Reconfiguring RD&D processes: Integrative Social Robotics

Christina Vestergaard
Research Unit for Robophilosophy and Integrative Social Robotics
Aarhus University
Aarhus, Denmark
etnocv@cas.au.dk

Malene Flensborg Damholdt
Dept. for Psychology and Dept for Health
Aarhus University
Aarhus, Denmark
ORCID: 0000-0001-9352-6703

Johanna Seibt
Research Unit for Robophilosophy and Integrative Social Robotics
Aarhus University
Aarhus, Denmark
ORCID: 0000-0003-3076-5912

***Abstract*—Integrative Social Robotics (ISR) is a full-scope interdisciplinary, value-driven 'RD&D paradigm: it reconfigures Research, Design, and Development processes in social robotics to ensure culturally sustainable applications or "positive social robotics." The research methodology of ISR promises advancements in HRI research, especially (a) with respect to a detailed modeling of the subtlety, complexity, and dynamics of human experience in interaction with 'social' robots, and (b) the descriptive and conceptual integration of HRI research. The paper summarizes central elements of ISR, and explains in greater detail the methodological execution of ISR, which involves a special mixed-method approach with iterations. We illustrate the method by describing how the integration of analytical perspectives and methods of different disciplines led to the development of a new model for understanding human sense-making of robots, "sociomorphing," replacing the more restricted model of "anthropomorphizing."**

***Keywords*—HRI methodology, Integrative Social Robotics, mixed methods, interdisciplinarity**

I. Introduction

Integrative Social Robotics (ISR) was introduced in 2016 as a new paradigm for the organization of the RD&D (research, design, and development) processes in social robotics that involves relevant expertise of the Humanities at all technology readiness levels. ISR shares basic motivations and methodological elements with other initiatives or so-called "design strategies"—such as "value-sensitive design", "design for values", "mutual shaping" "care-centered value-sensitive design" [1]. But there are also substantive differences, some of which we would like to present here for discussion. After a brief rehearsal of the main characteristics of ISR in section II, we will in section III briefly elaborate on and illustrate the interplay between quantitative, qualitative, phenomenological, conceptual/axiological, and constructive research within ISR, with focus on the specific form of a mixed-method approach used in ISR. In section IV we introduce four central notions of the descriptive framework used in ISR (OASIS) to support the conceptual integration of HRI.

II. The Five Principles of ISR

Social reality is arguably the most complex dynamic domain we are familiar with. Thus it is no surprise that 15 years of multidisciplinary HRI research have not yet delivered results that are sufficiently reliable and precise to imply legal regulations. What is surprising, however, is that expertise in the analysis and description of the symbolic and normative spaces of human social interactions has not been constitutively included in HRI research from the very beginning. This has generated a conceptually fragmented research landscape where quantitative methods are foregrounded that are unsuitable to trace behavioral or phenomenological dynamics, nor individual differences therein, and long-term effects are unknown. As we explained elsewhere [1, 2, 3], this methodological restriction gridlocks the productive cycle of research-based technology regulation, and ISR was developed to address this problem head-on by means of a new RD&D paradigm. ISR approach operates with five principles that ensure relevant conceptual and practical reconfigurations of the research, design, and development of applications of social robotics (whether these be performed under the label of "social robotics" or "HRI"):

1) The Process Principle: Social robotics produces not objects (robots) but processes—a new type of social interactions. The focus on processes (interactions) instead of objects (or actors or networks) restructures the data domain in fundamental ways, rendering human experiences constitutive ingredients of the target of the R&D process.—Since human experience, with its phenomenological and symbolic complexity and dynamics, is outside the scope of engineering and the behavioral sciences, further expertise is required:

2) The Quality Principle: The RD&D process must be carried by researchers of all disciplines that are directly relevant for the description and evaluation of the social interaction(s) involved in the envisaged application. This implies in particular the involvement of researchers from the social sciences and the Humanities, who can supply expertise in conceptual analysis, phenomenological analysis, the analysis of symbolic (socio-cultural) practices, and ethics.—In order to integrate all relevant expertise, RD&D processes for social robotics applications need to acknowledge the complexity of social interactions:

3) The Principle of Ontological Complexity: Any social interaction I is a composite of several (at least seven) realizations of interaction conceptions. These interaction conceptions represent how each agent (as well as an external observer) understands what they are engaged in from various perspectives, as we will further explain in section III. Here it is only important to note that on the ISR approach neither 'subjective' interpretations nor 'objective' observations can lay claim to the 'truth' of the interaction.

