



Charting User Experience in Physical Human–Robot Interaction

HASTI SEIFI, School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA

ARPIT BHATIA and **KASPER HORNBAEK**, Department of Computer Science, University of Copenhagen, Denmark

Robots increasingly interact with humans through touch, where people are touching or being touched by robots. Yet, little is known about how such interactions shape a user's experience. To inform future work in this area, we conduct a systematic review of 44 studies on physical human–robot interaction (pHRI). Our review examines the parameters of the touch (e.g., the role of touch, location), the experimental variations used by researchers, and the methods used to assess user experience. We identify five facets of user experience metrics from the questionnaire items and data recordings for pHRI studies. We highlight gaps and methodological issues in studying pHRI and compare user evaluation trends with the Human–Computer Interaction (HCI) literature. Based on the review, we propose a conceptual model of the pHRI experience. The model highlights the components of such touch experiences to guide the design and evaluation of physical interactions with robots and inform future user experience questionnaire development.

CCS Concepts: • **Human-centered computing** → **Interaction design theory, concepts and paradigms**;

Additional Key Words and Phrases: Physical human–robot interaction, tactile human–robot interaction, haptics, user experience, systematic review

ACM Reference format:

Hasti Seifi, Arpit Bhatia, and Kasper Hornbæk. 2024. Charting User Experience in Physical Human–Robot Interaction. *ACM Trans. Hum.-Robot Interact.* 13, 2, Article 27 (June 2024), 29 pages.

<https://doi.org/10.1145/3659058>

1 Introduction

Robots are increasingly moving from caged areas in factories into public and private spaces. This transition creates situations where robots touch people on purpose or accidentally. For example, a robot may fetch an object and hand it over to a user, correct a person's pose during physical therapy, or touch someone to provide emotional support. Conversely, the ability of people to touch a robot also has important use cases. For instance, the user may teach the robot a physical maneuver, correct its movements, or touch a robot to convey emotions. In the **Human–Robot Interaction**

This research was supported by the National Science Foundation (Grant 2301335).

Authors' Contact Information: Hasti Seifi (Corresponding author), School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA; e-mail: hasti.seifi@asu.edu; Arpit Bhatia, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: arbh@di.ku.dk; Kasper Hornbæk, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: kash@di.ku.dk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-9522/2024/6-ART27

<https://doi.org/10.1145/3659058>

(HRI) literature, such touches are called tactile or **Physical HRI (pHRI)**. In this article, we refer to these interactions as pHRI.

Past research has supplied initial evidence that pHRI can impact **User Experience (UX)** and behavior [98, 110, 111]. For example, research has shown that users form an impression of the robot based on its timing and movement parameters [98], and robot touches can increase user motivation and effort in repetitive tasks [110]. Some studies have varied physical interaction parameters (e.g., touching force or location) [9, 98], whereas others have focused on the relation to other modalities (e.g., verbal announcement [32]) or the context of interaction [73]. Also, prior pHRI studies have used a wide range of evaluation metrics ranging from the ease of interaction [31, 42], to the touch sensation [124] or emotion [126] or user appraisal of the robot's social attributes in such interactions [19, 98].

Despite its importance, little is understood about the UX of pHRI and the factors influencing it. In particular, the above studies have not been pulled together systematically, and their integration into a conceptual model of pHRI has not happened. The lack of a conceptual model for the experience of robotic touch hinders progress on pHRI in several ways. First, researchers may find it difficult to clearly define an interaction's goals and measure the UX. Second, discussing the impact of interaction parameters on UX outcomes across studies is challenging. Third, the relation between the pHRI experience and existing work in fields such as UX of other forms of computing [15, 67, 106] is unclear. Thus, one cannot pinpoint how the growing area of physical interactions with robots, as embodied autonomous agents, may borrow from or complement the literature on the UX of other technologies.

We review work on pHRI to highlight gaps in the literature and provide a conceptual model of the UX of touching and being touched by robots. As a step toward this model, we ask (1) What physical interactions, experimental variables, and data collection methods are used in pHRI studies? and (2) What metrics do pHRI researchers use, and how do they relate to UX metrics reported for other technologies?

To answer these questions, we conducted a systematic review of a sample of pHRI studies that were published between 2010 and 2021. We identified 44 empirical studies of human–robot touch experience by screening the literature based on inclusion and exclusion criteria. Two authors coded various aspects of these studies, such as their goals, physical interactions, data collection methods, and measurements. Our analysis of the trends in these studies highlights underexplored areas in pHRI experience, such as a gap in studying accidental touch. For the second question, we collected all the questionnaire items and data recordings used in the studies and created an affinity diagram. This analysis led to 25 UX metrics that we further divided into 5 facets of pHRI experience: (1) *overall*, (2) *usability*, (3) *sensory*, (4) *personal and interpersonal*, and (5) *experiential* facets. We report the prevalent UX metrics as well as methodological issues in evaluating and reporting the interactions in our sample of 44 studies.

Based on the above analysis, we propose a conceptual model of the pHRI experience with three components (Figure 1): (1) *design parameters* of pHRI revolve around the three entities of the user, robot, and their interaction, (2) the pHRI *timeline* includes the physical interaction as a subcomponent that happens once or is repeated, and (3) the *UX metrics* capture the outcome of the interaction with the five facets of pHRI experience. The details of these components are derived from our systematic review of the 44 pHRI studies. We discuss how this conceptual model and the results of our review can inform pHRI research and practice. This article contributes:

- Trends and gaps in pHRI interactions and user evaluation practices from 44 studies published between 2010 and 2021.

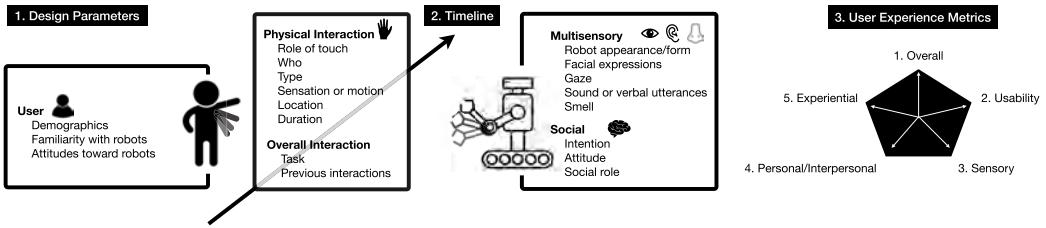


Fig. 1. A conceptual model of pHRI experience with three components: (1) design parameters can vary around the user, the robot, and their interaction, (2) the interaction timeline includes physical interactions, which are short and possibly repeated episodes during an overall HRI timeline, and (3) the UX outcomes of pHRI are shaped by the design parameters and interaction timeline. The UX can be measured according to five UX facets of *overall*, *usability*, *sensory*, *personal and interpersonal*, and *experiential* metrics. We devised this conceptual model based on our systematic review of the pHRI studies.

- Five facets of UX metrics and their prevalence in the pHRI studies.
- A first conceptual model of the pHRI experience based on our review.

2 Related Work

We present an overview of previous surveys in pHRI followed by research on defining the UX in the **Human–Computer Interaction (HCI)** and HRI literature.

2.1 Surveys on pHRI

Previous work refers to physical interactions with robots as tactile or pHRI. In the *Springer Handbook of Robotics*, Haddadin and Croft categorize pHRI as a form of proximate interaction where humans and robots are collocated, and the robot has autonomy in performing (part of) a task [43]. Others use tactile HRI to refer to physical interactions with robots. For example, Argall and Billard discuss that tactile HRI is at the intersection of two areas: (1) tactile sensing and (2) interactions between humans and robots [11]. No clear distinction exists between tactile HRI and pHRI in the literature. While “touch” has more social and experiential connotations than pHRI, both tactile HRI and pHRI have been used to refer to either technical or social aspects in the HRI literature. For example, when the purpose of interaction was purely pragmatic such as in object handover or kinesthetic teaching, the users could associate social and emotional attributes to the robot or interaction. Pan et al. conducted an instance of such a study where users evaluated the social attributes of the robot based on an object handover task [98]. We adopt the definition by Haddadin and Croft and treat pHRI as any collocated interaction that involves an exchange of haptic signals (e.g., force, contact, accelerations) between a human and an autonomous embodied agent. Specifically, in our definition, pHRI is the umbrella category that includes direct tactile interactions where humans are touching or being touched by a robot as well as indirect physical interactions through an object (e.g., during a handover). Also, in our definition, pHRI can refer to both pragmatic and experiential aspects of physical contact.

