Article

Human-Robot Interaction in Groups: Methodological and Research Practices

Raquel Oliveira [1,2,*] **, Patrícia Arriaga** [1] **and Ana Paiva** [2,3]

[1] Centro de Intervenção e Investigação Social (CIS-Iscte), Iscte-Instituto Universitário de Lisboa (CIS-IUL), 1100 Lisbon, Portugal; Patricia.Arriaga@iscte-iul.pt
[2] INESC-ID (GAIPS), 1100 Lisbon, Portugal; ana.paiva@inesc-id.pt
[3] Instituto Superior Técnico, University of Lisbon, 1100 Lisbon, Portugal
[*] Correspondence: rsaoa@iscte-iul.pt

Abstract: Understanding the behavioral dynamics that underline human–robot interactions in groups remains one of the core challenges in social robotics research. However, despite a growing interest in this topic, there is still a lack of established and validated measures that allow researchers to analyze human–robot interactions in group scenarios; and very few that have been developed and tested specifically for research conducted in-the-wild. This is a problem because it hinders the development of general models of human–robot interaction, and makes the comprehension of the inner workings of the relational dynamics between humans and robots, in group contexts, significantly more difficult. In this paper, we aim to provide a reflection on the current state of research on human–robot interaction in small groups, as well as to outline directions for future research with an emphasis on methodological and transversal issues.

Keywords: human–robot interaction; group interactions; research methodology; social robotics

1. Introduction

If you look at the field of robotics today, you can say robots have been in the deepest oceans, they've been to Mars, you know? They've been all these places, but they're just now starting to come into your living room. Your living room is the final frontier for robots.

Cynthia Breazeal (Retrieved from https://cyberbotics.engineering.osu.edu/ (accessed on 16 June 2021).)

Living rooms and homes everywhere offer a fertile ground for the implementation of social robots, but also a particularly challenging one. The constant movement of people, both in physical and interpersonal terms, hardly offers a routine that is easy to choreograph in advance, and presents a level of unpredictability, inherent to social interaction, that is hard to factor in with today's technology.

In particular, group interactions are pervasive forms of social interaction that are at the core of our everyday life. Following the recognition of their importance, researchers in the field of Human-Robot Interaction (henceforth, HRI) and social robotics have been increasingly concerned about understanding the behavioral dynamics of groups of humans and robots, as reflected by the growth of published academic research on this topic in the last two decades (see Figure 1). Robots are no longer utopian machines of the future restricted to science fiction scenarios. Instead, they can now be found in schools [1,2], museums [3–5] and shopping malls [6–8] and many scholars believe that social robots hold the potential to further revolutionize the way we live and interact with each other [9].

In this context, keeping up with the fast pace of technology and the growing introduction of social robots in everyday contexts has become a complex task for researchers working in social robotics [9] .

To address this issue, in recent years, multidisciplinary teams have been working together to develop better and more comprehensive methodologies to assess HRI in groups.

Nonetheless, the measurement and analysis of small group interactions still presents a challenge due to their inherent complex and malleable nature.
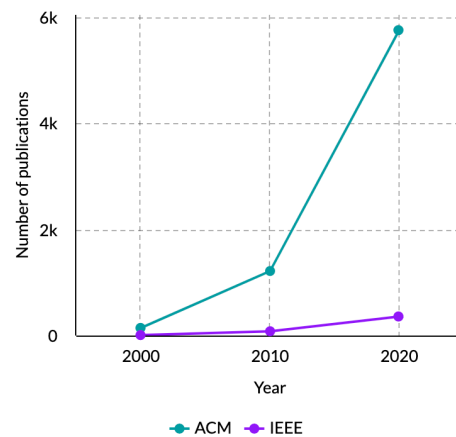
Figure 1. Cumulative growth in the number of publications indexed in ACM and IEEE resultant of the crossing of the keywords "group" or "team", and "social robot" in the abstract of papers published until 2020 (search last conducted on 16 June 2021).

Framework and Goals

In this article, we will start by providing a brief account of the psychological literature on groups (see Section 2) and link it to the current state of the art of the research on human–robot groups (Section 3). Furthermore, we will explore methodological and transversal concerns related to this field of research (Sections 4 and 5) and discuss potential avenues of development for future research (Section 6). For a schematic representation of our approach to these issues in this article, please consult Figure 2).

Our goal with this paper is to identify methodological and research-related (or transversal) issues, with a focus on HRI research, in which there is room for improvement; and to analyze how these current methodological and transversal shortcomings can impact the quality and reproducibility of the results of said research.

In this context, we must acknowledge that the research produced under the umbrella of social robotics is too heterogeneous and diverse to be generalized, and that, as a result, many articles in this area will have already addressed some of the issues outlined in this article. Similarly, we also acknowledge that many of the issues addressed in this paper are not unique to the field of HRI, and are present, to some degree, in research published in other areas. Nonetheless, we believe that the issues outlined in this paper are still present in published research in HRI, to an extent that justifies the analysis presented.

It is our goal to approach group HRI and the aforementioned issues through the specific lenses of psychology, and to analyse them, where possible, in regards to their specific impact on HRI research. With this paper, we seek to contribute to the field of HRI research by emphasizing some of the current shortcomings and challenges of social robotics research, and to outline some possible methodological avenues to address those challenges. In addition, we also reflect on some transversal (or research practices-related) concerns that underline research in HRI.

Although the definition of a group will be explored in-depth in Section 2, for the purpose of this article, we will consider group HRI to be any type of interaction between at least three group members who share a significant goal and that exert some type or degree of mutual influence over one another [10] (e.g., complete a task, play a game). These group members can include one or more social robots and/or one or more persons, and are not limited by the context they operate in (e.g., schools, museums, shopping malls).
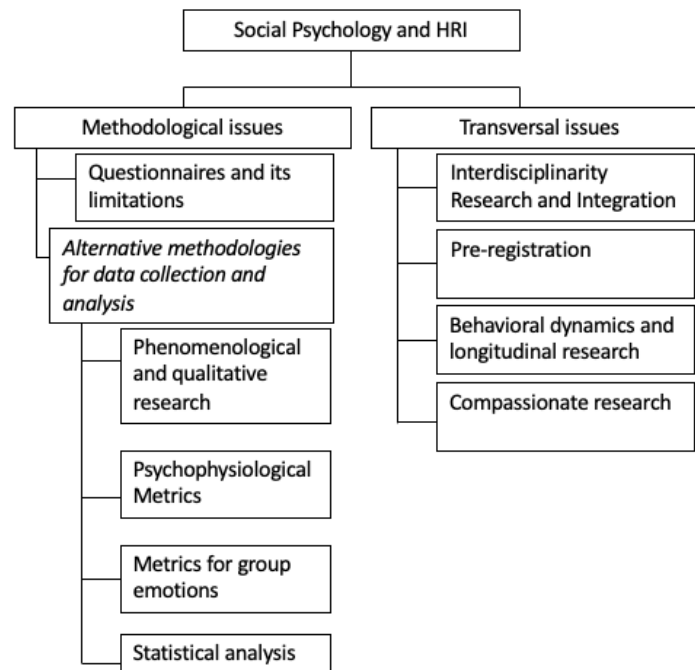
Figure 2. Schematic representation of the framework of this article.

A small group, for the purpose of this paper and in accordance with previous research, will be defined as any group that satisfies the aforementioned conditions [10], and that is composed of at least 3, and at the most 12 members (The definition of small groups in what regards their specific size has been fluid, and some authors suggest other limits for the definition of a small group (see, for example [11]).)

In addition, for the purpose of this paper, a social robot will be considered a socially embodied agent to which social interaction plays a key role [12]. This can include social robots which are used mainly to achieve social goals (e.g., provide company or conversation), but also robots which main functions rely heavily on its socially abilities (e.g., playing a competitive game). In this context, we will exclude industrial robots and robots who do not feature any communication abilities.

We would also like to emphasize that our goal is not to review or summarize the research in group HRI as a whole. An extensive review of important facets of group HRI was already published and we recommend its reading [13]. Instead, we seek to focus on the specific methods, methodologies and transversal issues that underline that research by taking a critical look at published research and outlining paths for future improvement.

2. Social Psychology: What Makes a Group?

A social group or team is more than a collection of people.

Imagine, for example, a collection of a dozen people waiting in line at a bus stop. Most would probably agree that these people do not actually constitute a social group or a team, as they are not significantly related to each other, nor do they share a significant common goal that requires their collaboration.

In this case, this collection of people is not a group because they lack *entitativity*, or, in other words, the perception by the group members and other people that those people together form a group. Entitativity is an important concept in this context, because it is a strong determinant of how we perceive and interact both with members of our ingroups and outgroups. For example, belonging to a group that is perceived as having a strong level of entitativity can help members face difficult circumstances [14] and achieve their psychological needs [15]. The level of entitativity can also affect the way people behave

towards outgroups. Paradoxically, people have been shown to be more xenophobic towards members of an outgroup that is perceived as having strong entitativity [16], but they have also been shown to be more generous towards members of groups with strong entitativity [17].

The level of entitativity, in turn, depends on a myriad of factors. For instance, similarity has been found to be positively associated with entitativity [18,19]. In fact, people often form or join groups precisely because they share significant similarities or goals with other members of that group. Perhaps they all enjoy playing card games, work together on the same project or all support the same football team; in other words, they all share something that *brings them together*.

Similarity, however, is not enough. Frequent interaction and communication also play an important role in increasing the entitativity of a group [20]. For instance, members of a workgroup are likely to be in frequent communication with each other and to share similar interests and goals, and thus to be considered a group.

In addition to frequent communication, members of a group are also likely to be interdependent to at least some degree, meaning that they need cooperation among group members to reach a specific goal [21]. For instance, a research team that wishes to build a social robot is likely to include individuals with different backgrounds (e.g., computer scientists, engineers, designers) who need to collaborate to achieve that goal.

Over time, groups are also likely to develop formal and informal group structures and rules, and assign different roles to each individual [22]. These formal and informal rules define what is acceptable and expected from each group member and, in general, are positively associated with the perception of entitativity.

In psychology, the study of group interactions has been roughly organized into six categories: (a) composition (*e.g., who is a member of the group and how does that affect the group dynamic?*), (b) structure (e.g., *does the group have a formal or informal structure?*), (c) performance (*how do the characteristics of a group affect its performance?*), (d) conflict (e.g., how does the group solve conflicts?), (e) the ecology of groups (e.g., how does the group interact with its environment?), and (f) intergroup relationships (e.g., how does one group interact with another?) [23–26].

3. Groups of Humans and Robots

Most research on HRI focused on examining one-to-one interactions between one person and one robot; however, the efforts that have been conducted to investigate HRI at a group level thus far provide convincing evidence that social robots can influence the dynamics of a group [27]. In particular, the presence or interaction with social robots can exert an influence on others in two significant ways: directly and indirectly (see Figure 3b,c).

