problems and identified certain presumptions that are shared by all participants of the debate, e.g., about which kind of ontological features (abstract, determinable etc.) can go



together and which cannot. Then I tried out some new feature combinations and investigated which ones would (a) solve the relevant problems and (b) still be acceptable as describing entities that speakers of natural languages can rationally think they are referring to. In this way one can better motivate the switch to process philosophy because one can show that certain definitions of entities that are afforded by process philosophy can help to resolve certain problems. The dissertation was, in essence, an analysis of the three big problems involved in the problem of identity in ontology: the problem of individuation, the problem of universals, and the problem of persistence. In each case I was able to show that the discussion of these problems in contemporary analytical ontology is driven by about 20 unreflected-upon presuppositions, which preclude a straightforward treatment of the task of defining numerical, qualitative, and temporal identity. Eventually, I showed that three of the so-called core problems of ontology arise only if we, due to theoretical habituations, restrict the way in which we construct ontological theories.

In the last chapter of the dissertation I argued for a fundamental revision of that research paradigm, and then showed in a sketch that once you give up on these 20 presuppositions you actually get a new ontological framework that satisfactorily deals with these three tasks.

Sellars was very sympathetic to this kind of project because he had also undertaken this sort of depth-structure analysis of ontological debates. There is one paper in particular that I found really inspiring, entitled "Grammar and Existence: A Preface to Ontology," that goes very, very deep into the presuppositions of "predicate-logical analysis," one of the standard tools of analytical philosophy. Together with Geach, Sellars was one of the first philosophers who identified unreflected-upon presuppositions in the very conception of a variable. The easiest way of illustrating this is to say that whenever we have a variable, x, that occurs in a formula with the 'existential quantifier,' so ' $\exists x \ (F \ x)$ ', we are used to reading this as 'there is an x such that x is F' or 'there is *something* that is F.' But not every entity needs to be a thing or thing-like. Geach argued at some point that we should read predicate variables in formulas such as: $(\exists f(f(a)))$ as 'somehow, there is somehow which a is.' Sellars developed a similar line of thought, saying that it is not only concrete particulars that we can quantify over, but also other kinds of entities, and that we must be careful about how we read quantified formulas, because the presuppositions introduced by our readings may distort the analysis. I often quote Sellars' pithy slogan summarizing this insight: "in logic we come always with dirty hands."



When Sellars died in 1989, I completed the dissertation with Nicholas Rescher as my first supervisor and Nuel Belnap as my second reader. Rescher had been sympathetic to process philosophy—he had written a short piece entitled "The process revolt" in the 1960s, and I believe my dissertation was an important inspiration for his later 1996 book *Introduction to Process Metaphysics*. (This booklet is a very effective introduction to the core intuitions of process philosophers, and it should be read in tandem with M. Weber's anthology *After Whitehead: Rescher on Process Metaphysics*, which contains constructive and critical interactions with Rescher's introduction from the perspective of other process metaphysicians.)

It was a wonderful, philosophically very intense and productive period of four years. Besides my dissertation I also wrote a book on Sellars' philosophy.

2. One recurring theme of your work concerns the theory called process ontology and the critique of the "myth of substance" or substance paradigm, namely the idea that the focus of ontological analysis should be on entities that appear to be static, unchanging, or persist through change. You have claimed that dynamic entities are entities in their own right and can be considered as the basic entities of our ontology. I wonder whether you can summarize the basic points of the process ontology that you propose and explain the task of an ontological account of dynamic beings.

JS: Let me try. We normally proceed from the assumption that the world is an assembly of things or objects, in the sense that these individual things or objects are unique "particulars," that is, they occur precisely in one place at any time at which they exist (unlike colors, or shape, or traffic etc.). So, thing-like objects can be individuated in terms of their location. This is practical, since then one can use a four-dimensional Cartesian coordinate system to refer to an object, and the door is open to a mathematical representation of the world. Almost all contemporary ontologists are committed to socalled *concrete particulars*. A concrete particular is something that is uniquely located, i.e., something that necessarily occurs only in one spatial region at any time at which it exists (in contrast to universals, which can exist multiply in space at a time). As long as one holds that all and only particulars (whatever has a unique location) are individual entities (a 'this' versus a 'that'), this implies that particulars are individuated in terms of their spatio-temporal location—unless one is prepared to embrace spooky 'individuating factors' ("haecceitas," "bare individuators") within a particular. I am breaking apart the traditional tie between individuality and particularity.