The literature on pHRI diverges into two areas, one investigating technical engineering challenges of pHRI in the robotics literature, while the second area focuses on the social experiential aspects of physical interactions with robots. Existing pHRI surveys primarily cover the former literature on the technical developments in the field. In their atlas of pHRI, De Santis et al. proposed the safety and dependability of robots as two key criteria for assessing physical interactions [35]. They provided an overview of work on robot hardware and software design toward these two criteria. Relatedly, Hadadin and Croft provided an overview of the technical pHRI literature focusing on human injury

analysis and safety standards for pHRI. They presented progress toward human-friendly hardware and algorithm design for robots [43]. Losey et al. reviewed techniques and algorithms related to sharing control between humans and robots in collaborative physical tasks [80]. Argall and Billard categorized the literature according to progress in tactile sensor development and types of physical interactions [11]. The categorization of tactile sensors was purely technical, focusing on the composition of the sensor and the applications of various approaches (e.g., hard sensors vs. soft sensors). They categorized the types of physical interactions into three groups where the touch (1) interferes with robot behavior execution such as in accidental touch, (2) contributes to behavior execution such as in collaborative assembly or robot-assisted therapy, or (3) contributes to behavior development, for instance, when the robot is learning a skill through physical contact.

Others have focused on one type of physical interaction. Most recently, Ortenzi et al. published a review of object handover studies [95]. They described two phases of a handover task as pre-handover and physical handover, noting that important cognitive and physical processes start before the physical part of the interaction. The authors reviewed the progress and gaps in the literature according to these two stages of the object handover task. In addition, they summarized the user evaluation metrics into two categories of objective task performance and UX metrics, and further divided the UX metrics into subjective and psycho-physiological metrics. The above surveys informed our initial codes for analyzing the studies.

Our work complements the above literature by providing a systematic review of UX in pHRI. In contrast to the above surveys, our review focuses on the UX of pHRI. Also, while informative, the above surveys do not explicitly report the sample of articles and the analytical process of the authors. In contrast, we present a systematic review of the literature where we define a sample based on clear inclusion and exclusion criteria and code the articles according to a code book to provide statistics on the trends observed in the sample. With this approach, we also present a data-driven categorization of subjective and objective UX metrics for pHRI across various physical tasks.

2.2 Definition and Metrics for UX

Previous work in the HCI domain (e.g., games and mobile phones) has outlined the definition and methods for evaluating the UX of interactive technology. UX research aims to provide a holistic view of human interactions and reactions to technology. Hassenzahl and Tractinsky divided the UX of an interactive product into pragmatic and hedonic factors [49, 50]. Pragmatic or usability factors provide a task-centered view, focusing on the effectiveness and efficiency with which target users can complete specific tasks with a given technology [55, 62]. In contrast, UX is a broader multidimensional concept that also encompasses the positive aspects of interactions such as user motivation, emotions, and esthetics, and reflects the context-dependent and dynamic nature of interacting with technology [15, 50]. Bargas-Avila and Hornbæk provided a systematic review of 51 empirical studies of UX, analyzing the range of *products* (e.g., websites, and games), *UX dimensions*, and data collection *methodologies* employed in their sample. The studies in their review showed a prominent focus on using qualitative methods (e.g., interviews) and evaluating emotions, enjoyment, and aesthetics as measures of UX. We follow a similar systematic review approach and compare evaluation trends in pHRI with those reported by Bargas-Avila and Hornbæk for other forms of interactive technology.

In the HRI domain, researchers have emphasized the need for adopting the UX design practices from HCI [7, 78] and proposed new frameworks to capture the unique nature of interacting with robots [129]. Lindblom et al. described three primary challenges for HRI as the need for robot designers to adopt an iterative process, incorporate UX goals in the development process, and learn about UX evaluation methods and theory [78]. Young et al. discussed adopting three perspectives when designing and evaluating HRI. These perspectives account for (1) visceral factors such as

emotions, (2) social mechanics such as gestures and facial expressions, and (3) social structures involving the interaction context [129]. Similarly, Weis et al. provided a framework and guidelines for the holistic evaluation of HRI with four key factors: usability, UX, social acceptance, and societal impact [119]. Grounded in the researchers' experience and knowledge of the literature, these frameworks highlight the complexity of capturing the UX of interacting with robots. Yet, the extent to which these frameworks or their underlying factors are employed in pHRI studies is unknown.

A few questionnaires exist for evaluating the UX of robots. HRI researchers commonly employ general-purpose questionnaires to measure user emotions [22, 87, 117] or workload [48]. In addition, the Godspeed questionnaire [16] and **Robotic Social Attributes Scale (RoSAS)** [29] are specifically designed and widely used for robots. The Godspeed questionnaire consists of 24 Likert-scale user ratings about the robot's anthropomorphism, animacy, likeability, intelligence, and safety. While the questionnaire is not validated, its wide adoption reflects the field's demand for established evaluation practices. In a series of online studies, Carpinella et al. analyzed the Godspeed questionnaire and developed the RoSAS with 18 items that capture a robot's perceived warmth, competence, and comfort. RoSAS has been validated in online studies using descriptions of robots or images of robot faces, but the questionnaire is also used in pHRI studies recently [98]. Other instruments such as the **Negative Attitude toward Robots Scale (NARS)** [91, 93] and **Robot Social Anxiety Scale (RAS)** [92] capture overall user beliefs and feelings toward robots. Building on this literature, we provide a detailed account of how empirical studies of pHRI measure UX and discuss how our results can inform future evaluation practices and questionnaire development in the field.

3 Methods

We describe our process for identifying and screening relevant studies for our review, iterative coding of the studies included in our final sample, and creating an affinity diagram of the UX metrics. Our review procedure is based on the **Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA)** Flow Diagram [97] for systematic reviews (see Figure 2).

3.1 Identification

To gather high-impact articles on pHRI user studies, we searched through three top HRI venues: the **ACM/IEEE International Conference on Human Robot Interaction (HRI)**, ACM; the **ACM Transactions on Human–Robot Interaction (THRI)**, ACM; and the **International Journal of Social Robotics (IJSR)**, Springer. We used the following query to search in the title and abstract of every publication in the selected venues:

```
(touch* OR tactile* OR hand* OR haptic* OR kines*) AND (study OR evaluat* OR participant* OR experiment*)
```

Since we focused on UX in pHRI, our query filtered for articles with both a physical interaction and a user study of the interaction. The first part of the query aimed to capture articles with touch interaction, while the second part included articles with a user study. The wild cards (*) helped us capture different possible wording such as "handover" or "handshake" and "evaluate" or "evaluation." For HRI and THRI, we could directly run our query using the ACM Digital Library search options. In the case of IJSR, Springer only supports searching on the full text of the article. Hence, we used a custom script to further filter the results obtained from the Springer website to ensure that the search terms only appear in the title or abstract. We included all full articles published before October 2021, when we ran the search query. The query returned 215 articles: 92 from HRI, 32 from THRI, and 91 from IJSR.

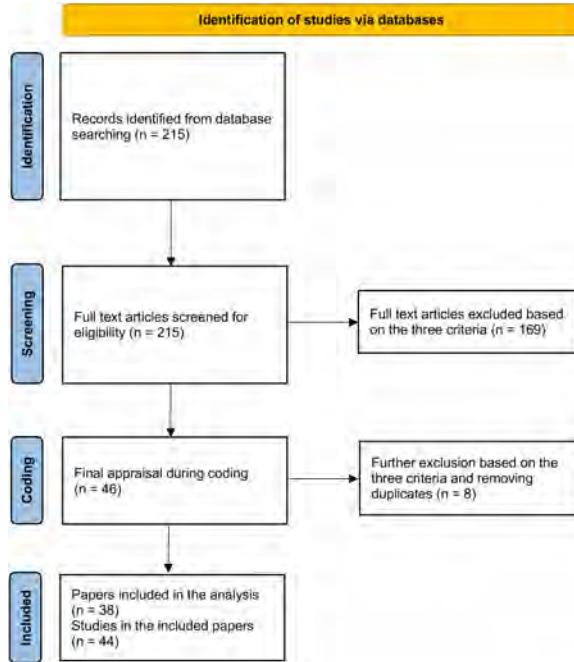


Fig. 2. The four stages we used to identify relevant studies for our review. These stages are based on the PRISMA Flow Diagram [97] and include the number of articles involved in each stage.

3.2 Screening

To further screen relevant articles from the search query results, we defined the following inclusion criteria:

(1) *Physical Autonomous Robots*: The article needs to involve a physical robot that appears as an autonomous agent to the user. This criterion ensures that the user thinks they are interacting with a robot rather than a human. We include studies using a Wizard of Oz setup [104] where a human operates the robot without the user's knowledge. Articles involving virtual avatars or simulations [86], haptic devices operated by a human [107], and teleoperated robots [113] are excluded according to this criterion.

(2) *Touch Interaction*: The article needs to have a direct touch interaction or indirect interaction through an object (e.g., handover) where physical forces or tactile feedback is exchanged between a human and a robot. The interaction should not be via an intermediate interface such as through a joystick [115], a touchscreen [84], or haptic devices [107, 131].