Social robots exert a *direct influence in group dynamics when they are active participants in that group* (see Figure 3, where the group and direct interactions among their members are denoted in box b)), regardless of their role. For instance, social robots can be effective conflict moderators in groups of children [28] and adults [29]. Social robots can also positively impact the performance of members of a group in a collaborative task, and can increase the perceptions of group cohesiveness [30].

These effects on members of the group, however, seem to be influenced by the group composition and by the characteristics of the robot(s) (see Figure 3a). In particular, group size, for example, seems to influence behaviors towards robots in cooperative games, with groups of people displaying more competitive behaviors towards robots in this context, than towards other humans [31]. In addition, the group size (specifically, the number of robots) also seems to have an interaction effect with other important characteristics, such as the type of embodiment of the robot (anthropomorphic, zoomorphic, or mechanomorphic) [32].

The specific composition of the group and the robot-robot interactions within that group also matter. Robots that interact socially with each other are perceived as being more anthropomorphic. Moreover, groups of robots that are perceived as having a high level of

entitativity are perceived more positively and users report a higher degree of intention to interact with them in the future [33].

Before the interaction, groups (in comparison to isolate individuals) can also have an important effect in the initiation of the interaction and the level of trust assigned to the robot. For instance, although groups of people (as opposed to single individuals) are more likely to interact with a robot [34] and more likely to trust it [35], research also shows that in group interactions with robots, people pay less attention to the robot [13].

However, social robots can also *indirectly influence* the behavior of a group by causing a *ripple effect*. In this context, we consider not how the presence of one or more robots influences group dynamics, but instead, how it influences *the interactions that the people in the group have with other people* .

For instance, in studies involving autistic children, interaction with social robots has shown to have beneficial effects in how those children interact with their peers [36], therapists [37] and caregivers [13,38].
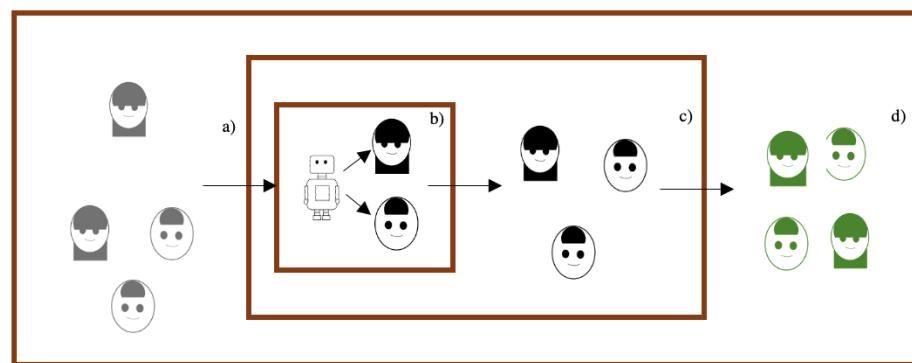
Figure 3. Schematic representation of the levels of analysis of HRI in groups: (**a**) group composition (e.g., size, individual characteristics of the members of the group), (**b**) direct interaction among robot(s) and members of the group, (**c**) indirect interaction among robot(s) and members of one group and members of another group and (**d**) interaction between mixed groups of humans and robots (which might have interacted directly or indirectly), and the wider society. The arrows represent unidirectional interactions. Please note that the gender distribution and colors implied by the icons is not significant, and was added to symbolize diversity within each group.

In the same line, research has shown that robot's expressions of vulnerability can have a beneficial impact on other member's of the group willingness to share their vulnerabilities among each other [39].

Finally, the introduction of social robots and the formation of mixed social groups is also likely to have an impact on their interaction with outgroups (see Figure 3d). For instance, in the context of a competitive game, participants have been shown to prefer and show fewer signs of aggression towards ingroup robots than outgroup humans [40,41]. Verbal support given by a robot to outgroup members seems to positively impact their participation in a joint task, but also reduces the verbal support given by other ingroup members [42]. In mixed group HRI, it has also been observed that the role and goal-orientation played by the robot can influence interaction among humans and robots, and also among humans who are members of an in-group or out-group [43] (see Figure 4 for an example of research on this topic). In addition, similar to what has also been observed in human-human interaction (see black sheep effect [44,45]) dissenting ingroup robots tend to be perceived less favourably than dissenting outgroup robots [46].
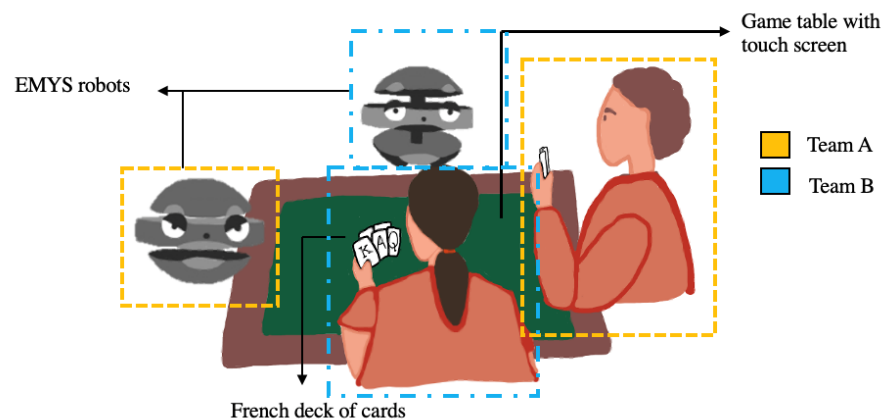
Figure 4. Representation of a scenario involving human–robot interaction in groups based on the work of Oliveira and colleagues [43]. A traditional Portuguese group card-game was implemented for a mixed group of humans and robots. The goal-orientation (collaborative vs. competitive) of robots was manipulated through the utterances they spoke, and the roles (partner vs. opponent) were manipulated through the sitting arrangement of participants (players sitting in front of one another played as partners, whereas players sitting to the sides were opponents). Interactions were recorded and coded according to a coding scheme used for group interactions [47].

In terms of the research in group HRI that has been conducted, a recent review has shown that most studies investigating group HRI involve scenarios where one robot interacts with two, three or more than four people [13]. Few studies investigated the interaction between one person or multiple people and more than one robot. Interestingly, this review also shows a good equilibrium between laboratory and field (or *in-the-wild*; approx. 54%) experiments, with a good portion involving autonomous robots [13]. This stands in contrast to the research conducted in other fields of HRI (e.g., [48]), in which we see a predominance of wizard-of-oz techniques.

4. HRI Research Methodology

In the area of social robotics, researchers have often borrowed concepts and methodologies from different areas of social sciences with the aim of improving research about human factors in HRI. The six categories presented at the end of Section 2 are, to some extent, present in HRI literature and present valid concerns for researchers investigating these types of interactions. Nonetheless, there are still some methodological issues standing in the way of the improvement and progress of HRI research on small groups. These limitations have been previously pointed out in several reviews concerning HRI (see, for example [13,49]).

Below, we provide a summary of these limitations, a discussion of their impact on our current understanding of HRI in groups and offer potential avenues to overcome them.

4.1. The Issue of Measurement

Assembling successful human–robot teams is no easy task. It requires the development of robots that can perform a myriad of social and functional tasks, and that can adapt and contribute to the establishment of healthy human environments [50,51]. In the context of HRI, collectively, research has demonstrated that robots can be effective teammates in mixed groups (i.e., groups involving more than one human and robot intervenients) [50].

Nonetheless, there are currently only a few metrics specifically developed to assess the functioning of groups and teams in HRI. Some of these quantitative metrics attempt to look at aspects of interaction that pertain to performance issues (e.g., [52,53]), whereas others focus on the measurement of the social aspects of interaction (see [54]). In this context, the need to develop more specialized measures (in the context of groups) has been pointed out by many authors [54,55] and remains a valid concern today.

To overcome this lack of specialized validated metrics for HRI in groups, researchers often opt to apply metrics developed for one-on-one interaction to group scenarios. However, the widespread use of metrics that were developed for the study of individual variables (such as robot perception [56,57]) in a group context might have several limitations. This is because these metrics are usually developed to capture information about an individuals' response to robots and can, as a result, fail to capture the effect of the *social situation* and of the dynamics that are created in this type of scenario. In this context, if we want to be able to model group emotions and other aspects of group interactions, to create robots that can interact in naturalistic ways, we need to gather high fidelity information on how groups of humans and robots interact.

One initial consideration that can help us address the challenges associated with the study of groups is to think of the different levels of analysis that have been associated with group research and to decide which one might be more useful in the context of HRI.

In this regard, social psychology traditionally distinguishes two levels of analysis: the individual-level approach and the group-level approach [58]. Researchers who argue in favor of the first tend to focus on the study of the individuals who compose the group [58]. This is in line with the approach that many studies in group HRI have taken and it defines group interactions as a collection of each individual group member' responses. However, social researchers who adopt the group-level approach tend to argue that groups are more than the sum of its parts in the sense that groups of people can produce behaviors and attitudes that none of its individual members would produce by themselves [58].

More recently, attempts to integrate these two approaches have given origin to more interactionist approaches that suggest that group behavior is both a result of the individuals' responses and the synergy between the individual and the group. This seems to be a particularly good approach because it focuses on the *group dynamics* and conceptualizes social groups as a system of reciprocal and ever-changing interactions between groups and individuals [58].

Assuming this perspective has the potential to enrich research in group HRI in many different ways. First, by enlarging our notion and concept of groups as a form of social interaction that is inherently distinct from interpersonal interactions. Second, by allowing the definition of a set of methodological tools that are developed to tackle different aspects of group interactions and that can be integrated to create a coherent picture of groups' dynamics. Finally, by providing researchers with a conceptual framework that might contribute to an improved understanding of the functional and social dynamics of human–robot teams and groups.

4.2. Moving Beyond Questionnaires

A summary of this section can be consulted in Table 1.

Surveys and questionnaires have always been considered a valuable method of data collection in social research [59]. They provide a transversal and straightforward way to collect information that is not only cheap, effective but also easy to report given its widespread use. However, despite its appeal, we must acknowledge that surveys offer only a small peek into an often complex and multi-layered reality.

Psychological tests or surveys are instruments that allow researchers to measure the psychological traits and states of individuals, and are used in a wide array of disciplines [60]. The results of these instruments are important because they lead researchers and other stakeholders to make decisions that can have outreaching consequences. However, the value of questionnaires is predicated on their psychometric qualities (namely, reliability and validity).

Reliability refers to the extent that a questionnaire can produce consistent and reproducible results (for a more in-depth explanation, see [60]). In this context, there are four main types of reliability that must be considered when developing or assessing the psychometric properties of questionnaires. Test-retest reliability refers to the extent that a survey can produce consistent results over a certain time interval. Interrater reliability refers

to the extent that individuals' observations of a certain phenomenon are consistent with other's observations, whereas interrater reliability refers to the extent that one individual's observations in two or more separate occasions are consistent. Internal reliability refers to the extent that individual's responses are reproducible and consistent across similar items of a scale.