So. I claim that when we talk about concrete individuals we are not talking about something that is individuated in terms of their spatio-temporal location, but in terms of what it 'does' in the widest sense of the term—we are talking about something that is functionally individuated. A good illustration for such functionally individuated entities are the so-called "subjectless activities" (the term is by C.D Broad, later taken up by W. Sellars). When we say "It is raining" or "It is snowing" there is no subject that does the raining or snowing, we talk about a dynamic quality. My processes or dynamics are non-particular individuals—to signal this I speak of "general processes." They may be so specific that there is only one space-time region in which this activity or dynamic happens to occur. This is well-familiar from Leibniz—he also claimed that individuals are infima species, the ultimately specific general entities. For instance, I am such a specific activity ('Johanna-ing') and there is only one spatial region where the activity occurs now. But that is contingent, as in principle this sort of dynamic could actually be multiply occurrent in space. You might say now: but what have we gained if we switch from location-individuated (i.e. particular) individuals to functionally individuated (i.e., non-particular) activity-like individuals? One very good reason to admit activity-like individuals into our ontology is that we get a new solution to the long-standing problem of persistence: dynamics are individuals that can be literally recurrent in time, so, the activity of Johanna-ing or being-Johanna can literally recur as the same (qualitatively and numerically) at different moments in time.

"General Process Theory," the ontology that I have been developing, is a theory that describes the structure of the domain(s) we talk about in common sense and science. The entities of this structure are dynamics (or 'general processes') and they stand in many different relationships (causal dependencies, composition with and without emergence, etc.) The basic relation amongst dynamics is a part-whole relation. But the relationship among processes is not transitive—it does not hold that if "a is part of b" and "b is part of c," then "a is part of c". So, I cannot use classical mereology, the usual formal theory of part-whole relations, to define the relationships among dynamics. I had to develop a new, non-transitive mereology, called "Leveled Mereology," where parthood relations can be defined for, and up to, different levels.

According to General Process Theory (GPT), whatever there is, is dynamics, but there are many different types of dynamics. To distinguish among these different types, to build a typology of processes, I believe we need a five-dimensional classificatory framework. Relative to this framework, we can then determine precisely what it is that we are talking about



when we talk about a computer, a tree, jam, a traffic jam, an organism, the hurricane Katrina, symphonies, a person, a group of persons, a university, the color blue, water, friendship, the football World Cup, an autocatalytic reaction, self-maintaining systems, and perhaps even field quanta. That is, dynamics or processes are ontological counterparts for our common sense talk about simple and complex things, persons, events, stuff, masses, and for many scientific concepts.

When I talk about an "ontological counterpart" I have the following relationship in mind: The concepts of common sense and science are defined in terms of their inferential role within a conceptual scheme, and it is these inferences that ontological theories need to 'hook up with.' For example, a dynamics of type X is the ontological counterpart of a coffee cup if the definition of type X entails the inferences licensed by a sentence about a coffee cup in common sense reasoning, and in this sense explains why we are justified in hanging on to these inferences. A methodological parenthesis: There is much talk about 'grounding' at the moment in the journal discussion, facts (something with predicative structure!) are taken to be elements of reality (as opposed to the world of experience or reality-for-us), and 'metaphysics' and 'ontology' are conflated. All of this seems to me a return to pre-Kantian metaphilosophy, a regress compared to the methodological sophistication of analytical ontology in the first half of the 20th century—as far as I can see, this is a purely sociological phenomenon that we will, I hope, soon overcome.

General Process Theory has wide scope in two senses of the term. On the one hand, it has the potential to be an *integrated* ontology, that is, an ontology that can be used in order to interpret what we reason about in commonsense reasoning and in scientific reasoning. On the other hand, it can be used to interpret commonsense reasoning as formulated in many different. and possibly all, of the world languages. The ontological theories we have been working with—ontologies that assume that all concrete individuals are particulars—can be used to interpret the entities we talk about in Indo-European languages where inferences about spatiotemporal occurrences are built into the grammar of the language (e.g., via the article). But such particularist ontologies cannot be used to interpret sentences of many other of the world's 6400 or so languages, of which the Indo-European languages form only a small part. Since GPT operates with non-particular individuals, it holds out the prospect of an ontology with wide 'typological' scope, in the linguistic sense of the term—i.e., an ontology that can be used across languages with widely divergent grammars, from English to Chinese, Samoan, or Cayuga.



3. Your work involves not just philosophy but different areas of study as linguistics, sciences, and lately, robotics. Focusing on the last one, how did you decide to get involved in robo-philosophy? Could you give us a definition of robo-philosophy and explain the contribution of robo-philosophy to ethical reflections on robotics?