(3) *User Study*: Physical interaction needs to be evaluated through a quantitative or qualitative user study by having a research question or a hypothesis about the touch interaction. The study should focus on user response and evaluation rather than hardware design [112] or algorithm development [33]. The participants in the study must either participate in the touch action or observe it.

Two authors independently determined and marked the inclusion of all the articles (0 for excluded, 1 for included) by skimming through the full-text articles. The overall agreement percentage was 90.69%, and Cohen's kappa was 0.746. Each article with a rating discrepancy was discussed among all three authors, with the third author providing an independent opinion. In this phase, 169 articles were excluded, resulting in a sample of 46 articles for further coding. During the coding process,

we further removed eight articles that did not meet our inclusion criteria or were duplicates due to a conference article being extended and later published in a journal. We included the journal version in our analysis.

The final set of 38 articles consisted of 44 unique studies, which we included in the review.

3.3 Coding the Included Articles

We coded the final set of 44 studies in two rounds. The first round provided *summary codes*, and the second round resulted in *detailed codes* for the user studies and physical interactions. In the first round, two authors extracted free-form text from the articles about the study goal(s), physical interactions and their context, user evaluation methods, and high-level outcomes of the article. We copied the article's text and paraphrased it for brevity whenever necessary. These text descriptions (i.e., *summary codes*) served as a memory aid for the coders and guided the next coding round.

In the second round, the same two authors defined *detailed codes* for the physical interactions and user evaluation methods. For the physical interactions, we defined five detailed codes describing the role, type, location, actor(s), and duration of the physical interaction (Table 1). For the user evaluation methods, we coded the time of data collection, methods of data collection, and the questionnaires used in the studies (Table 3) similar to the review by Bargas-Avila and Hornbæk [15]. Also, we further extracted all the subjective user-reported data (e.g., questionnaire items, interview questions) and measurements (e.g., task completion time and user behavior coded from videos). Finally, we extracted parameters that were varied in the studies (i.e., independent variables) as well as the robot model, number of participants, and the participant background data that were collected in the studies (Table 2). We established an interrater agreement for the *detailed codes* on a random 25% subset of the studies (i.e., 11 studies) before coding the rest of the studies. The interrater reliability was full agreement on 78.03% of the codes, partial agreement on 10.61% of the codes, and a mismatch on 11.36% of the codes. The two authors discussed the differences and updated the descriptions in a coding sheet. The rest of the studies ($n = 33$) were coded by one author. The aggregated results of this coding process are reported as numbers, percentages, and instances in the 44 studies (Tables 1–3). The final coding scheme and the codes for the 44 studies are included as supplementary materials.

3.4 Identifying User Experience Metrics through Affinity Diagramming

We further analyzed all the measurements and questionnaire statements from the previous coding phase to identify UX metrics in the pHRI studies (see Supplemental Materials). These measurements and questionnaire statements capture the aspects of UX that the pHRI researchers aimed to cover in their study. Thus, they provided a rich data source for analysis. In two sessions, the three authors created affinity diagrams of the measurements and statements.

We grouped the questionnaire statements that users rated in the studies in one session. We discarded all the ratings where the exact statement or question was not reported in the article (53 rated statements and questions from 14 studies). This left 333 statements and questions for the affinity diagramming session. Before the session, all the statements were printed on article, cut individually, and mixed into a random order. During the affinity diagramming session, the three authors read each statement aloud, discussed the underlying metric, and placed it on a table near statements referring to the same metric. When a statement was unclear, it was set aside for checking in the future. After going through 100 statements, the authors labeled the affinity groups using sticky notes. This process was repeated until all the statements were placed in the groups. If an author proposed a change in the grouping of a statement, all the authors discussed and accepted or

Table 1. Parameters of the Physical Interactions in the 44 pHRI Studies

Physical Interaction Parameters	Description and Examples	N	%
1. Role of touch			
Support completing a task	Physical contact is needed for completing a task [41, 100].	8	18
Communicate or influence	Physical contact is mainly provided to communicate to or influence users or robots (e.g., emotions, effort, and judgment) or socially support them [12, 54].	26	59
Teach or guide movement	Physical contact is used to teach or guide movement of the robot or user [74, 79].	10	23
Unintended contact	Physical contact is not intended in the interaction, but it happens (or appear to happen) as a result of an error [64].	1	2
2. Who			
Human	The human initiates and is active in the physical contact [42, 76].	19	43
Robot	The robot initiates and is active in the physical contact [32, 73].	10	23
Mutual	Both the human and the robot participate in the physical interaction (e.g., handover, handshake, and hug) [18, 109].	20	45
3. Type of touch			
Touch (general)	The action is generally reported as “touching” in the article. This category often involves brief or static contact [12, 124].	13	30
Move	Holding onto and moving a body part in space [5, 38].	6	14
Handover	Passing objects without direct physical contact between the actors [8, 98].	7	16
Handshake	Taking hold of and shaking each other’s hand [9, 14].	4	9
Hug	Embracing or being embraced in one’s arms [18, 111].	5	11
Stroke	Actions that were described as stroking or wiping in the studies [32, 110].	2	5
Other	Push, pull, tap, hand clapping, or when users could select any touch actions from a set of available actions [31, 42].	11	25
4. Body location			
Hand or end effector	Anywhere on or below the wrist for the user as well as the robot’s end effector [39, 79].	19	43
Arm	Forearm or upper arm of the user, a humanoid robot, or any location on a robotic arm [5, 124].	7	16
Whole body	The physical contact involved multiple body parts (e.g., during a hug) or the contact could be applied to any body part (e.g., touching a robot anywhere on its body) [19, 109].	10	23
Other	Other body locations included shoulder, upper back, waist, buttock, and the robot’s tray [69, 73].	11	25
5. Duration			
Brief	≤60 seconds [8, 88]	13	30
Long	>60 seconds [109, 123]	2	5
Unlimited	No time limit was imposed on the physical interactions and the timing varied across the users [52, 59].	8	18
Not reported	Duration of the contact is unclear from the article [76, 79].	21	48

We coded the studies according to the role of touch, who initiated the touch (actors), type of touch, location of touch, and duration of the physical interaction. The numbers and percentages may not add up to 44 and 100, respectively, because one study can involve multiple parameters.

rejected the proposed change. In this process, we discarded 17 statements that were either open-ended questions (e.g., “What would you name the robot?”) or the exact measures or their meaning was unclear (e.g., “the participant’s impression of the researcher who proctored the experiment”).

Table 2. Factors That Were Systematically Varied (i.e., Independent Variables) in the pHRI Studies

Independent Variable	Description and Examples	N	%
Physical interaction	Touch/no touch ($n = 10$), role of touch ($n = 2$), who ($n = 4$), location ($n = 1$), sensation or motion parameters ($n = 13$), duration or timing of touch ($n = 5$) [69, 110].	35	80
Visual	Facial expressions ($n = 2$), gaze behavior ($n = 1$), or visual appearance of the robot ($n = 1$) [18, 39].	4	9
Sound/utterances	Verbal utterances or noises made by the robot as part of the physical interaction [32].	1	2
Task	The study task to be completed [31, 42].	6	14
Intention	Robot or human's attitude or social role [63, 126].	9	20
Demographics	Gender or other demographic characteristics of the robot or participants [12, 109].	3	7
Other	Previous interactions, task outcome (success or failure), human vs. robot [52, 10].	6	14
None	This category includes qualitative studies without independent variables and studies where the researchers collected numerical data but no parameters were varied systematically [14, 100].	5	11

The numbers and percentages may not add up to 44 and 100, respectively, because one study can involve multiple independent variables.

This process resulted in 24 groups (i.e., UX metrics). After the session, one of the authors entered the UX metric for each of the 316 statements (333 statements—17 discarded items) into an Excel sheet. Another author reviewed all the statements and their UX metrics and flagged 31 items that were not consistently categorized. In a subsequent meeting, the three authors discussed these statements and revised the UX metrics for the statements as needed.

In a second session, the authors grouped 72 data measurements (e.g., through datalogs or video coding) from the 44 studies. The three authors followed the same procedure as above to group the measurements. This process led to five groups (i.e., UX metrics). Four of the UX metrics, including *accuracy*, *time*, *social traits or behavior*, and *emotion*, overlapped with the 24 metrics from the rated statements. One new metric was identified as *descriptive measures* (e.g., number of actions).

Finally, one of the authors counted the number of studies with a rated or measured item for each UX metric. Table 4 presents the results of the above process.

4 Results

In this section, we present trends and gaps in the pHRI studies based on the detailed analysis of our sample. Specifically, we report the physical interaction parameters, independent variables used in the studies, data collection methods, and user study metrics in our sample of 44 studies.

4.1 Physical Interaction

We discuss the physical interactions according to five parameters that were varied in previous studies: role of touch, who initiates and is active in touching (actor), type of touch, body location, and duration of the physical interaction. Figure 3(f) shows six example physical interactions with different combinations of these parameters. Table 1 gives an overview of physical interaction parameters for all the studies in our sample.