4) The Context Principle: The identity of any social interaction is relative to its (spatial, temporal, institutional, etc.) context. This implies that the RD&D process must operate with participatory design, involving all stakeholders, with frequent feedback loops, to arrive at a joint evaluative understanding of the new interactions introduced by the robot. This must be extended beyond 'technology placement' into the phase when 'new normality' has established itself.—The final principle of ISR, however, holds the key to the gradual unraveling of the mentioned current triple gridlock in research-based regulation.

5) The Values First Principle: Applications of social robotics must be developed in compliance with the Non-Replacement Maxim: social robots may only do what humans should but cannot do. Applications should preserve or maximize values relative to cultural axiological orderings. More precisely, the Non-replacement Principle says that robots may only afford social interactions that humans should do, relative to values V_i, but cannot do, relative to constraint C. The open formulation of the principle is intentional since it forces careful value analysis (empirical and conceptual axiology) and joint value deliberation at the beginning of the RD&D process, as well as throughout. ISR operates with a pragmatist conception of values as perceived aspects of interactions, thus value dynamics must be traced and normatively adjudicated by the ethicists in the developer team but also, see principle 4, all stakeholders.

Due to the fifth principle, ISR operates akin to a value-geared "design strategy", such as "value-sensitive design" or "design for values," investigating how putatively valuable ideas for applications fare in the field: are the new interactions afforded by the robot indeed, by all stakeholders, experienced as value-conducive as envisaged? There are, however, three features that set ISR apart from familiar value-geared participatory design-strategies for technology development.

First, in combination the demand for value maximization and the Non-Replacement Maxim in principle 5 amount to a *substantive normative restriction* already on the research targets of social robotics—simply put, as long as long-term effects are unknown, we should only investigate applications that are uncontroversially desirable.

Second, since the RD&D process includes axiological and ethical expertise, one can determine more precisely whether an application is *culturally sustainable* (value-preserving) or enhances central values in ways that could not be achieved without robotics, thus qualifying as *positive social robotics.*

Third, ISR is a *research methodology* that redirects HRI towards the topically required full interdisciplinarity and beyond. As we shall explain in the following two sections, the first four principles create a research environment that—unlike standard quantitative behavioral HRI research—allows us better to do justice to the dynamic complexity of social reality. In section III we sketch the concrete transformations of research practices generated by the ISR principles, and suggest that these are characteristic for the transition from interdisciplinary to transdisciplinary research.

III. THE RESEARCH METHODOLOGY OF ISR

The ISR approach has grown out of more than five years of research activities in HRI. The content (hypotheses, research designs) and the basic principles of the methodology were developed in the course of research interactions among researchers from 11 disciplines (anthropology, cognitive science, communication studies, conflict studies, engineering, linguistics, management, neuroscience, philosophy, psychology, sociology) using quantitative methods (experimental behavioral and neuroscientific research) qualitative methods (field research) and philosophical (phenomenological and conceptual) analysis. (A 'light' version of ISR was explored by another research team [4]).

The ISR paradigm stems from but extends the traditional uses of Mixed Methods that often combine quantitative and qualitative methods sequentially when collecting and analyzing data in a single study, based on the premise that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone [5, 6]. By contrast, it is the backbone of Mixed Methods within the ISR paradigm that all relevant disciplines contribute to the planning, execution, and analysis of empirical research. This is necessitated by the realization that the complexity of human interactions with so-called "social robots" is beyond the analytical scope of any one discipline—such interactions happen not only in the physical environment but as 'social' encounters of sorts occur in the *symbolic spaces* of human-social interactions, i.e., in the immensely complex domain of 'human social reality'.

More in detail, the mixed method approach we pursue in ISR is, first, a convergent design approach [5] where experimental and 'in the wild' research is conducted through quantitative and qualitative methods in the same timeframe. Second, both research methods in addition are informed by analytical categories of cognitive science and philosophical research on social cognition, collective intentionality, and social phenomenology. Third, while the more common sequential combination of quantitative and qualitative methods has one method feed into another, the disciplinary interaction in the ISR method is much more dynamic (see principle 4, requiring frequent feedback) and involves numerous disciplines that traditionally are not invited into development, execution, and analysis of empirical data.