JS: As a process philosopher, I tried to understand in greater detail what an interaction is, focusing first on causal and emergent interactions. Some interactions are interactions of people, and then they come in various forms of collective actions, which are investigated in so-called Social Ontology and Action Theory. In 2007, as I started to investigate social interactions and conflict processes in more detail, just at the time when the first so-called social robot came onto the scene. Marco Nørskov, now a young colleague and close collaborator, approached me with a dissertation topic on the philosophy of social robotics. Marco and I realized increasingly in our joint discussions that the very notion of an artificial social agent challenges many fundamental assumptions in Western metaphysics and epistemology. So-called robo-ethics, that is, the reflection on social robotics from the point of view of ethics, needed, we found, a more fundamental, ontological investigation of precisely what is happening when a robot and a human interact. That created the idea that we need not only robo-ethics, but also theoretical reflections on our current conceptualizations of the phenomena of human-robot interaction and the larger cultural changes social robotics may entail. However, matters are in fact even more complex, since social robotics not only calls for philosophical reflection of the conceptual and cultural implications of social robotics—these clarifications are also of use for social robotics, and finally, since empirical research with social robots can also be used to do "experimental philosophy", I defined robo-philosophy as philosophy of, for, and by social robotics.

If I may, I'd briefly elaborate on this last variety of robo-philosophy, philosophy by social robotics. Social robotics offers the possibility of doing experimental philosophy in a new and perhaps methodologically more responsible way, since it is part of an integrated interdisciplinary approach and is supervised by researchers from empirical disciplines who have, unlike philosophers, the required training to do empirical (quantitative or qualitative) research. For example, as a philosopher, when I ask whether the philosophical criteria for the ascription of consciousness have all practical relevance—which capacities does an 'agent' need to display in order to be judged as 'conscious'? To what extent *should* philosophy, which is essen-



tially not a descriptive but a normative discipline—reconstructing conceptual norms of a certain conceptual scheme—, be influenced by such empirical data about how users of the conceptual scheme *de facto* implement these norms (or fail to do so)? This is another and difficult metaphilosophical question of course.

To adduce another example for the new methods of robo-philosophy, consider the dispute between utilitarians, deontologists, virtue ethicists, and care ethicists in so-called metaethics. Here, the core question is "Who has the better account of what is a morally good action?" If we now build two self-driving cars, one that uses a utilitarianist decision module and another one with a decision module that implements deontological principles, we can set up experiments with both cars in various decision making contexts to see which one makes the 'intuitively better' ethical decision in which context.

Robo-philosophy is also fascinating because it actually challenges us to develop a new theory of social interaction that gets rid of the traditional core assumption, namely that social interaction always requires consciousness and intentionality in both interaction partners. It forced us to pay attention to phenomena of what I call "asymmetric" sociality where only one interaction partner has these capacities, and to aspects of sociality that are below the level of consciousness (e.g., turn taking, or the creation of joint attention by gaze-cues, or reference fixing by pointing). In this way philosophy gets in touch with a more recent effort undertaken by researchers in linguistics, anthropology, and cognitive science to turn sociality into an interdisciplinary area of investigation.

There are also fascinating epistemological and phenomenological issues involved in robo-philosophy. For example, there is the question of the role of temporality in human cognition, the fact that we have a past, present, and future—how does this influence what we perceive of a scene and how we interact with other agents in the scene? What is the role of the human capacity of imagination for how we interact with objects and agents in our environment, and would artificial agents have to have (functional analogues of) temporality and imagination in order to be truly social agents?

Let me add here an interesting difference between philosophy of A.I. and philosophy of social robotics. Questions of A.I. discussed in philosophy so far are always focused on the capacities of the artifact. Social robotics introduces a new relational focus onto philosophy—it is no longer the capacities of the *relata*, or the interaction partners that is at the center, but the *interaction* as such. This recalibrates all questions of A.I.: here we begin with the interactions, and only then turn to the capacities. The question is no



longer whether agent X is conscious, intelligent, or has emotions. The question is whether the interaction between X and Y is one that is more or less asymmetrically conscious, intelligent, or emotionally involved. It is radically asymmetric if only Y has these capacities and otherwise merely projects such capacities onto X without any foundations for such a projection, and less asymmetric when X simulates the relevant capacities to a degree that justifies the projections relative to the empirical criteria that we use for each other (which are less demanding than philosophers think).

4. You are a member of the PENSOR group, a multidisciplinary research group based in Aarhus, that collaborates with the Hiroshi Ishiguro Lab (ATR) in Kyoto. You are also a member of the TRANSOR Network, a research network for transdisciplinary studies in social robotics. What are the main issues addressed by these two research groups? What are the positive aspects of interdisciplinarity and internationalization?