Table 3. Timing and Methods of Data Collection and Questionnaires Used in the Reviewed Studies

Categorization Scheme	Description and Examples	N	%
1. Time of data collection			
Before	Electrodermal arousal before touch [76]; NARS before any interaction with the robot [130]	11	25
During	Video recordings of each touch event [126]; time to complete task [31]	25	57
After	RoSAS after interaction with the robot [98]; post-study interview [79]	37	84
2. Method of data collection			
Questionnaires	Self-developed questionnaire on robot friendliness [110], social perception of human-to-robot handovers with RoSAS [98]	38	86
Video recordings	Video analysis of emotions exhibited by participants during clap interaction [39], timing and frequency information from videos of handovers [45]	15	34
Datalog	Motion tracking to collect average of speed, step-length, and cadence [74], temperature and tactile sensor data [14]	7	16
Interviews	Interviews regarding attitudes toward assistance from a robot in homes [100], interviews regarding the whole experience of touch interactions with a humanoid robot [130]	4	9
Physiological signals	Skin conductance response [130], respiration rate [121]	3	7
Other	Expert rating of task outcomes [5], manual timing [42], and think aloud [38, 45]	6	14
3. Questionnaire type			
Existing questionnaires (validated)	NASA TLX [48], SAM [22], PANAS [117], PAD [87], and RoSAS [29]	16	36
Existing questionnaires (not validated)	GodSpeed [16], RAS [92], and NARS [93]	12	27
Self-developed (items known)	Likert scale rating on robot's social qualities [12], Likert scale rating on user's convenience while receiving an object during a handover [8]	16	36
Self-developed (items unknown)	Open-ended questions regarding kinesthetic teaching methods [5], questionnaire regarding subjective experience of interacting with the robot via touch [10]	4	9

The numbers and percentages may not add up to 44 and 100, respectively, because one study can involve multiple methods. SAM, Self-Assessment Manikin; PANAS, Positive and Negative Affect Schedule; and PAD, Pleasure-Arousal-Dominance.

Role of Touch. pHRI has been used in multiple domains such as healthcare, manufacturing, education, technical assistance, and domestic help. Within these domains, we identified four primary roles for physical interactions. First, in most studies ($n = 26$), the main purpose of the physical interaction was to *communicate or influence* the user's or the robot's emotions, judgment, or behavior. For example, Law et al. studied if participants' trust in a robot is impacted after observing the robot touch a person on the shoulder in a video [73]. Shiomi et al. investigated the effect of

Table 4. The 25 Metrics and the 5 Facets of the pHRI Experience

UX Metric	Definition	Example Measurements and Rated Statement	N (Rated)	N (Measured)
F1—Overall Overall evaluation	Assessment of overall experience as positive or negative, including statements about user preference and liking.	“I think using the robot is a good idea.” [19], “I would have preferred that the robot did not touch my arm” [32].	13 (21)	-
Descriptive measures	Summary statistics describing the task/interaction without a positive or negative connotation	Number of actions [110], gesture intensity [126]	-	11 (15)
F2—Usability				
Time	Time needed to complete a task	Completion time [42], response time [76], “I am satisfied with the time it took to complete the task using the interface.” [31], “Efficient” [42]	2 (2)	8 (13)
Accuracy	The accuracy with which a task is completed, that is some quantification of error	Number of collisions [31], if the robot accidentally dropped the object [45], “accurate” [88]	1 (1)	8 (10)
Ease of use	General satisfaction with using the interface	“I think the robot is easy to use.” [18], “I was worried that I might break the robot using the interface” [31].	8 (22)	-
Understanding the task*	Understanding or learning of information in the interface	“I found the voice of the robot easy to understand.” [41], “The interface was intuitive to use to complete the task.” [31]	6 (12)	-
Workload	The physical (e.g., energy) and/or mental resources users spend on the interaction, including NASA Task Load Index (TLX) as an established instrument	“I really had to concentrate to use the robot.” [41], “Was the handling physically exhausting?” [123]	3 (4)	-
Feedback	The amount and quality of information given to the user during the interaction	“The instructions from the robot were sufficient.” [41], “Do you think the feedback was helpful?” [123]	3 (4)	-
Learnability	User attitude toward how easy it is to learn to use the interface	“It was easy to learn how to use the touching interface.” [31], “How difficult was to learn how to use the robot?” [74]	2 (3)	-
F3—Sensory				
Visual	Qualities judged based on appearance	“Large/small” [124], “laid-back/busy” [124]	7 (15)	-
Physical sensation	Qualities judged through touch	“Smooth/rough” [124], “The robot looks very strong” [32]	3 (16)	-
Auditory	Qualities judged through sound	“Quiet/noisy” [124]	1 (1)	-
F4—Personal and Interpersonal				
Social traits or behavior	Traits or behavior that relate to interactions with others	Percentage of eye-contact [14], face distance [14], and frequency of prompted/unprompted touches [34], “This hug made the robot seem (unfriendly–friendly).” Block et al. [18], “Likeable” [42]	14 (32)	7 (21)

(Continued)

Table 5. Continued

UX Metric	Definition	Example Measurements and Rated Statement	N (Rated)	N (Measured)
Personal traits	Qualities that typically belong to a person	“I think the robot went out of its way to help the person.” [12], “Principled” [73]	9 (37)	-
Capability	Assessment of the skills or ability of an entity which may or may not refer to a specific task	“I felt that the robot was very capable of performing its job.” [12], “I trust the robot to do the right thing at the right time.” [45]	8 (19)	-
Active or passive	Assessment of the overall activity of an entity. This includes statements about the speed or frequency of action or reaction.	“The robot moves its arms too slowly.” [63], “The robot showed an passive behavior” [41].	8 (17)	-
Intelligence	A subset of capability that focuses on mental skills or ability	“I feel understood by the robot.” [19], “The robot understood what I explained to it.” [63]	8 (11)	-
Predictability	Qualities of reliability, consistency, and anticipating the next action of an entity	“I always knew what the robot was going to do next.” [42], “The robot worked the way I expected it to.” [41]	7 (11)	-
Teamwork	A subset of capability that focuses on joint abilities or skills between two or more entities	“The robot has specialized capabilities that can increase our performance.” [12], “Someday I could work with this robot to build something of interest.” [63]	5 (14)	-
<hr/>				
F5—Experiential				
Safety	Feelings of fear, being threatened or nervous, and danger	“I felt safe.” [8], “I feel threatened by the robot.” [19]	16 (26)	-
Enjoyment	Comfort, enjoyment, or engagement	“It was enjoyable when the robot was touching my arm.” [32], “I feel uncomfortable with the robot.” [45]	15 (25)	-
Emotion	Affect instruments (PAD [87], SAM [22], and PANAS [117]), emotion labels in Russell’s circumplex model of affect [105], or reference to user feelings	Arousal level from Galvanic Skin Response [76], facial expressions [39], “I found it exciting to interact with the robot.” [41], “Interesting/Boring” [14]	8 (18)	5 (8)
Symbolic	Referring to the value or meaning of something in the society (among people) [6]	“People would be impressed if I had such a robot.” [19], “I would feel nervous operating a robot in front of other people.” [130]	2 (2)	-
Motivation	Internal desire or external pressure to do something	“I was motivated to walk.” [74], “I felt pressure or resistance for walking faster/slower.” [74]	1 (2)	-
Autonomy	Sense of control or independence in the interaction	“I felt independent to walk, even being supported by the platform.” [74]	1 (1)	-

The metrics are derived from the rated statements and measured items reported in the 44 studies. The numbers outside the parentheses reflect the number of studies and the numbers. The numbers inside the parentheses reflect the number of items. The metrics in each facet are sorted from high to low prevalence in the studies.

touch on user behavior. In a study, they examined how touching or being touched by a robot influences user effort in a monotonous task and their judgment of the robot’s friendliness [110]. In another study, the same authors examined if reciprocating the user’s hug by a teddy bear robot can increase the user’s interaction time and amount of self-disclosures [111]. Fitter and Kuchenbecker studied user perception of the social attributes of a Baxter robot during a playful hand-clapping