Validity, on the other hand, refers to a scale's ability to measure what is intended to measure [60]. To assess validity, researchers typically look at three main types of validity: criterion, construct and content validity. Criterion validity, which can be concurrent or predictive, refers to the extent that one measure of a certain construct that is closely associated with a neighbour construct. Construct validity measures the extent to which the items of a scale accurately measure the underlying construct of interest and are thus positively correlated with other measures of the same construct (convergent validity) and negatively correlated with measures of unrelated or opposite constructs. Content validity, which is measured through the assessment of experts in the domain, evaluates the degree to which the behaviors, traits or beliefs included in the scale's items adequately cover the domain of interest.

In addition, to the psychometric properties of the scale, researchers must also take into consideration the quality of their study designs, namely by evaluating the study's internal and external validity [60]. Internal validity measures the extent to which the design of a study allows researchers to establish strong cause-and-effect inferences, whereas external validity refers to the extent to which the study design of a study allows the results to be generalized to the wider population of interest.

The over-reliance on questionnaires might threaten the external validity of findings and hinder their potential for generalization. For instance, questionnaires are often vulnerable to cultural influences, with scale translations and validations often presenting different structures of factors or different arrangements of items per factors (e.g., [61]).

In the specific context of HRI research, these issues take on heightened importance. For instance, given that the purpose of much of HRI research is to generate insights that are meaningful for those developing and implementing social robots in real-life contexts (i.e., "in-the-wild"), guaranteeing that the results of such research have external and internal validity is of paramount importance. In this context, there are many research-related aspects that can influence participants' responses and that introduce a source of bias in the data collected (e.g., researchers' characteristics, participants' desire to please the researcher or to be agreeable [62]).

Similarly, other aspects related to the participants' characteristics can also influence their responses to social robots. For example, factors such as the participants' prior interaction with robots (i.e., novelty effect; [2,63,64]), their a priori comfort and willingness to interact and accept new technologies or their attitudes towards robots and their introduction in society [65–67] are all non-interaction-related variables that can have an impact on how participants perceive the robots they interact with. These outside influences are often hard to control, and can be a source of contamination that influences the conclusions gathered from the research efforts made by researchers. However, these potential confounding variables are not unique to the use of questionnaires.

The presence of these biases, which is often transversal to different types of measures, emphasizes the importance of the triangulation, which has long been regarded as a reminder of the limitations of each individual method and has endured as one of the most satisfactory answers to this problem [68].

Methodological triangulation refers to the combination of different types of methods (qualitative and quantitative) to study one particular subject, thus allowing researchers to overcome the specific limitations of each method [68]. In group contexts, the use of triangulation is particularly important because it allows researchers to tackle the increased level of complexity present in this type of interaction .

Below, we summarize different methodological alternatives or complements that have been used broadly in social sciences and that can be implemented in research on group HRI.

4.2.1. Phenomenological and Other Types of Qualitative Research

A growing number of researchers has begun to emphasize the importance of incorporating user input into the development of social robots [69,70]. The reason behind this logic is that allowing potential users to have an active voice in the process of developing robots that are targeted at interacting with groups of people with similar characteristics to them, can be an important factor in creating technology that is adapted to the users' specific needs [70]. In a group interaction scenario, this is particularly important because it has been stated that perception is a social phenomenon and thus, an individual's perception of a robot can be affected, not only by the behavior and characteristics of the robot, but also by the behavior and perceptions of other people [43,59,71].

In this context and in congruence with the concerns presented in the previous section, qualitative and phenomenological research can be useful tools for researchers interested in unravelling the *meaning* behind patterns in quantitative data [59].

For example, focus groups with representative groups of a target population might yield useful information that the researcher or developer might have not considered in the first place. For example, in [70] the authors conducted a focus group to explore the perceptions of blind people regarding robots. In particular, the authors collected information regarding possible situations in which blind users thought robots could be useful and also what characteristics (e.g., size) robots should have. Then, based on the information collected, the researchers conducted an experiment with blind users using a task that involved some of the aspects (i.e., moving objects around and assembling things) mentioned by the participants in the focus group.

In addition, conducting focus groups (rather than individual interviews for example) might be more effective in generating novel and diverse information due to the fact that participants have the opportunity to build and develop each others' ideas [72].

Moreover, other qualitative techniques such as diary keeping, might be useful for researchers investigating factors related to the longitudinal effects (i.e., effects of the interaction with robots that outlive an initial interaction and that are durable in time) of robots in home or school-like environments [73,74]. This technique requires participants to keep diaries in which they describe their personal experiences when undergoing a specific treatment or experience. Content analysis might then be used to extract relevant information on a wide range of factors [73]. In particular, in group scenarios, it might be useful in understanding how participants in the study perceive the robots, other members of the group and how their perception evolves over time or when compared to specific events that can be introduced by the researcher [2].

Finally, techniques that involve the direct observation of group behavior can also be of importance. Borrowing observational coding schemes and group behavior models from other disciplines (e.g., the Interaction Process Analysis for small groups [43,47], can be a useful alternative to understand the specific nature of HRI in groups. In this line of thinking, the use of these observational tools can be particularly useful for those interested in analyzing the content of interactions as well as its distribution in time and in different tasks (e.g., entertainment or problem-solving interactions).

Although this technique has its limitations (see [75]), we believe it can be a useful addition to the methodological toolbox of researchers interested in studying HRI in groups.

Table 1. Summary of data collection methods, some of their respective advantages and shortcomings and examples of application to HRI research problems.

Data Collection Method	Advantages	Shortcomings	Example of Application
Questionnaires	-Cost-friendly [76];	-Questionnaire fatigue [76,77];	A research team is interested in evaluating whether the level of competence displayed by a robot in a group competitive task influences participant's perceptions of the robot and their willingness to interact with the robot again. They can employ pre-developed questionnaires (e.g., RoSAS [56] for the perception) or create ad hoc questions (e.g., [78], for the willingness to interact again in the future).
	-Easy application [76];	-Lack of nuance [77];	
	-Widely used (familiar) [76,79];	-Accessibility issues [77];	
	-Scalability [77]	-If not properly validated and evaluated, can produce biased data [77].	
Focus groups	-Less time consuming than other similar methods (interviews; [76]);	-More time consuming than questionnaires [80];	Researchers intend to develop a robot for therapy, so they consult a group of experts (therapists) to get their feedback about key development issues (e.g., [81]).
	-Allows the exploration and in-depth discussion of important topics [76,80];	-Researcher has less control over the data generated [76,82];	
	-It can reach many participants simultaneously [82].	-Data can be difficult to analyze and interpret [82].	
Dary-keeping	-Allows us to see how participants' perceptions evolve over time [83,84];	-If the goals are not well-defined and transmitted to participants, relevant information might not be recorded [83];	Researchers are interested in evaluating the acceptance and user's opinions about a social robot that has been implemented in the users' home (e.g., [85]).
	-Experiences and opinions are recorded closer to when they happen, and not in hindsight [84];	-Participants might not be motivated to journal frequently [86];	
	-Allows us to capture external factors that can influence user's feelings and opinions [84,86].	-Data can be difficult to interpret and analyze [84,86].	
Interviews	-Allows exploration of user's opinions, feelings and experiences [87];	-Time consuming [87,88];	Researchers seek to develop a social robot that can help blind users with daily tasks, so they conduct interviews with blind users, in which they obtain their feedback about desired functionalities (e.g., [70]).
	-It provides flexibility in the topics explored [87]	-Interviewers must be trained and develop an interview script a priori [87,88];	
	-Interviewer can take into account the non-verbal behavior of the interviewee [76].	-Can be costly due to the need for having dedicated facilities (i.e., rooms) and the possible need for dislocation to meet participants [76,87].	

Table 1. *Cont.*

Data Collection Method	Advantages	Shortcomings	Example of Application
Observation of behaviors	-Allows direct observation of behavior;	-Internal states (e.g., motivation for a specific behavior) are not observable [89];	A research team is interested in understanding how different roles (partner vs. opponent) and different goal-orientations (competitive vs. collaborative) can influence group interactions in entertainment settings [43];
	-Allows for the accounting of different types of behavior (e.g., verbal, non-verbal [89]);	-Can be costly and time-consuming [90,91];	
	-Allows us to collect data from several individuals simultaneously [90,91].	-It can result in a large amount of data that can be difficult to analyze and interpret [89,90].	
Psychophysiological metrics	-Allows for real-time data recording [92];	-Requires very specific expertise to collect and analyze data [92];	A research team wants to implement context-sensitive robotic behaviors according to participants' level of anxiety, thus achieving improved implicit communication between user and robot [93].
	-Psychophysiological responses are not under the voluntary control of participants, so they are difficult to fake or manipulated [92];	-Can be costly given that they require specific apparatus and tools;	
	-Humans are not always accurate in making judgements about their cognitive or internal states (e.g., [94]).	-Collection of phychophysiological data can feel intrusive to the participant [92].	

4.2.2. Psychophysiological Metrics

Physiological metrics (e.g., heart-rate variance, skin conductance) have often been used in psychology to measure individuals' bodily reactions to stimuli. In general, physiological indicators provide an account of the degree of arousal, and can be influenced by psychological constructs [92,95]. This is thought to be a good way to measure people's responses because it provides an alternative to self-report measures (that can be biased) and it allows researchers to assess certain aspects of social cognition and emotions that are not always accessible to the individuals (e.g., reaction times).

To evaluate each individual response within the group, the group dynamics and the associated emotional, cognitive and behavioral responses, verbal and nonverbal responses may be recorded and their coding could be facilitated by the use of multi-modal sensors for real-time and off-line data collection and analysis.

Thus, depending on the research questions, relevant measures for the assessment of group dynamics in HRI may include non-invasive psychophysiology sensors to detect specific emotional responses or other emotional processes, including indexes of stress and emotional regulation through the evaluation of heart rate variability, complemented with electrodermal activity, respiration (see, for example, [96]).

Other non-invasive biomarkers such as salivary hormones of cortisol, testosterone, and/or oxytocin could be used to complement emotional and behavioral responses. For example, cortisol release has frequently been measured in humans to evaluate their responses to social stress events [97]. Testosterone levels seem to increase when individuals anticipate conflict and competitive situations [98,99]. In contrast, oxytocin has been related to social bonding and attachment, cooperation, trust, and several other measures of prosociality (e.g., [100–102]).

In HRI research, these variables have been measured primarily through questionnaires (e.g., [103–105]), and thus, offer interesting areas for the employment of alternative data collection methodologies.