According to classifications in science studies, this cotemporaneous involvement of different disciplines in all phases of the research process is the hallmark of the transition from multi-disciplinarity to interdisciplinarity: An inter-discipline, such as biomedical engineering, generates a new understanding of the domain and new modeling resources by the "integration of concepts, methods, materials, models" [7, p. 719]. Within the given research context of human-robot interactions, the ISR requirement that all relevant disciplines should be involved in interaction (principles 2 and 4) has the effect that the development of research design, questionnaires and interview guides also draws on relevant research for an in-depth analysis of human experience (i.e., research in philosophy (social phenomenology, collective intentionality, and social ontology). The ISR methodology thus combines the analysis of behavioral data (including self-reports) and neuroscience data, with interview data that can tease out the distinctive differences and dynamics of human experiential 'sense-making' in the encounters with social robots.

However, the integration afforded by the ISR approach leads beyond interdisciplinarity. As we have argued elsewhere [1], there are some indications that social robotics will take the format of pluridisciplinary collaboration that science studies call a 'transdiscipline': in a transdiscipline, such as integrative systems biology, "each field in the adaptive …problem space will likely penetrate and change significant practices in regions of the collaborating field" [7, p. 723]. The path towards transdisciplinary integration, where collaboratively gained research results impact the participating disciplines, goes via research practices and a type of investigative discourse that, curiously, in combination are best captured by the technical notion of 'dialogue' [8]—as joint exploration of a phenomenon that is unavailable to any single interlocutor. At the level or research practices this entails the commitment to 'listening' to other disciplines and their methods in order to understand more fully the phenomena under exploration, as well as epistemic humility concerning one's own discipline.

More concretely, the transition to greater interdisciplinary integration on the path to transdisciplinarity requires (a) *joint development of all material from experimental design to questionnaires, interview guides, analysis, and research papers*, but also the deep methodological conviction of productive interdependence (rather than mere supplementation): 'none of us could ask these questions on our own or produce the same analytical answers on our own'--'our disciplines act together not as sum but as a team'. (b) This awareness of the 'collaborative surplus' effect arises typically during *iterations on data analysis*. In fact, the second component of 'dialogical' research practices follows the model of "iterative process," familiar from the disciplines of engineering and design, where design, methods and analysis are continuously put in question to identify and overcome the limitations of one-track approaches and to pursue more ambitious solutions.

An example of how research and theoretical conceptualizations can sprout and evolve within this interdisciplinary milieu can be seen in our early discussions where we noticed that research on social robotics and HRI was all but forgetting about the 'human side' in human-robot interaction. This early realization led us to explore the human experiences of human robot interaction, not in the sense of "user experience" focused on instrumental aspects and safety, but with particular attention to the question of how humans experience the 'sociality' of their interactive engagements with 'social' robots. In these early iterations of our research dialogue, we developed two new questionnaires, the "ASOR: Attitudes towards Social Robots" questionnaire [9, 10], to gauge which characteristics are ascribed to robots: mental capacities, socio-practical capacities, and socio-moral capacities, and the "AMPH: Anthropomorphism questionnaire" [11,] to gauge the human tendency to anthropomorphize animals and artifacts vs. natural events. Both new assessment tools were developed in an interdisciplinary setting with the aim of getting a more fine-grained understanding of human experiences in human-robot interaction. Upon utilizing the questionnaires, we quickly discovered discrepancies in the results stemming from the questionnaire-based data and observational data: participants with seemingly low tendencies to anthropomorphize would nevertheless go through the motions or scripts of a social interaction with the robot. Thus, reliance on either of the two data sets alone (or given precedence of one data set over another) could have led us to either conclude that our participants did not anthropomorphize or that they did! Disengaging ourselves from analysis based on one discipline alone and discussing all available data using the discourse model of interdisciplinary dialogue brought us to new realizations: (a) Participants engaged in social interactions with the robot without 'anthropomorphizing', i.e., without projecting *human* capacities onto the robot. The triangulation of our methods, the quantitative studies using the two new questionnaires, as well as qualitative research with focus interviews, suggested that participants, while engaging with the robot in a social interaction, tried to accommodate the new experience, not by anthropomorphizing but by generating a new model of a sociality experience with the robot as a social other. We termed this process of interactive sense-making *sociomorphing*. Humans sociomorph a robot when they engage with it socially without projecting human capacities to it. The qualitative interviews revealed that participants were grappling with how to interpret /understand the robot, yet trying to engage with it following normal interaction rules, as suggested by the interview excerpts below:

- "Uhmm, well at first I didn't really know if there was any interaction…at some point I realized that it sort of reacts to what I am saying, so it was a weird experience, and I just had to adjust to talking to it."
- "But I wasn't sure, I am thinking: Is this how I should greet a robot? Will it think about it if I don't greet it or something like that?"
- "Mostly I was unsure of what I could say to it and how much it expected me to say to it."