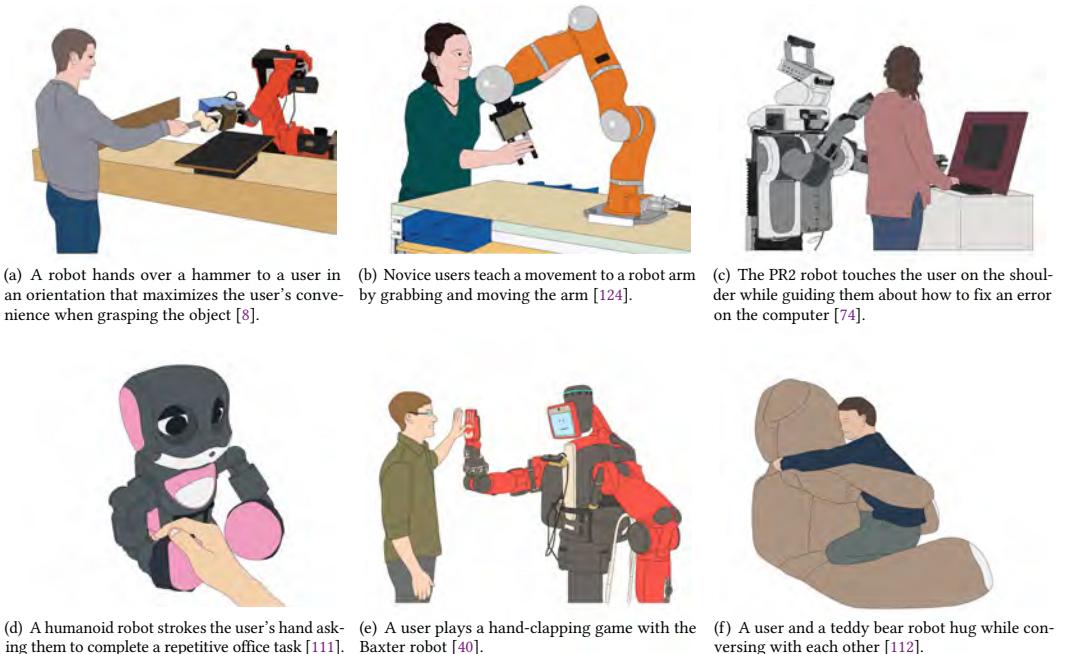


Fig. 3. Example of physical interactions in our review that show a variety of touch parameters (Table 1).

game [39]. Second, in several studies ($n = 8$), physical contact was required to *support completing a task*, such as object handover [8], nursing [32], or opening a path for robot navigation [64]. Third, in a subset of the studies ($n = 10$), the physical contact served to *guide or teach movement* of the robot or the user. In particular, five studies had the user involved in guiding a robot to a position through direct touch [31, 38, 42, 63], four studies examined UX of programming movements of a robot by physical demonstration [5, 52, 79, 123], and one study focused on a robotic platform supporting user movement in a walking rehabilitation task [74]. Fourth, only in one study ($n = 1$) the physical contact was *unintended*, meaning that the contact appeared to be the result of an error [64]. Given that accidental contact with robots can happen as a result of errors or noise in the robot hardware or algorithms, the lack of research on the effects of accidental contact on UX is surprising. We discuss this gap in Section 5.

Who Touches. The studies also differed in whether the human ($n = 19$), the robot ($n = 10$), or both ($n = 20$) were actively engaged in the touch interactions. In this case, analyzing the link between who touches and the role of touch in the reviewed studies revealed interesting patterns. When the human was touching, the role of touch could be any of the four categories mentioned above. Joint touches were used to communicate or influence users ($n = 12$), support completing a task ($n = 7$), or teach or guide movement ($n = 1$). In contrast, when the robot was touching, the role of the touch was mainly to communicate or influence the participant ($n = 9$). For example, Yeufang et al. investigated how a robot's touch impacts the user's emotion and attitude toward the robot [130]. Only in two study conditions did the robot-initiated touch support completing a task. In one study, the robot cleaned the forearm of the user in a nursing context [32], and in another study, the robot touched the participant to open a navigation path [64]. Robot-initiated touches were not used to teach or guide the user's movement, and a robot did not initiate an unintended accidental contact. These scenarios are imaginable in physical interaction with robots, but they are underexplored in the literature.

Type of Touch and Body Location. The type of touch interaction was described with various phrases and levels of detail. Several studies ($n = 13$) referred to the interaction generally as touching without specifying it in detail. Others used a label (e.g., tapping) to describe the touch. However, we could not discern whether the use of these labels was consistent across different studies. The location of touch was mainly on the hand ($n = 19$) or arm of the users or the robots ($n = 7$). Others involved multiple areas of the body ($n = 10$) such as during a hug, or focused on other body parts ($n = 11$) such as the shoulder or waist. The focus on hands and arms is similar to the haptics literature, reflecting the importance of hands in sensing and manipulation. In contrast, haptic devices targeting multiple or other body parts are rare.

Duration. The duration and timing of the touch showed trends and some methodological issues in the reviewed articles. The duration of physical contact was reported in only about half of the studies ($n = 23$). When reported, the duration ranged from a few seconds ($n = 13$) to minutes ($n = 2$). In some cases, the duration was not limited in the study and differed between participants ($n = 8$). In all cases, the touch event was part of a longer interaction timeline, usually spanning a single session. Only one study was conducted over several sessions. Huijnen et al. ran a four-session study with children with autism over 4 weeks to compare their attention and interaction with the KASPAR robot and a teacher [59]. The time of the robot-initiated touches ($n = 10$) was either implicitly or explicitly communicated to the participants. In two studies, the participants could not anticipate the exact timing of the touch. In one study, the robot occasionally touched the participant during a scary movie [121], and in another study, the robot touched the participant's shoulder to open a navigation path while passing from behind the participant [64]. In other cases, the participants could either anticipate the contact from the robot's verbal announcement or the study protocol ($n = 3$), or they only observed videos of a robot touching someone ($n = 5$). Studies of the robot touching the user without explicit notice or permission can improve the efficiency of human–robot teaming, but they are underexplored in the reviewed studies.

4.2 User Study Variables and Methods

Most reviewed studies ($n = 39$) used controlled experiments to study the effect of one or more independent variables on UX or task performance. We examined the independent variables (Table 2) and the data collection methods (Table 3) that the authors employed in their studies.

Independent Variables. Most studies varied the parameters of the physical interaction. Some compared direct physical contact with the robot to no contact or to interacting through an interface such as a joystick ($n = 10$) while others varied who touched the other ($n = 4$). Variations in sensation or motion parameters included the type of touch and robot forces [9, 18, 19, 39, 79, 123, 125], body materials [124], temperature [19], the existence of clothes [10], as well as robot's position or trajectory of motion [8, 25, 45, 69, 88, 98] for the physical interaction. Other studies examined timing variables such as the touch duration [18, 19, 45] or robot's movement speed [98] or reactivity [39]. Two studies varied the role of the touch interaction [32, 64]. Specifically, in one study, the robot stroked the user's forearm to either clean it (i.e., completing an action) or to support the user emotionally [32]. In another study, the robot directly touched the user to open a path for its navigation (*support completing a task*), or the user accidentally collided with the moving robot (*unintended contact*) in two different conditions [64]. The location of touch was only varied in one study. The focus on studying the impact of physical parameters is not surprising for pHRI studies.

Other studies varied factors such as visual and auditory modalities, the interaction tasks, and the user or robot's background or intent. Only one study varied the robot form by asking users to rate videos of two hugging robots [18]. All the other studies used a single robot. Similarly, verbal utterances or sounds were also only varied in one study. In this case, Chen et al. manipulated the timing of verbal utterance in relation to touch [32]. The results of these studies confirmed

that various contextual parameters can mediate the UX of touch. Finally, a few studies did not have any independent variable; they examined the correlation between different variables and measurements (e.g., number of verbal utterances and data from temperature sensors) without systematically varying them [14] or collected user interaction data and experience for only one physical interaction condition [34, 100].

Study Methods. We analyzed the user study methods based on the time and methods of data collection and the questionnaire types and sources following Bargas-Avila and Hornbæk’s analysis of the HCI literature [15].

The results of coding the time of data collection show clear similarities between the pHRI and the HCI literature (Table 3). Researchers often collected the data after ($n = 37, 84\%$) or during ($n = 25, 57\%$) the physical interaction. A small set of studies also collected data before the physical interaction ($n = 11, 25\%$). The before data provided baselines for physiological recordings ($n = 3$) or user impressions of a robot before physical contact ($n = 5$). The video recordings helped the researchers code the start time of the physical interactions in two studies. Similar to these trends, the review of UX methods in the HCI literature reported that researchers often collected data after the user interaction (70%), followed by during (58%) and before the interaction (20%).

In contrast, the data collection methods in our review (Table 3) are less diverse and more focused on quantitative data than those reported in the above-mentioned review. Most pHRI studies in our review used a questionnaire with subjective ratings ($n = 38, 86\%$), followed by video recordings ($n = 15, 34\%$) and datalogs ($n = 7, 16\%$). Interviews and other data collection methods, such as measurements of body movements and think-aloud protocol, were rarely used. In contrast, Bargas-Avila and Hornbæk noted that only 33% of their reviewed studies mainly used quantitative methods. Also, they reported a wider range of data collection methods in their sample, such as live observation, diaries, probes, collages or drawings, and photographs.

Many studies ($n = 20$) use self-developed questionnaires. In a subset of the studies ($n = 4$), the authors do not report the exact statement or questions presented to the participants. Validated questionnaires are often used for evaluating one’s emotions or workload. Similar methodological trends and issues are previously reported in other fields [15] and are known to the community. Such patterns may be inevitable in a developing field and suggest the need to develop questionnaires and best practices for pHRI studies.