In addition, eye-tracking technologies allow the assessment of several eye gaze metrics to capture and understand approach and avoidance behavior, including the time spend looking at each member of the group, the avoidance of contact, which can be of interest to study HRI in group scenarios. This technique has been adopted before (e.g., [106,107]); however, its use can be extended to other contexts of group interactions.

Another advantage of these types of measures is that they allow the continuous recording and assessment of the individual's responses (as opposed to their posterior assessment, achieved, for instance, through the application of questionnaires), and the fact that they can be used in contexts in which the application of other measures could be disruptive of the real-time experimental task [92]. This can be particularly relevant for HRI research due to the importance of creating naturalistic interaction experiences among humans and robots [108].

Despite the fact that the collection and analysis of some of these metrics (e.g., heart rate variability) requires very specialized knowledge and a higher level of discomfort for the participants (in comparison to other metrics, such as surveys), this methodology has the potential to significantly improve our knowledge of HRI in groups by complementing and advancing previous findings.

Moreover, similar to other methods of data collection, bio-physiological measures also present some limitations, particularly the difficulty in making inferences regarding covert states based on psychophysiological data and the fuzzy patterns of the psychophysiological responses associated with some emotions (for more information, see [92,109]). As such, these metrics provide the most value if used in complement with other measures (both physiological and non-physiological), and if interpreted within the contextual framing in which they occur [92].

4.2.3. Metrics for the Analysis of Group Emotions

Physiological measures can be useful tools in analyzing variables related to groups' emotional processes and dynamics. However, other, more non-invasive measurements of emotions (for example, through the use of emotion recognition software; for a more in-depth review of this topic and its limitations, see [110–112]), can also provide an adequate way to analyze the role of emotions in HRI in groups.

This type of method would allow researchers to measure and collect information in real-time about the responses of individuals within the group, as well as other non-verbal responses. In this context, some authors have already begun to incorporate these techniques in HRI research to categorize variables such as facial emotional expressions [113] and voice [114], (see [115] for a review).

Nonetheless, the employment of these techniques also has limitations that should be taken into consideration. In particular, they often rely on the recognition of facial markers that are observed in the prototypical manifestation of certain emotions [116,117]. However, because people do not always express emotions in the same way (especially, when considering more complex emotions), the use of these tools can be complemented by the use of human coders or by the analysis of other indicators (e.g., body language).

In contrast with bio-physiological responses, other behavioral responses which are mostly under the voluntary control of the individual can also provide an interesting source of information regarding an individual's responses to a certain stimulus. However, like any other measure, they need to be interpreted according to the specific situation and cultural context in which they occur [95].

In addition, although there have been attempts to map specific patterns of physiological responses associated with specific emotions, researchers still need to rely on several simultaneous physiological and other subjective measures to ensure reliability [95].

4.3. Towards Improved Statistical Methods

The statistical and methodological advancements witnessed in recent decades in social sciences have been phenomenal [118]. More specifically, in the context of group interactions, some authors suggested that groups should be regarded as complex systems that interact with smaller systems (e.g., the members), as well as other systems equally larger (e.g., other groups) or even larger than themselves (e.g., the society around them). In addition, groups also tend to have fuzzy boundaries that can simultaneously distinguish them and connect them to other groups and individuals around them [23,118] (see Figure 5).

In this context, the statistical techniques we use to analyze this type of interaction must be able to adequately mirror this complexity and thus, need to be different from the statistical techniques used to analyze other types of interaction [119]. Ignoring this complexity during the process of data analysis can lead to a higher risk of incurring in type I errors by increasing the likelihood of obtaining spurious significant results [119].

One technique that has been consistently pointed out as being a useful tool to accommodate this increased complexity is multi-level modelling (MLM) [120,121]. MLM can be useful for those exploring the effects of an intervention in a group's behavior and it allows the researcher to analyze these effects on an individual, group and organizational or cultural level, simultaneously [120]. In addition, it provides an alternative to traditional statistical methods that assume the independence of observations (e.g., t and F tests), which is often not the case in scenarios that involve group interactions (for more information, see [121,122]).

Similarly, other nested approaches for the analysis of certain group behavior dimensions (such as behavior duration) have been suggested and can be useful tools for future research on multi-party HRI. MLM, for example, has been used in the past in HRI research on small groups (e.g., [43]), however, its use is still not widespread enough for it not to be mentioned here as a *future trend*, rather than a current one.
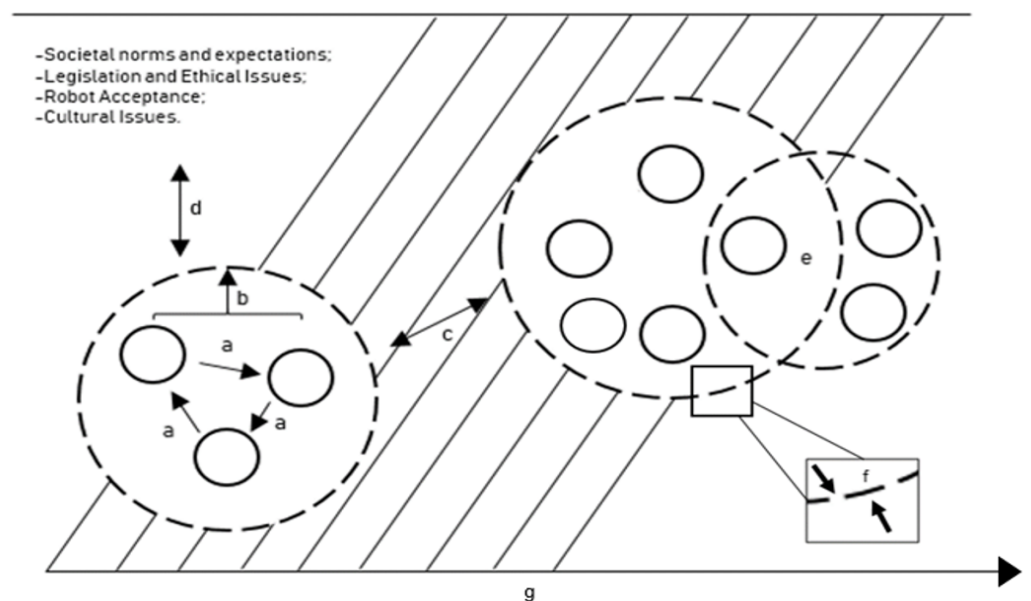
Figure 5. Schematic representation of the complexity and different relational levels of group interactions, including (**a**) interactions between members of a group, (**b**) interaction between a group and its members; (**c**) intergroup interactions; (**d**) relationship between the group and the wider social environment; (**e**) overlapping group membership; (**f**) representation of the fuzzy boundaries of social groups and (**g**) the progress in time of group processes.

Furthermore, in line with some of the issues on behavioral dynamics already pointed out, there is also a growing interest in statistical techniques that allow researchers to analyze the temporal sequencing of events in a more interactive way (rather than through the traditional input-progress-output model) [59]. This, among other things, would allow researchers to study and develop probabilistic models of turn-taking in social HRI in groups (i.e., predict *who* is more likely to intervene next and *what* situations facilitate different types of interactions) [123].

5. Transversal Concerns

A summary of this section can be consulted in Table 2.

Although methodological quality is the cornerstone of valid, reliable and useful research, research involves much more than the methods it employs. The research practices that underline the research conducted in any academic field can also have important consequences to the quality of the work produced. In this context, we define transversal concerns as concerns related to the research practices that underline academic research efforts. These transversal concerns, some of which will be explored more in-depth in the following sections, can include aspects related to the management of different interdisciplinary contributions and perspectives, aspects related to the reproducibility of the research output and the importance of the collaboration among social robotics research centres spread around the world.

Table 2. Summary of the main transversal concerns discussed in this paper, and their relation to the field of HRI research in terms of the advantages and challenges they present.

Transversal Concerns	Importance for HRI Research	Main Challenges
Interdisciplinarity	-Research on human psychology and other social sciences can be a good starting point for research in HRI;	-Communication between academics of different fields can difficult;
	-Research methods common in social sciences can be used to improve HRI research;	-Collaborations might be hard to establish due to a lack of network opportunities with academics from other fields.
	-It offers new sources of insight and new perspectives that can be beneficial to the development of social robots.	
Pre-registration	-Increases the transparency, rigor and reproducibility of research;	-Requires a substantial amount of time to be dedicated to the planning and study preparation process;
	Reduces bias and opportunities for dysfunctional research practices;	
	-Puts an emphasis on the careful planning of important aspects (e.g., data collection methods, sample size estimation) of research studies;	
	Pre-registered reports that are subject to peer-review can increase the likelihood of the publication of negative or null results.	
Longitudinal research	-Presents an opportunity to better understand how group HRI develops over time;	-Can be costly to implement and monitor, both in terms of time and money;
	-Large-scale longitudinal studies "in-the-wild" offer useful insight on how to better develop social robots suited for this type of interactions.	-It might be difficult to keep participants engaged in the research process for such lengths of time.
Compassionate research	-Contributes to the development of better social robots by focusing on the needs of prospective users;	-It can be difficult to reach and conduct research with some groups of users;
	It can increase the societal value of social robots by making them more valuable and useful to users.	-Users' needs and expectations of social robots can vary widely across culture and demographics.

5.1. Interdisciplinarity Research and Integration

HRI is a multi-disciplinary area of investigation that includes the contributions of engineers, social scientists and other researchers interested in developing social autonomous robots. One particular area that has largely contributed to the development of many studies in HRI in groups is the area of social psychology. This discipline can be broadly defined as the study of the impact of the presence of others (whether it is real, imagined or perceived) on our behavior [71] and it has been at the core of the development of many theoretic and conceptual models on human groups' behavior (e.g., social cognitive approach, for a review of these theories, see [124]).

For this reason, psychology and other disciplines concerned with collective behavior, such as sociology or political science, can serve as a good starting point for those interested in HRI in groups. Although the study of groups itself is a multi-disciplinary discipline and one that is characterized by the existence of multiple different perspectives (e.g., sociocultural, social cognitive), it is usually agreed that group behavior differs significantly from individual behavior in many instances [125]. For example, in one of the most influential series of experiments in psychology, Asch [126] demonstrated the effect of conformity to the majority in group situations, and how this effect was amplified by the size of the group (In the mentioned experiment, Asch [126] surrounded participants with a group of other people, who, unbeknownst to the participants, were confederates (actors). He then presented an image of lines with different lenghts, and asked the group (one naïve participant and the confederates) to identify which line matched a target line in terms of length. Although the answer was obvious, participants still conformed to the opinion of the group of confederates (which was wrong) more than 70% of times. This experiment has been replicated in HRI research, confirming its existence among human members of a group, but not in mixed groups [127]).