These reports support the insight that conscious social agency inextricably involves multiple perspective-taking (see principle 3 above) --here humans were taking a 2nd-person perspective, looking at their own actions from the robots 'point of view.' We arrived at the hypothesis that different forms of sociomorphing result in different *types of experienced sociality* with robots, which can be descriptively anchored in different anticipations of the other's interactive capabilities [13]. In further cycles of the iterative process of the ISR methodology this hypothesis will be tested in new experimental designs that combine behavioral, neuroscientific, and qualitative research. If there are indeed significant correlations between types of experienced sociality and anticipated coordination capabilities of a 'social other', then the new analytical concepts of sociomorphing and types of experienced sociality amount to important additions to a descriptive framework that may facilitate the conceptual integration of HRI research, as we shall sketch in the following section.

IV. Towards Descriptive Integration

HRI displays terminological obstacles ranging from vagueness to dissociation to equivocation—with "social", "experience", "anthropomorphism", "presence," "simulation," "action" as glaring examples. The ISR approach, which calls for full scope inclusions of relevant disciplines ("Quality Principle"), might seem to exacerbate the problem. Instead, it is, we submit, the road to a solution. Within our research team we realized early on the need for a conceptual framework that would (i) allow for precise descriptions of human interactions with social robots and (ii) be accessible from many different disciplines. The first step towards conceptual integration consists in descriptive integration. Drawing on conceptual analysis, phenomenology, social and process ontology, we devised the descriptive framework OASIS (Ontology of Asymmetric Social Interaction) to articulate the insights gained from our empirical research, and to support the descriptive integration of HRI. The details of OASIS can be found in [12, 13], here we highlight three core elements.

Interaction matrix. According to OASIS the 'reality' of human interaction with social robots can only be captured by combining descriptions from at least seven perspectives, collected into a so-called "interaction matrix". In a 2-person human-human interaction each agent understands the action A she/he is performing from (i) the 1st person perspective as her/his intentional doing (e.g. 'I am holding out my hand'); (ii) from the 2nd person perspective (e.g. 'you will see offering a handshake'): (iii) from the 3rd person perspective of an imagined observer representing the normative conception of the interaction that A is a part of (e.g. 'I and you will be seen as engaged in greeting'). In addition, there is the perspective of a *de facto* external observer of the interaction between the two agents, who may be a passer-by (e.g. 'they are breaking the COVID regulations') or a researcher (e.g. 'they enter into the dynamics of interpersonal communication'). In human-robot interaction, the robot's three perspectives are the roboticist's, translated into the robot's functional, physical, and kinematic design.

Levels of sociality. The action descriptions in the interaction matrix are formulated in the (common-sense and scientific) idioms of ten different *levels of sociality*, relating to coordinative capacities ranging from socio-biological mechanisms (level 1) to conscious planning with mutual recognition and all registers of collective intentionality (level 10). Thus, intentional action descriptions in self-reports of participants (at levels 7-10) can be listed in the interaction matrix with external 3rd person descriptions using observational vocabulary from scientific disciplines focusing on lower levels of co-ordination capacities, e.g., 'mechanisms' of proxemics, acoustics, emotional alignment, etc. That social (inter-)actions can be attributed to coordination capacities at very different 'levels' matters if the other agent, B, (e.g. a robot or animal) is said to simulate a 'social action'. A robot may *realize* pre-conscious coordination mechanisms in a greeting but can only simulate intentional speech production.

Simulatory expansion matrix. Simulation can be defined as a structural similarity relation on processes, which in turn allows for distinctions of five degrees of simulation: approximating, displaying, mimicking, imitating, and functionally replicating. According to the current state of technology, the following correlation appears to hold: the higher the level of social coordination implies by an action description, the lower the degree of simulation of that action by a robot; and vice versa, low-level coordination capacities can be simulated to a high degree, or even—for coordinations due to evolutionary mechanisms—realized. The "simulatory expansion matrix" of an action A lists all degrees of simulation for each part of the action (e.g., a_1, a_2, a_3, implying coordination capacity at sociality levels 4, 1, and 7); thus two robot models which 'simulate' the same action in different ways, appear on different lines in the simulatory expansion matrix.