JS: The positive aspects of interdisciplinarity are very clearly that one has the feeling that one has a much better grasp on an extremely complex phenomenon such as social interaction. We often tend to forget that social interaction is the most complex phenomenon there is in the world—you need a host of disciplines—physics, biology, sociology, psychology, cognitive science, philosophy, and more—in order to describe fully what it is going on when two people are having a conversation, like the one we are having right now.

We all know, as researchers, that we are limited when we use the methodology and terminology about our own discipline, we sense our own limitations, and interdisciplinary research is a welcome opportunity to overcome these limitations. Of course, it also requires patience and humility—interdisciplinary communication is sometimes difficult, especially when it is compounded with cultural differences. But there are also immensely revealing, and ultimately positive aspects. This is what we experienced with our distinguished Japanese collaborators; sufficient reflectedness on both sides, openness, trust, and good will can create very productive ties.

I think that the PENSOR research group—an abbreviation for "Philosophical Enquires in Social Robotics," now we actually call ourselves the Resarch Unit for Robophilosophy (R.U.R.)—has been standing out for some time, since we combined from the beginning conceptual, phenomenological, qualitative, and quantitative research methods. In the new project that we have just launched, 25 researchers from 11 disciplines will collaborate. Our aim is to implement a new 'method paradigm' for social robotics, "Integra-



tive Social Robotics," by way of developing several new applications of social robots. Here, it will surely take some time before all members of the research team can understand each other's terminologies, but we have the ambition of working together towards the transdisciplinary integration of the field—roboticists currently produce robots, researchers in integrative social robotics will produce human-robot social interaction, using participatory design and the knowledge and methodologies from a wide scope of disciplines, including the Humanities.

As regards internationalization, social robotics will also open up a new angle on the benefits of multicultural research teams. The social robotics market will be global, no question, but roboticists write their cultural assumptions into the interactive *affordances* of the artifact they produce—that is, physical and kinematic features that invite or block human tendencies to interact with the robot in accordance with certain socio-cultural roles. So, we need to be able to understand cultural expectations and the link between interactive affordances with socio-cultural roles and ethical significances in greater detail on a global social robotics market. As a policy maker, for instance, I need to be clear on which interaction affordances a robot has for a specific application context—what is a sign of recognition in some cultural context might be interpreted as impoliteness in another. In short, social robots may repeat familiar cultural misunderstandings. On the other hand, there is also great innovative potential in introducing agents with unexpected behaviors.

5. I find the experiment you did with the other people in your research, group analyzing the interaction between Telenoid [a portable teleoperated android robot] and elderly people, very interesting. Do you plan to run more experiments with robots?

JS: We plan to run experiments with robots of different design and different interactive capabilities. Some robots look mechanical, others look like copies of human beings, the so-called *androids*—for example Hiroshi Ishiguro's *Geminoid* robot. In between these two extremes, there are robots that in their physical features or, in particular, in the way in which they move, invite people to treat them as social agents. There is much talk, in HRI (Human Robot Interaction Research) about a human tendency to "anthropomorphize" robots, which strikes me as wrong—as with a cat or dog, robots are not treated as human beings, but as social agents. We want to investigate how the interplay between the design of the robot and its behavioral repertoire can be used to enable social interactions with robots that humans expe-



rience as valuable—as enhancing their rationality or autonomy or creativity, etc. Often, we don't know whether the employment of a robot will have more positive or more negative effects, especially since we have no long-term studies so far. But there are some applications that are uncontroversially valuable, and we focus on these (I cannot say here what they are).

What distinguishes our approach, the approach of Integrative Social Robotics, from other research in HRI or social robotics is that we begin with values, not just with utilities.