To study the specificity of group behavior, several authors came up with a multitude of different models and theories (for a review, see [125]) that organize themselves around factors such as social identity (i.e., focus on the relations between social groups, by considering social processes such as stereotyping and intergroup conflicts [128]), distribution of power (i.e., focus on the distribution of power and resources among unequal members of a group, which looks at phenomenon such as negotiation and consensus building [129]) and functionality (i.e., focus on understanding the aspects that influence the effectiveness of a group, by considering factors such as the internal structure of the group, the characteristics of the task and the environment in which the group interacts [130]) .

Although these approaches differ from each other to the extent they focus on different factors to explain group behavior (e.g., social identity [128], distribution of power [129], functionality [130]), all of them provide useful lenses through which to look at group interactions. Indeed, despite the overlap in the study of different topics (i.e., the same topic can be analyzed from different perspectives), all of these conceptual perspectives (e.g., sociocultural, evolutionary, social learning, social-cognitive) highlight different facets of the same phenomena and can be useful for guiding research on HRI [125]. For example, in psychology, collaboration and competition have been both looked at from a sociocultural perspective (which puts an emphasis on social norms and culture) and an evolutionary perspective (which emphasizes the role of genetics and inheritance in the development of social behaviors) [125]. Despite the focus on different aspects, these two approaches both provide complementary hypothesis about how and when individuals and groups choose to behave collaboratively or competitively.

In methodological terms, this interdisciplinarity implies a broad conjugation of methods that stem from research on different academic areas (e.g., ethnographic research) and of different conceptual perspectives (e.g., sociocultural). This, in turn, allows the existence of a more varied menu of methodological tools that can be used to improve research in HRI in groups and supports the future establishment of more interdisciplinary research projects and works.

5.2. Pre-Registration

Pre-registration has the potential to increase the transparency, rigor and reproducibility of published research by decreasing existing biases, motivations and opportunities for dysfunctional research practices [131,132]. Currently, some online platforms for preregistration of studies already exist, offering a variety of templates that can be used by researchers to establish a priori what their goals, hypotheses, data collection and analytic strategies will be (For instance, clinicaltrial.gov, osf.io, aspredicted.org, and PROSPERO (https://www.crd.york.ac.uk/prospero/) for health-related systematic reviews of literature.).

Similarly, owing to the recognition of the importance of pre-registration of studies, some academic journals have also already begun to offer the possibility of submitting pre-registered reports (For instance, Nature offers the possibility to submit pre-registration reports: https://www.nature.com/nathumbehav/registeredreports (accessed on 7 July 2021). This type of submission typically involves the peer-review of study protocols that detail, before the start of data collection, all relevant details regarding the goals, hypotheses, methods and data analysis plan. These records can then be *provisionally accepted*, and later published (given that the authors adequately follow their pre-registered plan), regardless of the results.

This system of pre-registration is, thus, important for many reasons [131,132]. First, as stated at the beginning of this section, because it helps reduce biases, motivations and opportunities for dysfunctional research practices. These can include postdiction (i.e., *"...the use of data to generate hypotheses about why something occurred..."*, p. 2600, [131]) which can be motivated by the desire to be published and the simultaneous opportunity to do so granted by the a priori lack of commitment to any set of predictions or hypotheses [131].

Second, pre-registered reports are also important because they might contribute to reducing publication bias. Publication bias refers to the tendency to prefer studies that

yield significant results (as opposed to nonsignificant or null results) for publication, and it has been a known tendency since at least the 1980s [133], being present in several fields of science [134]. The review and conditional acceptance of pre-registered reports for studies involving HRI, thus, presents a seductive way forward and solution to the problem of publication bias.

In the same line, pre-registration is also widely acknowledged as offering important contributions to efforts related to increasing the reproducibility of research [134]. The importance of reproducibility has been emphasized in the last decade, with several international teams of scientists conducting replication efforts on several influential studies across different fields of study. Of these efforts, however, emerged disappointing results, with estimates for successful reproductions ranging between 11% and 50% [135–139]. Although to the best of our knowledge, no large-scale efforts to investigate the reproducibility of HRI studies exist, pre-registration and the a priori review of study protocols can, as in other areas, increase reproducibility and help generate better quality research.

Another issue with current HRI research that can affect its validity and reproducibility regards the employment of adequate methods of sampling and sample (i.e., pool of participants) sizes. Indeed, sampling and sample size are crucial aspects of quantitative research in general, which seeks to make statistics-based generalizations for a wider population. In this context, it is important, on one hand, to guarantee that the sample employed is representative of the target population that is deemed to be of interest, by controlling for aspects that have been shown to impact HRI. These aspects can include, for instance, culture, prior interaction [140] and attitudes towards robots [141].

On the other hand, it is also important that the sample size used is sufficient to detect the effect sizes expected in any given experiment. Using adequate sample sizes reduces the chance of type I and type II errors, and thus is an important concern to have in mind when conducting quality research [142]. Some tools have already been developed for this purpose. These tools allow researchers to calculate sample sizes according to the expected effect size of their independent variables, as well as their study designs and methods for data analysis. These tools include programs such as GPower (GPower is available for download here: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower (accessed on 7 July 2021)) [143,144], and guidelines for sample size estimation that have been advanced by other authors (e.g., [145–148]). However, further development and refinement of sample size estimation techniques for the context of group interactions is still necessary [149,150], as many of the existing tools focus on an individual level of analysis.

5.3. Behavioral Dynamics and Longitudinal Research

The importance of analyzing how interactions unfold over time has been recently identified as one of the major issues in small groups research [2,59]. This concern emerges as an answer to the increased level of complexity present in group scenarios and has the potential to yield important insights on how groups appear, develop and change as time passes.

In this context, longitudinal methods present an exciting opportunity to address the issue of how groups of humans and robots behave over time, how they adapt to new circumstances and create bonds within each group or team [2]. This is particularly relevant because groups often have fuzzy boundaries (see Figure 3), thus allowing members to join or leave the group as time progresses. However, this is not the only time-bound process in group interactions. In the context of group HRI, others might include the analysis of how engagement with the robot varies over time, how trust is developed, adjusted and maintained or how the exploration of humans and robots can learn, overcome obstacles and developed attachment relationships over time.

These concerns are congruent to those presented by other authors stating the dangers of generalizing conclusions yielded by single-interaction studies. Aylett [151], for example, warns that studies involving short interactions, with participants that do not interact regu-

larly with robots, can cause an atypical pattern of behavior. In the same line, the novelty effect has been consistently identified as one of those factors that can endanger the generalization of the results of a study and must be controlled for [152]. In addition, we must also consider if one-time short interactions in mixed groups are sufficient for those groups to develop any level of entitativity, as they lack many of the factors that are important to create that sense of *groupiness* (e.g., frequent interactions or communications, similarities) [153]. Furthermore, the employment of longitudinal measures can also be useful when analyzing the transient nature of individual gains (e.g., training and improving a new skill) achieved through interacting with robots.

This issue can be addressed by investing and designing large-scale longitudinal studies that allow researchers to measure the effect of HRI across time and assess the stages of *development* and *growth* of these interactions.

Despite its many benefits, longitudinal research is often neglected because of its costs (both financial, logistical and time-related) and because of the difficulty in keeping participants engaged for such long periods of time. Nonetheless, it remains a necessary step towards a better understanding of HRI that must be undertaken if we want to respond to the question of how social robots have the potential to affect people's lives over time and thus, justify the importance of its introduction in social contexts.

5.4. Compassionate Research

Part of the transversal argument that justifies the research, development, and creation of social robots is that developing better, more socially effective robots, can ease their acceptance and help improve people's lives. Indeed, whether it is in educational contexts, helping children learn a new subject or in care contexts, helping people with disabilities to achieve a higher level of autonomy, robots are developed because they give people some kind of advantage.

Compassionate research, a recent trend in social sciences, argues for the development of research that is grounded in (and motivated by) the need to help other people and in the desire to improve their lives [154]. Although this concern might be perceived to be universal to all areas of HRI (and perhaps all areas of human studies), it is particularly important here due to the pervasiveness of group interactions and their effect on social and emotional well-being.

In the context of group interactions among people, we know that there are several interventions that are based on the power of groups and the social support provided by them. Support groups for various traumatic experiences, as well as for other day-to-day activities are very important in peoples' lives in the sense that they allow them to connect to others, contextualize their experience and share their emotions. Because more and more robots are being employed in care contexts and working with special populations, this seems like a promising future field for research that can be supported by the employment of this framework.

6. Discussion and Future Endeavours

Our goal has been to put forward some thoughts for consideration regarding the advancement and future of small group research in HRI, with a focus on methodological issues. As a broad discipline, social HRI has emerged in the past few years as an exciting field of research that triggers the interest of academics from many different backgrounds. Nonetheless, we believe that there is still much to do with regard to the study of human–robot groups. In this context, we see groups as *complex, adaptive, dynamic systems* (see Figure 5), often embedded in hierarchical structures and involving multiple simultaneous bi-directional and non-linear causal relations [24,26]. Groups also do not constitute isolated or static entities. They are intricate, require constant mutual adaptation and operate through processes that unfold and change through time [25,26,155].

To tackle these complex issues, researchers should have available a set of methodologies that allow them to tackle this complexity without adopting a reductionist approach.

By developing and applying sound methodologies, researchers will be better equipped to solve problems and develop robots that are suited for group interactions.

Parker and colleagues suggested that *"innovation in theory needs to be matched by innovation in method"* ([156], p. 434, emphasis added). While there is still a lot to do towards the development of consistent theories of HRI in groups, that development might be aided by the creation of adequate metrics of research and mixed-method methodologies. To do so, we must focus on strengthening the methodological aspects upon which we support the validity of our findings and, ultimately, the guidelines we draw from the literature on how to build better robots.

For this purpose, we would like to call in more work on the analysis, development, and application of metrics in social HRI in groups; both in regards to the human factor in HRI, but also in regards to the measurement and evaluation of robot performance. In this paper, we suggest some possible paths and future methodological trends that have the potential to aid in the process of measuring human behavior in situations that involve HRI, by increasing the broadness of ways we can look at and measure HRI in groups. Although these metrics are not specific to HRI, but instead, borrowed from other disciplines that attempt to explore the different characteristics of human behavior, they can still be useful resources for researchers in social robotics, to the extent that they offer new methodological perspectives. In this context, and to the best of our abilities, we tried to enrich the text by providing examples of how these alternative data collection strategies and methodologies could be relevant to the field of HRI research in specific.

In summary, with this article, we sought to contribute to the advancement of research in HRI by outlining some of the challenges currently present in this field, and by proposing alternative and under-explored methodologies that could greatly benefit the quality of research in this field going forward.