4.3 UX Metrics

To capture the metrics of the pHRI experience, we collated the rated statements and measurements in the reviewed articles, created an affinity diagram, and grouped the resulting metrics into five facets. Table 4 shows the 25 UX metrics from this process. One can note the variety and distribution of the UX metrics in the reviewed studies. No metric was used by more than half of the studies ($n > 22$), and some metrics such as *autonomy* and *motivation* were included in only one study. Among the UX metrics, *accuracy*, *time*, *social traits or behavior*, and *emotions* are captured with both rated and measured items. In contrast, the rest of the metrics are only collected via rated statements, and *descriptive measures* is only among measured items. We further divided the 25 UX metrics into the following five facets:

F1—Overall. These UX metrics provide an overall interaction perspective without focusing on any specific component. The metrics include descriptive measures and overall evaluation ratings.

F2—Usability. This facet provides a task-centered perspective and focuses on the users’ performance and opinion in relation to completing a task. These metrics are known for evaluating the usability of an interface in the HCI literature [55]. Here, the robot is regarded as a computer interface. *Ease of use*, *accuracy* of the outcome, user *workload*, or extent of *understanding the task* are covered in several studies, whereas *learnability* and *feedback* are less common in our sample.

Hornbæk reports a wider range of usability metrics for the HCI literature [55], some of which are more detailed subsets of the metrics found in our sample. For example, Hornbæk reports mental effort, communication effort, and information accessed, but we include all of them in *workload*. Other usability metrics from Hornbæk’s review, such as *quality of outcome*, *binary task completion*, and *input rate*, were rare or absent in our sample.

F3—Sensory. The metrics in this facet evaluate the interaction or robot in relation to the human basic senses such as user perception of *visual*, *auditory*, or *physical* aspects. A few studies ask about attributes that are judged visually, while only 3 studies (out of 44) ask users to rate touch sensory properties. Only one study includes a rating of auditory components [124]. Our sample of studies did not include any metrics related to the sense of smell or taste. We could not find an exact match for this facet in the HCI literature. Metrics related to esthetics or appeal in the HCI literature [15] and the autotelics from recent work on haptic experience [67] seem the most related to this facet.

F4—Personal and Interpersonal. This facet evaluates the robot as an autonomous being. The metrics cover the judgment of the robot’s characteristics, capability, or social attributes and the judgment of the robot’s joint interaction with the user or *teamwork*. Only 5 studies ask users to rate *teamwork* when 20 studies have a mutual touch between the user and the robot. Overall, the personal and interpersonal metrics are well represented in our sample with many of the rated statements (142 out of 317). In contrast, these metrics are not present in the HCI literature.

F5—Experiential. These metrics deal with the user’s feelings, are more holistic, and are not focused on a task. The metrics in this facet overlap with the UX metrics reported by Bargas-Avila and Hornbæk, but we note a different distribution in our sample. Among the metrics, *enjoyment*, *emotion*, and *motivation* are present in both our and their HCI sample. In contrast, the UX of *safety* is the most common metric in our sample, but it is absent from their review. The focus on *safety* in the pHRI literature is perhaps due to the possibility of user injury in physical interactions with robots compared to the minimal risk of interacting with visual and auditory interfaces common in HCI. Surprisingly, the experience of *autonomy* is only included in one study in our sample.

The *experiential* (F5) and *personal and interpersonal* (F4) metrics are included the most in our sample, followed by the *usability* (F2) and *overall* (F1) metrics. The sensory metrics (F3) are evaluated the least in our sample.

Analyzing the distribution of UX metrics employed in the pHRI studies highlights that researchers are often interested in more than a single aspect of UX but have difficulty delineating their UX goals. While some studies focused on a single UX metric such as emotions (e.g., [9, 125]) or social traits or behavior (e.g., [34]), most studies ($n > 30$) in our sample collected data on more than one UX metric. In several cases, the studies collect data on UX metrics spanning across both task-centered *usability* metrics (F2) as well as metrics in the *personal and interpersonal* or *experiential* facets (F4, F5). For example, for human–robot collaboration in an assembly task, Gleeson et al. collected data across 11 UX metrics ranging from ease of use, time, and workload of the task to the physical sensation, the robot’s personal and interpersonal traits, and user enjoyment [42]. However, the goal statements are often either broadly defined or are not directly linked to the study’s metrics. For example, some studies aimed to assess if a touch condition was “positive” or “favorable” but collected user responses for several UX metrics (e.g., 16 metrics) across all the facets. Defining the metrics and facets of the pHRI experience can provide researchers with a lexicon to clearly specify and link their goals to data collection.

5 Discussion

Below, we present a conceptual model for the pHRI experience based on the above analysis. Then, we discuss implications for future work based on the trends and gaps in the literature and reflect on the limitations of this work.

5.1 A Conceptual Model of pHRI Experience

We propose a conceptual model for pHRI (Figure 1) with three main components: (1) design parameters, (2) interaction timeline, and (3) UX metrics. While this article focuses on the UX of pHRI, the first two components can directly impact UX and thus are included in our conceptual model. These components can help designers and researchers describe an interaction, use it to generate ideas about alternative designs and evaluate the UX of the interaction. Below, we briefly describe each component and discuss its utility for pHRI research.

(1) *Design Parameters*: The first component of the model depicts the pHRI design parameters around the user, robot, and their interaction. Researchers can manipulate the design parameters to study their impact on UX. For instance, one may design the pHRI for a specific demographic user group or compare the UX of two user groups. For example, Law et al. studied the impact of a user’s gender on their trust in a robot after watching a video of the robot tapping a person on the shoulder [73]. Others could study the impact of user expertise or beliefs (e.g., negative attitude toward robots [63]) on the pHRI experience. Studies of pHRI frequently manipulate physical and overall interaction parameters (Table 2). Finally, several design parameters exist for the robot, such as its multisensory presentation (e.g., robot’s appearance, gaze, and sound) or its social parameters of the robot’s role, intention, or attitude.

Design parameters have both descriptive and generative power by nature. The list of parameters and their values can be used to clearly describe what is manipulated or kept constant in an interaction. The parameters can also help generate ideas for new interactions. For example, how does the robot’s appearance or form impact UX outcomes? As another example, all the studies in our sample focus on the interaction between one human and one robot. Yet, one can examine UX where a team of humans or robots are present [68].

(2) *Timeline*: The second conceptual component is the interaction timeline, which enables researchers to describe the sequence and timing of physical and overall interactions. The temporal description of pHRI in our sample of studies is surprisingly incomplete. Duration of touch stimulation is a salient parameter for a haptic signal [127]. Yet, many studies in our review do not report the duration of their physical interactions. The interaction sequence is often described qualitatively and sometimes mixed with the description of the study procedure. The timeline component provides descriptive power for pHRI studies and emphasizes the impact of previous interactions, such as verbal utterances, on the pHRI experience.

Second, a timeline view of pHRI has generative power. A few studies in our sample showed the importance of temporal parameters (e.g., motion speed and duration) and the sequence of verbal and touch modalities [32] on UX outcomes. The temporal view is in line with work on robot sensing and planning algorithms that are sequential in nature. In robot sensing, the difference between touch gestures is in the spatiotemporal signature of the touch signals obtained from the sensor [24, 26]. Robot planning frameworks such as partially observable Markov decision processes [72] and reinforcement learning are inherently sequential. Considering interaction as sequential decision-making is also gaining traction in computational models of human interaction with technology [96]. Sequential descriptions of pHRI can facilitate knowledge transfer from pHRI studies to computational simulations of pHRI interactions and results.

(3) *UX Metrics*: The third component of the model is the UX metrics and their five facets. This component captures the complex and multidimensional aspect of UX in pHRI. The UX metrics are obtained from the statements and measurements in the studies and divided into overarching facets.

In contrast to the design parameters, the relationship between the UX metrics and the user, robot, and interaction entities is complex. Specifically, the *overall* (F1) metrics can apply to any of the human, robot, or interaction entities, the *usability* (F2) metrics can apply to the robot or

the interaction, the *sensory* (F3) and *personal and interpersonal* (F4) metrics mainly apply to the evaluation of the robot or the joint work, and the *experiential* metrics (F5) can apply to the human or the robot (e.g., evaluating emotion of the robot). Thus, we represent them separately from the user, robot, and interaction entities in the conceptual model. Among these metrics, the first three facets (*overall, usability*, and *sensory*) can apply to interactions with computers, *personal and interpersonal* metrics only apply to robots or virtual agents, and some metrics from the *experiential* facet such as *safety*, and *autonomy* mainly apply to pHRI or interactions with **Artificial Intelligence (AI)**.