6. How did you become interested in conflict resolution and global dialogue?

JS: My research interest in cultural conflicts, and intercultural dialogue as a way to prevent or transform them, was mainly triggered by the so-called cartoon crisis—an international crisis and culture clash created by the decision of a Danish newspaper to print cartoons that were offensive to Muslim readers. Cultural and ethnic conflicts are large-scale processes involving cultural or ethnic self-understanding that is often called (cultural or ethnic) 'identities.' The 'identity' idiom is quite unhelpful, however, since it suggests essential traits and persistent irreconcilability. In contrast, once we realize that our personal, social, ethnic, and cultural self-understanding is always dynamic, transitional, and continuously generated in a dialogue with what is 'other' to this self-understanding, it is easier to transform conflicts and to prevent them. That identities are a comfortable illusion of the lazy mind is a message that process philosophy has been trying to convey since its beginnings—but it is a difficult message that throughout history has drowned in the ever-rising simplicities of populism. Since Danish philosophers are not allowed to sit in the ivory tower, but need to justify their profession by showing its societal utility, I began to explore the intersection between the philosophy of dialogue, on the one hand, and dialogue as a method of conflict resolution on the other, and found both of them connected to recent cognitive science research on spatial orientation and orientation as a distinctive form of cognition. I found that philosophers of dialogue, notably Martin Buber, describe a type of searching cognition that we know from our spatial *orientation*. Similarly, conflict resolution by dialogue works insofar as both conflict parties reach the state of 'listening,' in a sense of the term that involves no longer perceptual judgments but another sort of taking in of another human being's reality that we can describe best as orientation. When we try to orientate ourselves, we are not yet performing classificatory judgment, we are taking in our environment; similarly, when we manage to



listen we withhold judgment for a while, taking in a space that contains the other.

So, triggered by the contingent incident of the cartoon crisis, I saw here a way of putting process philosophy to good use—I tried to explain why and how dialogical conflict transformation can be successful, based on an account of (personal, social, cultural) self-understanding that is established by the cognitional capacity of orientation, rather than on classificatory judgments.

7. One last question. Could you tell us what your current projects are?

JS: I want to finally finish a manuscript that offers a comprehensive exposition of General Process Theory and its applications. In connection with our research project in Integrative Social Robotics, I am finishing up an edition on *Robo-philosophy—Philosophy of, for, and by Social Robotics* (MIT Press), but I also hope to write a longer piece on the ontology of simulated sociality in human-robot interaction, and the responsibilities that we must bravely face in this new world, to twist a familiar title.

References

- Damholdt M. F., Nørskov M., Yamazaki R., Hakli R., Vesterager Hansen C., Vestergaard C., Seibt J., 2015, «Attitudinal Change in Elderly Citizens Toward Social Robots: The Role of Personality Traits and Beliefs about Robot Functionality», *Frontiers in Psychology*, 6.
- Rescher N., 1962, «The Revolt against Process», *Journal of Philosophy*, 62, 15, pp. 410-417.
- Rescher N., 1996, *Process Metaphysics: An Introduction to Process Philosophy*, SUNY Series in Philosphy, New York, State University of New York Press.
- Seibt J., 1990, Towards Process Ontology: A Critical Study in Substance-Ontological Premises, Tesi di Dottorato, University of Pittsburgh.
- Sellars W., 1960, «Grammar and Existence: A Preface to Ontology», *Mind*, 69, 276, pp. 499-533.



Weber M., 2004, *After Whitehead: Rescher on Process Metaphysics*, Frankfurt/Lancaster, Ontos Verlag.

Aphex.it è un periodico elettronico, registrazione nº ISSN 2036-9972. Il copyright degli articoli è libero. Chiunque può riprodurli. Unica condizione: mettere in evidenza che il testo riprodotto è tratto da www.aphex.it

Condizioni per riprodurre i materiali --> Tutti i materiali, i dati e le informazioni pubblicati all'interno di questo sito web sono "no copyright", nel senso che possono essere riprodotti, modificati, distribuiti, trasmessi, ripubblicati o in altro modo utilizzati, in tutto o in parte, senza il preventivo consenso di Aphex.it, a condizione che tali utilizzazioni avvengano per finalità di uso personale, studio, ricerca o comunque non commerciali e che sia citata la fonte attraverso la seguente dicitura, impressa in caratteri ben visibili: "www.aphex.it". Ove i materiali, dati o informazioni siano utilizzati in forma digitale, la citazione della fonte dovrà essere effettuata in modo da consentire un collegamento ipertestuale (link) alla home page www.aphex.it o alla pagina dalla quale i materiali, dati o informazioni sono tratti. In ogni caso, dell'avvenuta riproduzione, in forma analogica o digitale, dei materiali tratti da www.aphex.it dovrà essere data tempestiva comunicazione al seguente indirizzo (redazione@aphex.it), allegando, laddove possibile, copia elettronica dell'articolo in cui i materiali sono stati riprodotti.

In caso di citazione su materiale cartaceo è possibile citare il materiale pubblicato su Aphex.it come una rivista cartacea, indicando il numero in cui è stato pubblicato l'articolo e l'anno di pubblicazione riportato anche nell'intestazione del pdf. Esempio: Autore, *Titolo*, << www.aphex.it>>, 1 (2010).