In this context, our main recommendations regarding data collection, methodologies and data analysis for future research are:

- Triangulation of different types of measures is key in avoiding biases that can influence the data collected or the researchers' interpretation of it;
- Some of the alternatives to questionnaires in terms of data collection include, for instance, psychophysiological metrics, focus groups, journal-keeping, observation and codification of behaviours occurring during HRI in groups;
- Developing, evaluating and validating instruments, particularly in the context of group HRI research, is fundamental if we want to ensure that we are measuring what we intend to be measuring and that our results are valid and generalizable to the population of interest;
- The application of adequate statistical analyses (e.g., multi-level modelling) is necessary to capture the complexity and dynamic nature of group interactions.

In terms of the transversal concerns (i.e., research-related practices) explored in this article, the main take-home messages include:

- Recognizing the interdisciplinarity of the research conducted under the umbrella of social robotics implies developing good management and integration strategies, that allow the inputs and insights of different fields of knowledge to be leveraged productively;
- Following the recognition of the importance of interdisciplinarity and the relevance of exploring the cultural specificities of group HRI, it becomes important to establish multi-country (and multi-lab) collaborations that can result in large-scale research projects;
- The spread of pre-registration practices and the sharing of data among researchers are important factors to ensure the reproducibility, transparency and rigour of the research produced;
- Increasing the efforts to investigate long-term and in-the-wild HRI in groups is fundamental for a better comprehension of how these relations initiate and develop

over time (in other words, we emphasize the importance of considering *human–robot relations*, as opposed to *human–robot interactions).*

- We emphasize the importance of conducting compassionate research that is motivated primarily by the needs of potential users, in order to better leverage the social potential of social robots.

Although these suggestions and considerations are not exhaustive, in this article, we sought to provide a starting point for the further discussion of how methodological and transversal issues can impact research in group HRI. In this context, we seek to add value to this field of research by recognizing some of the challenges that researchers in HRI face today when conducting their research and draw attention to potential alternatives that can improve their research.

References

1. Alemi, M.; Meghdari, A.; Ghazisaedy, M. Employing humanoid robots for teaching English language in Iranian junior high-schools. *Int. J. Hum. Robot.* **2014**, *11*, 1450022.
2. Leite, I.; Castellano, G.; Pereira, A.; Martinho, C.; Paiva, A. Long-term interactions with empathic robots: Evaluating perceived support in children. In *International Conference on Social Robotics*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 298–307.
3. Fuentes-Moraleda, L.; Lafuente-Ibañez, C.; Alvarez, N.F.; Villace-Molinero, T. Willingness to accept social robots in museums: An exploratory factor analysis according to visitor profile. *Libr. Hi Tech* **2021**, doi:10.1108/LHT-07-2020-0180.
4. Yamazaki, A.; Yamazaki, K.; Burdelski, M.; Kuno, Y.; Fukushima, M. Coordination of verbal and non-verbal actions in human–robot interaction at museums and exhibitions. *J. Pragmat.* **2010**, *42*, 2398–2414.
5. Pang, W.C.; Wong, C.Y.; Seet, G. Exploring the use of robots for museum settings and for learning heritage languages and cultures at the chinese heritage centre. *Presence Teleoperators Virtual Environ.* **2018**, *26*, 420–435.
6. Aaltonen, I.; Arvola, A.; Heikkilä, P.; Lammi, H. Hello Pepper, may I tickle you? Children's and adults' responses to an entertainment robot at a shopping mall. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 53–54.
7. Niemelä, M.; Heikkilä, P.; Lammi, H.; Oksman, V. A social robot in a shopping mall: Studies on acceptance and stakeholder expectations. In *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 119–144.
8. Niemelä, M.; Heikkilä, P.; Lammi, H. A social service robot in a shopping mall: Expectations of the management, retailers and consumers. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 227–228.
9. Share, P.; Pender, J. Preparing for a robot future? Social professions, social robotics and the challenges ahead. *Ir. J. Appl. Soc. Stud.* **2018**, *18*, 4.
10. Wilson, G.L.; Hanna, M.S. *Groups in Context: Leadership and Participation in Small Groups*; McGraw-Hill: New York, NY, USA, 1990.
11. James, J. A preliminary study of the size determinant in small group interaction. *Am. Sociol. Rev.* **1951**, *16*, 474–477.
12. Fong, T.; Nourbakhsh, I.; Dautenhahn, K. A survey of socially interactive robots. *Robot. Auton. Syst.* **2003**, *42*, 143–166.
13. Sebo, S.; Stoll, B.; Scassellati, B.; Jung, M.F. Robots in groups and teams: A literature review. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–36.
14. Bougie, E.; Usborne, E.; de la Sablonniere, R.; Taylor, D.M. The cultural narratives of Francophone and Anglophone Quebecers: Using a historical perspective to explore the relationships among collective relative deprivation, in-group entitativity, and collective esteem. *Br. J. Soc. Psychol.* **2011**, *50*, 726–746.
15. Crawford, M.T.; Salaman, L. Entitativity, identity, and the fulfilment of psychological needs. *J. Exp. Soc. Psychol.* **2012**, *48*, 726–730.
16. Ommundsen, R.; Yakushko, O.; Van der Veer, K.; Ulleberg, P. Exploring the relationships between fear-related xenophobia, perceptions of out-group entitativity, and social contact in Norway. *Psychol. Rep.* **2013**, *112*, 109–124.

17. Smith, R.W.; Faro, D.; Burson, K.A. More for the many: The influence of entitativity on charitable giving. *J. Consum. Res.* **2013**, *39*, 961–976.

18. Crump, S.A.; Hamilton, D.L.; Sherman, S.J.; Lickel, B.; Thakkar, V. Group entitativity and similarity: Their differing patterns in perceptions of groups. *Eur. J. Soc. Psychol.* **2010**, *40*, 1212–1230.

19. Lickel, B.; Hamilton, D.L.; Wieczorkowska, G.; Lewis, A.; Sherman, S.J.; Uhles, A.N. Varieties of groups and the perception of group entitativity. *J. Personal. Soc. Psychol.* **2000**, *78*, 223.

20. Igarashi, T.; Kashima, Y. Perceived entitativity of social networks. *J. Exp. Soc. Psychol.* **2011**, *47*, 1048–1058.

21. Brewer, M.B.; Hong, Y.; Li, Q. Dynamic entitativity. *Psychol. Group Percept.* **2004**, *19*.

22. Forsyth, D. Group Dynamics 5th ed Belmont CA Wadsworth. *Cengage Learn.* **2010**. Available online: https://www.worldcat.org/title/group-dynamics/oclc/882092375 (accessed on 30 September 2021).

23. Wittenbaum, G.M.; Moreland, R.L. Small-Group Research in Social Psychology: Topics and Trends over Time. *Soc. Personal. Psychol. Compass* **2008**, *2*, 187–203.

24. Levine, J.M.; Moreland, R.L. Progress in small group research. *Annu. Rev. Psychol.* **1990**, *41*, 585–634.

25. Levine, J.M.; Moreland, R.L. Small Groups: An Overview. *Key Readings Soc. Psychol. Press.* **1998**. Available online: https://psycnet.apa.org/record/2006-12496-001 (accessed on 30 September 2021)

26. Moreland, R.L.; Hogg, M.A.; Hains, S.C. Back to the future: Social psychological research on groups. *J. Exp. Soc. Psychol.* **1994**, *30*, 527–555.

27. Jung, M.; Hinds, P. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.* **2018**, *7*, 1–5. doi:10.1145/3208975.

28. Shen, S.; Slovak, P.; Jung, M.F. "Stop. I See a Conflict Happening." A Robot Mediator for Young Children's Interpersonal Conflict Resolution. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 69–77.

29. Jung, M.F.; Martelaro, N.; Hinds, P.J. Using robots to moderate team conflict: The case of repairing violations. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015; pp. 229–236.

30. Short, E.; Mataric, M.J. Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions. In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Lisbon, Portugal, 28 August–1 September 2017; pp. 385–390.

31. Chang, W.L.; White, J.P.; Park, J.; Holm, A.; Šabanović, S. The effect of group size on people's attitudes and cooperative behaviors toward robots in interactive gameplay. In Proceedings of the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 845–850.

32. Fraune, M.R.; Sherrin, S.; Sabanović, S.; Smith, E.R. Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015; pp. 109–116.

33. Fraune, M.R.; Oisted, B.C.; Sembrowski, C.E.; Gates, K.A.; Krupp, M.M.; Šabanović, S. Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. *Comput. Hum. Behav.* **2020**, *105*, 106220.

34. Gockley, R.; Forlizzi, J.; Simmons, R. Interactions with a moody robot. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, Salt Lake City, UA, USA, 2–3 March 2006; pp. 186–193.

35. Booth, S.; Tompkin, J.; Pfister, H.; Waldo, J.; Gajos, K.; Nagpal, R. Piggybacking robots: Human–robot overtrust in university dormitory security. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 426–434.

36. Kim, E.S.; Berkovits, L.D.; Bernier, E.P.; Leyzberg, D.; Shic, F.; Paul, R.; Scassellati, B. Social robots as embedded reinforcers of social behavior in children with autism. *J. Autism Dev. Disord.* **2013**, *43*, 1038–1049.

37. Zubrycki, I.; Granosik, G. Understanding therapists' needs and attitudes towards robotic support. The roboterapia project. *Int. J. Soc. Robot.* **2016**, *8*, 553–563.

38. Scassellati, B.; Boccanfuso, L.; Huang, C.M.; Mademtzi, M.; Qin, M.; Salomons, N.; Ventola, P.; Shic, F. Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* **2018**, *3*, doi:10.1126/scirobotics.aat7544.

39. Strohkorb Sebo, S.; Traeger, M.; Jung, M.; Scassellati, B. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human–robot teams. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 178–186.

40. Fraune, M.R.; Šabanović, S.; Smith, E.R. Teammates first: Favoring ingroup robots over outgroup humans. In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), , Lisbon, Portugal, 28 August–1 September 2017 2017; pp. 1432–1437.

41. Fraune, M.R.; Šabanović, S.; Smith, E.R. Some are more equal than others: Ingroup robots gain some but not all benefits of team membership. *Interact. Stud.* **2020**, *21*, 303–328.

42. Sebo, S.; Dong, L.L.; Chang, N.; Lewkowicz, M.; Schutzman, M.; Scassellati, B. The Influence of Robot Verbal Support on Human Team Members: Encouraging Outgroup Contributions and Suppressing Ingroup Supportive Behavior. *Front. Psychol.* **2020**, *11*, 3584.

43. Oliveira, R.; Arriaga, P.; Alves-Oliveira, P.; Correia, F.; Petisca, S.; Paiva, A. Friends or Foes?: Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 279–288.

44. Marques, J.M.; Yzerbyt, V.Y.; Leyens, J.P. The "black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. *Eur. J. Soc. Psychol.* **1988**, *18*, 1–16.

45. Pinto, I.R.; Marques, J.M.; Levine, J.M.; Abrams, D. Membership status and subjective group dynamics: Who triggers the black sheep effect? *J. Personal. Soc. Psychol.* **2010**, *99*, 107.