The five pHRI facets are meant to provide an initial guide, rather than fixed categories, to evaluate various aspects of UX. Thus, the metrics in the five facets can relate and overlap. For example, *ease of use* can refer to the evaluation of an interface or task and also convey a subjective feeling of ease, making the metric relevant to both the *usability* (F2) and the *experiential* (F5) facets. Similarly, the usability and UX metrics are not strictly separate and overlap in the HCI literature. In the proposed model, we included *ease of use* in the *usability* metrics for two reasons. First, the *ease of use* statements in our sample refer to the evaluation of the task or the robot as a tool when completing a task, in line with our definition of the *usability* facet. Second, this categorization is aligned with existing HCI literature where *ease of use* is regarded as a usability rather than an UX metric [55], enabling the development of shared theories across the pHRI and HCI fields.

The UX metrics and their facets provide descriptive and evaluative power for pHRI research. The metrics can help researchers better describe the goal(s) of their pHRI interactions and evaluate the relevant metrics in empirical studies. Also, the metrics and their underlying statements (see supplementary materials) provide a starting point for creating a validated questionnaire for pHRI experience. The design parameters and the interaction timeline influence the UX outcomes. Yet, how these parameters impact UX outcomes cannot be currently established due to the methodological variations and gaps in the pHRI studies.

Using the Conceptual Model of pHRI Experience. We anticipate three ways that future work can use and build on this conceptual model to further highlight nuances of designing and evaluating physical interactions with robots.

First, researchers can use the model to identify gaps and define research directions. For example, an open question is the impact of different design parameters on the pHRI experience. Touch is often described in the literature as a personal and emotional communication channel when compared to the audio and visual modalities [83]. Future work can test whether this assumption holds in pHRI by designing experiments where a robot communicates emotional support through physical contact vs. other channels (e.g., gaze, facial expressions) and comparing user evaluation of personal/interpersonal and experiential metrics (F4, F5). Another important area for future work would be to investigate the impact of temporal parameters such as the duration, repetition, and influence of previous interactions on the UX of pHRI. Also, the combination of the conceptual model and the review of pHRI studies in our sample highlight gaps and open challenges for future research as we discuss in the next section (Section 5.2).

Second, the model can serve as an initial structure for promoting a shared understanding and communication among researchers with various engineering and social experiential research directions in the field. By charting the pHRI space with a set of parameters, the model outlines what and how to report pHRI experiments. The model can also help teaching and education in the field, allowing students to identify and discuss the design considerations in pHRI experiments.

Third, we hope pHRI researchers improve the model based on their empirical results. With improved reporting of design parameters and UX metrics, future research can revise the conceptual model to establish stronger links between different components. For example, our model currently

divides the design parameters according to the user, the robot, and their interaction, but future empirical studies may lead to a hierarchical structure for the design parameters that conveys the relative importance of the parameters on UX outcomes. Also, future studies of the temporal parameters of pHRI can help further specify the timeline component. Eventually, these descriptive models can form the basis for predictive and computational models of pHRI experience.

5.2 Implications for Future Research

We summarize gaps in the pHRI studies and discuss avenues for further charting the UX of pHRI.

Physical Interaction Parameters. Our review highlights three gaps in studying the physical parameters of pHRI. First, little work exists on the effect of unintended and accidental robot contact on users. Only one study in our sample included unintended contact. As robot hardware and algorithms are susceptible to error, studying the UX of accidental contact and other haptic errors (e.g., force variations) is an important area for future work. Studying accidental contact in an ecologically valid way is challenging and would require a careful study design. We anticipate advances in predicting user movement trajectories [30, 44] and detecting anomalies from robot sensor data [99] can help in conducting such studies in the future.

Second, our sample did not have any studies where robot-initiated touch guides or teaches movement to humans. Such physical interactions are particularly relevant for rehabilitation and skill training scenarios. Robotic devices for physical rehabilitation are often not (perceived as) autonomous systems. Also, rehabilitation studies often focus on hardware design and clinical outcomes [40, 81]. Future work should assess the UX of being guided by robot-initiated touch.

Third, the studies in our sample often involved short physical interaction episodes (e.g., seconds to minutes). We conjecture that longer interactions may be possible depending on the form of the robot. For example, users may hold and touch robotic pets such as PARO for longer (e.g., 30 minutes) and multiple sessions over several weeks or months [102]. Also, autonomous vehicles or automatic beds that simulate the feel and breathing of a human body [51] may be viewed as a form of pHRI; these technologies can sense the environment and “touch” the user, and the duration of physical contact can extend to hours or days. Do users perceive these systems as robots? The relationship between the UX of touching and being touched by these technologies and conventional robots is an open question for future work.

UX of Observed vs. Real Touch. Our sample included five studies of observed touch where participants watched videos of physical interactions between a human and a robot. A few studies in haptics and HRI show that people can infer tactile sensations and their emotional associations through vision [17, 57, 108, 114, 120]. These studies are in line with neuroscience research suggesting that a subset of motor neurons is fired for the observation or execution of actions, offering a neural basis for human empathic response [65]. On the other hand, a recent study showed that observing robot touch in a video vs. receiving robot touch in the lab can result in different emotional reactions in users [70]. Little is known about the extent to which user evaluations of observed and actual pHRI are consistent or different for various facets and metrics of the pHRI experience. Thus, future work should further investigate the relationship between observed vs. felt pHRI.

Generalizing Beyond a Single Robot. The impact of robot form and body materials on the pHRI experience is another open question. For example, users may experience the same tapping gestures differently depending on whether the robot is human or machine-like or whether the robot’s body is made of hard or soft materials. The studies in our sample used a single robot (except for Block et al. [18]) while acknowledging that their results may not generalize to other robots. The variety of robots used in the pHRI studies prevents the community from deriving generalizable findings and formulating guidelines for pHRI design. One obvious solution is to test the same physical

interaction(s) with several robots. Yet, robots can vary across many parameters. Researchers often have access to one or a few robots, and adding a new robot can notably increase the time or number of participants for a study. These factors make this problem intractable in a lab setting and for a single research group.

We anticipate two possible solutions to this problem. One solution would be to crowdsource this task across multiple HRI research labs that follow one predefined reproducible interaction protocol. Results of studies from different research labs can be compiled in a database and pulled to assess generalizable trends. An alternative solution can aim to identify archetypal robot forms, create visual proxies of the physical interactions (e.g., with video, virtual reality), and collect data on UX at scale. Recent collections of robots [1, 2, 3, 4, 108] can help HRI researchers formalize archetypal robot forms and their variations. Then, designers create virtual models of archetypal robot forms and program the interactions with a human avatar to collect user responses. Since physical feedback is not present, the collected data can only be an estimate of UX. They must be combined with smaller-scaler in-lab studies to assess the validity of the estimates. Progress in haptic feedback in virtual environments can further help with addressing this problem in the future [116].

Developing and Validating a Questionnaire for pHRI Experience. Our review highlights the need and provides a starting point for developing an UX questionnaire for pHRI. Existing questionnaires are either adapted from other fields or cover a small subset of the long list of metrics that matter to pHRI researchers (Table 4). On the other hand, the relationships between different pHRI metrics are unknown. Which ones correlate? What tradeoffs exist among different UX metrics? Future work on developing a pHRI questionnaire can shed light on these relationships and tradeoffs.

Establishing a valid questionnaire has an extensive process with several rounds of development and validation with many users [94]. Our work provides a list of UX metrics with their relevant statements for developing the questionnaire and a summary of use cases against which the instrument could be validated. The list of rated statements can be expanded through common user research methods (e.g., brainstorming, scenarios) to cover underrepresented dimensions (e.g., autonomy) in the initial phases of questionnaire development. Recently, HRI researchers have successfully used crowdsourcing platforms such as Amazon Mechanical Turk to develop and validate questionnaires such as the RoSAS based on images of robots [29]. Similarly, pHRI researchers may use crowdsourcing with videos of physical interaction use cases to collect responses at scale for the development and initial validation of a pHRI questionnaire. However, given the importance of haptic feedback in pHRI, such a questionnaire must be further validated through in-lab studies.

Going Beyond Ratings of UX. The pHRI literature can further develop thick qualitative descriptions and objective measures of UX. Most pHRI studies in our sample use subjective ratings to capture UX of their physical interactions. While valid, this focus is limiting. This approach falls between the quantitative measurement-based view common in robotics and the qualitative social science view that provides in-depth descriptions of a complex phenomenon. We argue that more pHRI research is needed on the two ends of the spectrum of the qualitative-to-quantitative data collection methods.

On the qualitative end of the spectrum, in-depth interviews and contextual inquiry can provide thick subjective descriptions of the UX of pHRI and shed light on design parameters and qualities that researchers may overlook. For example, the sense of *autonomy*, *teamwork*, and *symbolic* experience of pHRI are hardly captured in the studies. Qualitative studies can further highlight the nuances of these dimensions to inform future work. Interestingly, the review by Bargas-Avila suggests that such qualitative studies are prevalent in the HCI literature [15], and they have been used to highlight nuances of UX with other technological artifacts [71].