Altogether, the OASIS framework allows us (i) to 'read off' types of experienced sociality (in a specific interaction situation with a specific robot model), from 2nd person action description reported by the human agent, and (ii) to relate it to differentiated descriptions of the simulatory profile of the robot (in perspectival action descriptions four through six), and (iii) further relate it to the observational scientific descriptions of the interaction (perspective seven), focused on different levels of sociality (coordination capacities).

V. Conclusion

The approach of ISR was developed in order to improve the descriptive adequacy and depth of HRI research by including relevant expertise from the Humanities. As we tried to sketch here, summarizing more detailed expositions provided elsewhere, the research methodology of ISR holds out the promise to reconfigure a multidisciplinary research landscape into a transdiscipline. An important step along this path, which needs extensive further research, is the discovery of phenomena of "sociomorphing," i.e., the perception or practical treatment of something as social other (but not as *human* social other), supported by a complex and partly pre-conscious process of social cognition. How we sociomorph an entity, and the type of sociality we experience in interacting with it, depends on which coordination capacities we attribute to it when we take the 2nd person perspective on what we do. Since the forms of sociomorphing can be anchored in empirical research, the ISR approach offers new tools for the descriptive integration of HRI.

VI. References

[1] J. Seibt, M. F. Damholdt, and C. Vestergaard, "Integrative Social robotics, value-driven design, and transdisciplinarity," Interact. Stud., vol. 21, pp. 111–144, 2020.

[2] J. Seibt, "Integrative Social Robotics—A new method paradigm to solve the description problem and the regulation problem?," in What Social Robots Can Should Do, J. Seibt, M. Nøsrkov, and S. Andersen, Eds. Amsterdam: IOS, 2016, pp.104-114.

[3] J. Seibt, M.F. Damholdt, and C. Vestergaard, "Five principles of Integrative Social Robotics," in Envisioning Robots in Society--Power, Politics, and Public Space, M. Coeckelbergh, M. Funk, J. Loh, J. Seibt, and M. Nørskov, Eds. Amsterdam: IOS, 2018, pp. 28–43.

[4] K. Fischer, J. Seibt, J., R. Rodogno, M.K. Rasmussen, A. Weiss, L. Bodenhagen, W. Juel, and N. Krüger, "Integrative Social Robotics Hands-on,". Interaction Studies, vol. 21, pp. 145–185, 2020.

[5] J. W. Creswell, Qualitative, Quantitative and Mixed Methods Approaches. Newbury, CA: SAGE, 2014.

[6] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner, "Toward a definition of mixed methods research," J. Mix. Methods Res., vol. 1, pp. 112–133, 2007.

[7] N. Nercessian and W. Newstetter, "Interdisciplinarity in engineering research and learning.," in Cambridge Handbook of Engineering Education Research, A. Jori and B. Olds, Eds. Cambridge: Cambridge University Press, 2013, pp. 713–730.

[8] R. Anderson, L.A.Baxter, and K. N. Cissna, Eds. Dialogue-Theorizing Difference in Communication, Studies. Thousand Oaks, CA: Sage Publications, 2004.

[9] M. F. Damholdt et al., "Attitudinal Change in Elderly Citizens Toward Social Robots: The Role of Personality Traits and Beliefs About Robot Functionality," Front. Psychol., vol. 6, 2015.

[10] M. F. Damholdt, C. Vestergaard, M. Nørskov, R. Hakli, S. Larsen, and J. Seibt, "Towards a new scale for assessing attitudes towards social robots: The attitudes towards social robots scale (ASOR)," Interact. Stud., vol. 21, pp. 24–56, 2020.

[11] M. F. Damholdt, C. Vestergaard, and J. Seibt, "Testing for 'Anthropomorphization': A Case for Mixed Methods in Human-Robot Interaction," in Human-Robot Interaction: Evaluation Methods and Their Standardization, C. Jost, B. Pedevic, M .Grandgeorge et al, Eds. Cham: Springer, 2020, pp. 203–227.

[12] J. Seibt, C. Vestergaard, and M. F. Damholdt, "Sociomorphing, not anthropomorphizing," in Culturally Sustainable Social Robotics, M. Nørskov et al., Eds. Amsterdam: IOS, 2020, pp. 51–67, 2020.

[13] J. Seibt, "Classifying forms and modes of co-working in OASIS," in Envisioning Robots in Society, M. Coeckelbergh, et al, Eds, Amsterdam: IOS, 2018, pp. 133–146.