46. Steain, A.; Stanton, C.J.; Stevens, C.J. The black sheep effect: The case of the deviant ingroup robot. *PLoS ONE* **2019**, *14*, e0222975.

47. Bales, R.F. *Interaction Process Analysis*; A Method for the Study of Small Groups: 1950. Available online: https://psycnet.apa.org/record/1950-04553-000 (accessed on 30 September 2021).

48. Oliveira, R.; Arriaga, P.; Santos, F.P.; Mascarenhas, S.; Paiva, A. Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Comput. Hum. Behav.* **2020**, *114*, 106547.

49. De Visser, E.J.; Peeters, M.M.; Jung, M.F.; Kohn, S.; Shaw, T.H.; Pak, R.; Neerincx, M.A. Towards a theory of longitudinal trust calibration in human–robot teams. *Int. J. Soc. Robot.* **2020**, *12*, 459–478.

50. Groom, V.; Nass, C. Can robots be teammates?: Benchmarks in human–robot teams. *Interact. Stud.* **2007**, *8*, 483–500.

51. Fong, T.; Kunz, C.; Hiatt, L.M.; Bugajska, M. The human–robot interaction operating system. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, Salt Lake City, UT, USA, 2–3 March 2006; pp. 41–48.

52. Kannan, B.; Parker, L.E. Fault-tolerance based metrics for evaluating system performance in multi-robot teams. In Proceedings of the Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD, USA, 14–16 August 2006.

53. Balakirsky, S.; Scrapper, C.; Carpin, S.; Lewis, M. USARSim: Providing a framework for multi-robot performance evaluation. In Proceedings of the PerMIS. 2006. Available online: https://www.nist.gov/publications/usarsim-providing-framework-multi-robot-performance-evaluation (accessed on 30 September 2021).

54. Pina, P.; Cummings, M.; Crandall, J.; Della Penna, M. Identifying generalizable metric classes to evaluate human–robot teams. In Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI 2008, Amsterdam, The Netherlands, 12–15 March 2008; pp. 13–20.

55. Burke, J.; Lineberry, M.; Pratt, K.S.; Taing, M.; Murphy, R.; Day, B. Toward developing hri metrics for teams: Pilot testing in the field. In Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI 2008, Amsterdam, The Netherlands, 12–15 March 2008; p. 21.

56. Carpinella, C.M.; Wyman, A.B.; Perez, M.A.; Stroessner, S.J. The robotic social attributes scale (rosas): Development and validation. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 254–262.

57. Weiss, A.; Bartneck, C. Meta analysis of the usage of the Godspeed Questionnaire Series. In Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 31 August–4 September 2015; pp. 381–388.

58. Stangor, C. *Social Groups in Action and Interaction*; Routledge: 2015. Available online: https://www.routledge.com/Social-Groups-in-Action-and-Interaction-2nd-Edition/Stangor/p/book/9781848726925 (accessed on 30 September 2021).

59. Keyton, J. The future of small group research. *Small Group Res.* **2016**, *47*, 134–154.

60. Furr, R.; Bacharach, V. Psychometrics and the importance of psychological measurement. In *Psychometrics*; Sage Publications Inc.: Thousand Oaks, CA, USA, 2008.

61. Piçarra, N.; Giger, J.C.; Pochwatko, G.; Gonçalves, G. Validation of the Portuguese version of the Negative Attitudes towards Robots Scale. *Eur. Rev. Appl. Psychol.* **2015**, *65*, 93–104.

62. Miyazaki, A.D.; Taylor, K.A. Researcher interaction biases and business ethics research: Respondent reactions to researcher characteristics. *J. Bus. Ethics* **2008**, *81*, 779–795.

63. Smedegaard, C.V. Reframing the role of novelty within social HRI: From noise to information. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea, 11–14 March 2019; pp. 411–420.

64. Vogt, P.; van den Berghe, R.; de Haas, M.; Hoffman, L.; Kanero, J.; Mamus, E.; Montanier, J.M.; Oranç, C.; Oudgenoeg-Paz, O.; García, D.H.; et al. Second language tutoring using social robots: A large-scale study. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea, 11–14 March 2019; pp. 497–505.

65. De Graaf, M.M.; Allouch, S.B. Exploring influencing variables for the acceptance of social robots. *Robot. Auton. Syst.* **2013**, *61*, 1476–1486.

66. Hameed, I.A.; Tan, Z.H.; Thomsen, N.B.; Duan, X. User acceptance of social robots. In Proceedings of the Ninth International Conference on Advances in Computer-Human Interactions (ACHI 2016), Venice, Italy, 24–28 April 2016; pp. 274–279.

67. Naneva, S.; Sarda Gou, M.; Webb, T.L.; Prescott, T.J. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *Int. J. Soc. Robot.* **2020**, *12*, 1179–1201.

68. Denzin, N.K. Triangulation 2.0. *J. Mixed Methods Res.* **2012**, *6*, 80–88.

69. Kawamura, K.; Pack, R.T.; Bishay, M.; Iskarous, M. Design philosophy for service robots. *Robot. Auton. Syst.* **1996**, *18*, 109–116.

70. Bonani, M.; Oliveira, R.; Correia, F.; Rodrigues, A.; Guerreiro, T.; Paiva, A. What My Eyes Ca not See, A Robot Can Show Me: Exploring the Collaboration Between Blind People and Robots. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 15–27.

71. Allport, F.H. The group fallacy in relation to social science. *Am. J. Sociol.* **1924**, *29*, 688–706.

72. Morgan, D.L. *Focus Groups as Qualitative Research*; Sage Publications: London, UK, 1996; Volume 16.

73. Bolger, N.; Davis, A.; Rafaeli, E. Diary methods: Capturing life as it is lived. *Annu. Rev. Psychol.* **2003**, *54*, 579–616.

74. Gunthert, K.C.; Wenze, S.J. *Handbook of Research Methods for Studying Daily Life*; Guilford Press: New York, NY, USA, 2012.

75. Cucu Oancea, O. Using diaries-a real challenge for the social scientist. *Soc. Behav. Sci.* **2013**, *92*, 231–238.

76. Williamson, C. Questionnaires, individual interviews and focus groups. In *Research Methods: Information, Systems, and Contexts*; Tilde University Press: Melbourne, Australia, 2013; pp. 349–372.

77. Patten, M. *Questionnaire Research: A Practical Guide*; Routledge: 2016. Available online: https://www.routledge.com/Questionnaire-Research-A-Practical-Guide/Patten/p/book/9781936523313 (accessed on 30 September 2021)

78. Oliveira, R.; Arriaga, P.; Correia, F.; Paiva, A. The stereotype content model applied to human–robot interactions in groups. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea, 11–14 March 2019; pp. 123–132.

79. Jones, S.; Murphy, F.; Edwards, M.; James, J. Doing things differently: Advantages and disadvantages of web questionnaires. *Nurse Res.* **2008**, *15*, 15–26.

80. Mansell, I.; Bennett, G.; Northway, R.; Mead, D.; Moseley, L. The learning curve: The advantages and disadvantages in the use of focus groups as a method of data collection. *Nurse Res.* **2004**, *11*, 79–88.

81. Winkle, K.; Caleb-Solly, P.; Turton, A.; Bremner, P. Social robots for engagement in rehabilitative therapies: Design implications from a study with therapists. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 289–297.

82. Acocella, I. The focus groups in social research: Advantages and disadvantages. *Qual. Quant.* **2012**, *46*, 1125–1136.

83. Corti, L. *Using Diaries in Social Research*; Sage: 1993. Available online: https://uk.sagepub.com/en-gb/eur/using-diaries-for-social-research/book219185 (accessed on 30 September 2021).

84. Day, M.; Thatcher, J. "I'm really embarrassed that you're going to read this...": Reflections on using diaries in qualitative research. *Qual. Res. Psychol.* **2009**, *6*, 249–259.

85. Frennert, S.; Eftring, H.; Östlund, B. Case report: Implications of doing research on socially assistive robots in real homes. *Int. J. Soc. Robot.* **2017**, *9*, 401–415.

86. Snowden, M. Use of diaries in research. *Nurs. Stand. (2014+)* **2015**, *29*, 36.

87. Opdenakker, R. Advantages and disadvantages of four interview techniques in qualitative research. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*; 2006; Volume 7. Available online: https://www.qualitative-research.net/index.php/fqs (accessed on 30 September 2021).

88. Hannabuss, S. Research interviews. *New Library World* **1996**, *97*, 9.

89. Mann, C. Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emerg. Med. J.* **2003**, *20*, 54–60.

90. Lindahl, K.M. Methodological issues in family observational research. In *Family Observational Coding Systems*; Psychology Press: 2000; pp. 39–48. Available online: https://www.taylorfrancis.com/chapters/edit/10.4324/9781410605610-7/methodological-issues-family-observational-research-kristin-lindahl (accessed on 30 September 2021).

91. Foster, P. Observational research. *Data Collection and Analysis*; 1996; pp. 57–93. Available online: https://methods.sagepub.com/book/data-collection-and-analysis/n3.xml (accessed on 30 September 2021).

92. Lohani, M.; Payne, B.R.; Strayer, D.L. A review of psychophysiological measures to assess cognitive states in real-world driving. *Front. Hum. Neurosci.* **2019**, *13*, 57.

93. Rani, P.; Sarkar, N.; Smith, C.A.; Kirby, L.D. Anxiety detecting robotic system–towards implicit human–robot collaboration. *Robotica* **2004**, *22*, 85–95.

94. Schmidt, E.A.; Schrauf, M.; Simon, M.; Fritzsche, M.; Buchner, A.; Kincses, W.E. Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accid. Anal. Prev.* **2009**, *41*, 1087–1093.

95. Mauss, I.B.; Robinson, M.D. Measures of emotion: A review. *Cogn. Emot.* **2009**, *23*, 209–237.

96. Willemse, C.J.; van Erp, J.B. Social Touch in Human–Robot Interaction: Robot-Initiated Touches can Induce Positive Responses without Extensive Prior Bonding. *Int. J. Soc. Robot.* **2018**, *11*, 285–304.

97. Michaud, K.; Matheson, K.; Kelly, O.; Anisman, H. Impact of stressors in a natural context on release of cortisol in healthy adult humans: A meta-analysis. *Stress* **2008**, *11*, 177–197.

98. Book, A.S.; Starzyk, K.B.; Quinsey, V.L. The relationship between testosterone and aggression: A meta-analysis. *Aggress. Violent Behav.* **2001**, *6*, 579–599.

99. Mazur, A.; Booth, A. Testosterone and dominance in men. *Behav. Brain Sci.* **1998**, *21*, 353–363.

100. MacDonald, K.; MacDonald, T.M. The peptide that binds: A systematic review of oxytocin and its prosocial effects in humans. *Harv. Rev. Psychiatry* **2010**, *18*, 1–21.

101. Heinrichs, M.; Baumgartner, T.; Kirschbaum, C.; Ehlert, U. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biol. Psychiatry* **2003**, *54*, 1389–1398.

102. Schultheiss, O.C.; Stanton, S.J. Assessment of salivary hormones. *Methods Soc. Neurosci.* **2009**, *17*, 17–44.

103. Schaefer, K.E. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Robust Intelligence and Trust in Autonomous Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 191–218.

104. Salem, M.; Dautenhahn, K. Evaluating trust and safety in HRI: Practical issues and ethical challenges. In *Emerging Policy and Ethics of Human-Robot Interaction*; ACM Press: New York, NY, USA, 2015.

105. Yagoda, R.E.; Gillan, D.J. You want me to trust a ROBOT? The development of a human–robot interaction trust scale. *Int. J. Soc. Robot.* **2012**, *4*, 235–248.

106. Staudte, M.; Crocker, M.W. Visual attention in spoken human–robot interaction. In Proceedings of the 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), La Jolla, CA, USA, 9–13 March 2009; pp. 77–84.

107. Palinko, O.; Rea, F.; Sandini, G.; Sciutti, A. Robot reading human gaze: Why eye tracking is better than head tracking for human–robot collaboration. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 5048–5054.

108. Fraune, M.R.; Šabanović, S.; Kanda, T. Human group presence, group characteristics, and group norms affect human–robot interaction in naturalistic settings. *Front. Robot. AI* **2019**, *6*, 48.

109. Stemmler, G. *Methodological Considerations in the Psychophysiological Study of Emotion*; Oxford University Press: Oxford, UK, 2003.

110. Landowska, A.; Miler, J. Limitations of emotion recognition in software user experience evaluation context. In Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), Gdańsk, Poland, 11–14 September 2016; pp. 1631–1640.

111. Landowska, A.; Brodny, G.; Wrobel, M.R. Limitations of Emotion Recognition from Facial Expressions in e-Learning Context. In *CSEDU (2)*; 2017; pp. 383–389. Available online: https://www.scitepress.org/Papers/2017/63579/63579.pdf (accessed on 30 September 2021).

112. Kołakowska, A.; Landowska, A.; Szwoch, M.; Szwoch, W.; Wrobel, M.R. Emotion recognition and its applications. In *Human–Computer Systems Interaction: Backgrounds and Applications 3*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 51–62.

113. Tscherepanow, M.; Hillebrand, M.; Hegel, F.; Wrede, B.; Kummert, F. Direct imitation of human facial expressions by a user-interface robot. In Proceedings of the 2009 9th IEEE-RAS International Conference on Humanoid Robots, Paris, France, 7–10 December 2009; pp. 154–160.

114. Scheutz, M.; Schermerhorn, P.; Kramer, J. The utility of affect expression in natural language interactions in joint human–robot tasks. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human–Robot Interaction, Salt Lake City, UT, USA, 2–3 March 2006; pp. 226–233.

115. McColl, D.; Hong, A.; Hatakeyama, N.; Nejat, G.; Benhabib, B. A survey of autonomous human affect detection methods for social robots engaged in natural HRI. *J. Intell. Robot. Syst.* **2016**, *82*, 101–133.

116. Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Lee, S.; Neumann, U.; Narayanan, S. Analysis of emotion recognition using facial expressions, speech and multimodal information. In Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, USA, 13–15 October 2004; pp. 205–211.

117. Gaspar, A.; Esteves, F.; Arriaga, P. On prototypical facial expressions versus variation in facial behavior: What have we learned on the "visibility" of emotions from measuring facial actions in humans and apes. In *The Evolution of Social Communication in Primates*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 101–126.

118. Fairbairn, C.E. A nested frailty survival approach for analyzing small group behavioral observation data. *Small Group Res.* **2016**, *47*, 303–332.

119. Janis, R.A.; Burlingame, G.M.; Olsen, J.A. Evaluating factor structures of measures in group research: Looking between and within. *Group Dyn. Theory Res. Pract.* **2016**, *20*, 165.

120. Krull, J.L.; MacKinnon, D.P. Multilevel modeling of individual and group level mediated effects. *Multivar. Behav. Res.* **2001**, *36*, 249–277.

121. Kenny, D.A.; Mannetti, L.; Pierro, A.; Livi, S.; Kashy, D.A. The statistical analysis of data from small groups. *J. Personal. Soc. Psychol.* **2002**, *83*, 126.

122. Grawitch, M.J.; Munz, D.C. Are your data nonindependent? A practical guide to evaluating nonindependence and within-group agreement. *Underst. Stat.* **2004**, *3*, 231–257.

123. Pavitt, C. An interactive input–process–output model of social influence in decision-making groups. *Small Group Res.* **2014**, *45*, 704–730.

124. Mullen, B.; Goethals, G.R. *Theories of Group Behavior*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.

125. Poole, M.S.; Hollingshead, A.B.; McGrath, J.E.; Moreland, R.L.; Rohrbaugh, J. Interdisciplinary perspectives on small groups. *Small Group Res.* **2004**, *35*, 3–16.

126. Asch, S.E. Effects of group pressure upon the modification and distortion of judgments. In *Documents of Gestalt Psychology*; University of California Press: Berkeley, CA, USA, 1961; pp. 222–236.

127. Brandstetter, J.; Rácz, P.; Beckner, C.; Sandoval, E.B.; Hay, J.; Bartneck, C. A peer pressure experiment: Recreation of the Asch conformity experiment with robots. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 1335–1340.

128. Tajfel, H. *Social Identity and Intergroup Relations*; Cambridge University Press: Cambridge, UK, 2010.

129. Lovaglia, M.; Mannix, E.A.; Samuelson, C.D.; Sell, J.; Wilson, R.K. Conflict, Power, and Status in Groups. *Theories of Small Groups: Interdisciplinary Perspectives*. 2005. pp. 139–184. Available online: https://sk.sagepub.com/books/theories-of-small-groups (accessed on 20 December 2013).

130. Wittenbaum, G.M.; Hollingshead, A.B.; Paulus, P.B.; Hirokawa, R.Y.; Ancona, D.G.; Peterson, R.S.; Jehn, K.A.; Yoon, K. The functional perspective as a lens for understanding groups. *Small Group Res.* **2004**, *35*, 17–43.

131. Nosek, B.A.; Ebersole, C.R.; DeHaven, A.C.; Mellor, D.T. The preregistration revolution. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2600–2606.

132. Nosek, B.A.; Beck, E.D.; Campbell, L.; Flake, J.K.; Hardwicke, T.E.; Mellor, D.T.; van't Veer, A.E.; Vazire, S. Preregistration is hard, and worthwhile. *Trends Cogn. Sci.* **2019**, *23*, 815–818.

133. Simes, R.J. Publication bias: The case for an international registry of clinical trials. *J. Clin. Oncol.* **1986**, *4*, 1529–1541.

134. va not Veer, A.E.; Giner-Sorolla, R. Pre-registration in social psychology—A discussion and suggested template. *J. Exp. Soc. Psychol.* **2016**, *67*, 2–12.

135. Begley, C.G.; Ellis, L.M. Raise standards for preclinical cancer research. *Nature* **2012**, *483*, 531–533.

136. Prinz, F.; Schlange, T.; Asadullah, K. Believe it or not: How much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **2011**, *10*, 712–712.

137. Hartshorne, J.; Schachner, A. Tracking replicability as a method of post-publication open evaluation. *Front. Comput. Neurosci.* **2012**, *6*, 8.

138. Wager, T.D.; Lindquist, M.A.; Nichols, T.E.; Kober, H.; Van Snellenberg, J.X. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage* **2009**, *45*, S210–S221.

139. Collaboration, O.S. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspect. Psychol. Sci.* **2012**, *7*, 657–660.

140. Bartneck, C.; Suzuki, T.; Kanda, T.; Nomura, T. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *Ai Soc.* **2007**, *21*, 217–230.

141. Gnambs, T.; Appel, M. Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Comput. Hum. Behav.* **2019**, *93*, 53–61.

142. Fox, N.; Hunn, A.; Mathers, N. Sampling and sample size calculation. In *East Midlands/Yorkshire: The National Institutes for Health Research. Research Design Service for the East Midlands/Yorkshire & the Humber*. 2009. Available online: https://www.semanticscholar.org/paper/Sampling-and-Sample-Size-Calculation-Fox-Hunn/ae57ab527da5287ed215a9a3bf5f542ae19734ea (accessed on 30 September 2021).

143. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A.G. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **2009**, *41*, 1149–1160.

144. YENİPINAR, A.; Şeyma, K.; ÇANGA, D.; Fahrettin, K. Determining sample size in logistic regression with G-Power. *Black Sea J. Eng. Sci.* **2019**, *2*, 16–22.

145. Bujang, M.A.; Sa'at, N.; Bakar, T.M.I.T.A. Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malays. J. Med. Sci. MJMS* **2018**, *25*, 122.

146. Bujang, M.A.; Baharum, N. Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiol. Biostat. Public Health* **2017**, *14*, 2.

147. Vasileiou, K.; Barnett, J.; Thorpe, S.; Young, T. Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Med. Res. Methodol.* **2018**, *18*, 148.

148. Schoemann, A.M.; Boulton, A.J.; Short, S.D. Determining power and sample size for simple and complex mediation models. *Soc. Psychol. Personal. Sci.* **2017**, *8*, 379–386.

149. Hox, J.J.; Maas, C.J.; Brinkhuis, M.J. The effect of estimation method and sample size in multilevel structural equation modeling. *Stat. Neerl.* **2010**, *64*, 157–170.

150. Lane, S.P.; Hennes, E.P. Power struggles: Estimating sample size for multilevel relationships research. *J. Soc. Pers. Relatsh.* **2018**, *35*, 7–31.

151. Aylett, R. Games Robots Play: Once More, with Feeling. In *Emotion in Games*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 289–302.

152. Leite, I.; Martinho, C.; Pereira, A.; Paiva, A. As time goes by: Long-term evaluation of social presence in robotic companions. In Proceedings of the RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; pp. 669–674.

153. Hare, A.P. *Handbook of Small Group Research*. 1976. Available online: https://books.google.com.hk/books/about/Handbook_of_Small_Group_Research.html?id=LRZHAAAAMAAJ&redir_esc=y (accessed on 30 September 2021)

154. Hansen, H.; Trank, C.Q. This is going to hurt: Compassionate research methods. *Organ. Res. Methods* **2016**, *19*, 352–375.

155. Levine, J.M.; Moreland, R.L. Group socialization: Theory and research. *Eur. Rev. Soc. Psychol.* **1994**, *5*, 305–336.

156. Parker, S.K.; Wall, T.D.; Cordery, J.L. Future work design research and practice: Towards an elaborated model of work design. *J. Occup. Organ. Psychol.* **2001**, *74*, 413–440.