On the quantitative end of the spectrum, we need more research on identifying objective behavioral measures of UX. User ratings are hard to collect and potentially disruptive during the interaction, limiting the data to before and after the interaction. As such, the temporal view of physical interactions with robots is still poorly understood. A few studies in our review incorporated measures of user gaze [39, 63] or contact force [34], yet objective metrics of pHRI experience are sparse and hard to link to subjective ratings. For instance, user ratings of *safety* are often collected in pHRI studies. Yet, the relationship between user perception of safety and existing definitions and safety standards in the robotic literature [43] is underexplored. Devising objective or behavioral pHRI metrics and linking them to subjective UX metrics remains an open challenge for future work.

Adapting Robot Behavior to UX. Metrics of UX are mainly used as an outcome measure to compare predefined interaction conditions. How can UX metrics guide robot planning and learning during the interaction? How can robots adapt their touch actions to the user’s state and decide when the touch could be useful or conversely recognize if touching is undesirable by the user? Past pHRI studies have adapted the robot motion to the user body (e.g., size) or movement (e.g., user speed) [18, 58] with positive outcomes. Yet, the extension of this approach to UX metrics is not apparent. As mentioned above, a possible direction is establishing a link between behavioral measures (e.g., user gaze) and subjective ratings and incorporating them as an optimization parameter in robot learning and planning algorithms. Another approach may use advances in conversational agents to obtain an estimate of UX from natural dialogues during the interaction. Finally, a third approach can build on the pioneering work of Card and Moran [27, 28] and its extensions [75, 96] to predict human behavior in a task. These studies attempt to model the human interaction partner as an agent with perceptual [66], cognitive [56], and motor constraints [82] to predict the human’s actions in a task. Extensions to model human emotion are also being developed [89]. While these computational models are still in their infancy and cannot capture the full range of UX (e.g., motivation), they provide a path for adapting pHRI algorithms to user behavior and experience.

In the last 2 years, 11 other pHRI studies were published in our 3 target venues, showing similar trends to studies covered in our review. Specifically, several studies focused on communicating emotions with robots [13, 20, 128] or predicting user perception of robots before and after touch [108]. In other studies, the physical interaction aimed at completing a task, including two studies on object handover between a human and a robot [36, 60] and a co-manipulation scenario where the user and robot hold and move an object together [85]. Finally, four studies focused on teaching or guiding movement [61, 77, 101, 122]. Interestingly, in three of these studies, the robot was guiding a human [61, 101, 122] each with a different robot form: a drone that corrects user motion during an exercise [122], a robotic cane that can guide a visually impaired user [101], and a humanoid robot that guides the user while walking hand-in-hand [61]. These studies target one of the gaps in the literature where a robot guides or teaches movement to humans. The UX methods and metrics in these 11 pHRI studies were also similar to trends reported in our review. Most of the studies ($n = 7$) used custom questionnaires, followed by NASA TLX questionnaire for workload, and a few studies reported measures of task accuracy, gaze or blinking behavior, or physiological recordings.

5.3 Reflecting on the pHRI Experience

The UX metrics used in pHRI vs. HCI suggest key differences and shared characteristics for user interactions with robots vs. other technologies.

Most UX metrics in our sample fall in the personal/interpersonal facet (F4) while these metrics were absent from prior surveys of UX in HCI. Perhaps the frequent inclusion of personal/interpersonal metrics in pHRI reflects the common belief that robots are social agents. Previous work has shown that people evaluate robots in the same way they evaluate other humans according to

competence, warmth, and comfort [103]. This social view of robots is also reinforced by popular fiction [37]. Relatedly, the design of robots' actions is often inspired by or modeled after human-human or human-animal interactions. Another important factor in pHRI experience is the robot's physicality and autonomy. Young et al. argued that the robot's physical presence and autonomy to act in the personal and social spaces of humans can convey a sense of agency and intentionality often attributed to living creatures [129]. These factors are emphasized in pHRI since physical modifications of the world can signal competence, and touching is often associated with personal or social behavior.

Other key differences are related to the UX of *safety* (from F5), *frustration*, and *sensory* (F3) metrics in the pHRI and HCI literature. The focus on safety in our sample is aligned with the technical research in pHRI that aims to quantify and reduce the potential for harm and injury to users [43]. These safety considerations are less relevant in HCI when the user interacts with software-based user interfaces. In contrast, a common UX in HCI is *frustration* during computer use [53], but this aspect of UX is absent in our sample of pHRI literature. HRI studies have started to investigate user *frustration* in HRI [90, 118], which can help develop relevant measures for pHRI research as well. Also, the *sensory* facet has been a focus in the multisensory HCI and haptics research communities for decades but is rarely present in our sample of the pHRI literature. This facet presents an opportunity for leveraging HCI literature to inform pHRI design and research.

The *overall* (F1), *usability* (F2), and *experiential* (F5) metrics were frequently captured in both the pHRI and HCI literature. The inclusion of usability metrics suggests that robots can also be perceived and evaluated as a computer interface. Thus, the social vs. device appraisal of robots may depend on the interaction context and applications. In the HCI literature, the experiential factors are described as the "third wave in HCI" focusing on aesthetics, affective interaction, and the embodied and contextual aspects of human activity and experience [21, 37, 47]. These considerations overlap with the mixing of technical and social considerations when designing pHRI.

5.4 Limitations

Our work has three main limitations that can be addressed in future work. First, we scoped our sample to publications in three top-tier HRI venues. This decision was pragmatic. Given no prior systematic review of pHRI, we found it important for our analysis to cover various aspects of the user studies (e.g., goals, data collection methods), parameters of physical interactions, and UX metrics. Thus, we narrowed our sample to these three venues that focus on human-subject studies in robotics. Yet, pHRI studies with UX metrics are sometimes published in other robotics, engineering (e.g., rehabilitation, haptics), or interaction design venues (e.g., HCI, affective computing). Future systematic reviews can complement our work by analyzing one aspect (e.g., physical interaction parameters) in a larger sample to assess whether our reported trends can generalize across venues. Second, we obtained the UX metrics and their facets through affinity diagramming and discussion. Affinity diagramming is an established method for qualitative data analysis [46]. We acknowledge the active role of the researchers in generating the clusters from the data and the inherent subjectivity of our analysis method [23]. To support future work, we provide all the rated and measured items annotated with their UX metrics from our analysis as supplementary material. Third, our review focused on interactions with robots that are autonomous or perceived as autonomous by users. Yet, robot autonomy is a spectrum from user-controlled (e.g., in grounded force-feedback haptic devices or teleoperation) to semi-autonomous (e.g., shared-control robots) to fully autonomous robots. Future work can extend the proposed conceptual model to capture interactions with robots with varying levels of autonomy.

6 Conclusion

Physical interactions with robots have been an active area of research for the last two decades. However, little is known about how to think about, design, and evaluate these interactions systematically. This work presents a systematic review of 44 studies that vary in their use cases of pHRI and provides a conceptual model of the pHRI experience. Our analysis highlights common trends and underexplored areas in the literature. We hope our results pave the way for future theories, empirical studies, and evaluation questionnaires for touch interactions with robots.

Acknowledgments

We thank Tor-Salve Dalsgaard for providing the script to filter results from the Springer website. We thank the reviewers and our colleagues at Arizona State University for their inputs on the manuscript.

References

- [1] ABOT. 2022. The Anthropomorphic Robot Database. Retrieved from <http://abotdatabase.info/>. [Accessed 5 December 2022].
- [2] ROBOTS. 2022. IEEE Robot Database. Retrieved from <https://robots.ieee.org/>. [Accessed 5 December 2022].
- [3] OSF. 2022. Stanford Social Robot Collection. Retrieved from <https://osf.io/hz7p3>. [Accessed 5 December 2022].
- [4] RobotHands. 2022. A Growing Database of Robot Hands. Retrieved from <http://robohands.org/>. [Accessed 5 December 2022].
- [5] Baris Akgun, Maya Cakmak, Jae Wook Yoo, and Andrea Lockerd Thomaz. 2012. Trajectories and keyframes for kinesthetic teaching: A human–robot interaction perspective. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction (HRI '12)*. ACM, New York, NY, 391–398. DOI: <https://doi.org/10.1145/2157689.2157815>
- [6] Nilgun Aksan, Buke Kisaç, Mufit Aydin, and Sumeyra Demirbukten. 2009. Symbolic interaction theory. *Procedia - Social and Behavioral Sciences* 1, 1 (Jan. 2009), 902–904. DOI: <https://doi.org/10.1016/j.sbspro.2009.01.160>
- [7] Beatrice Alenljung, Jessica Lindblom, Rebecca Andreasson, and Tom Ziemke. 2019. User experience in social human–robot interaction. In *Rapid Automation: Concepts, Methodologies, Tools, and Applications*. IGI Global, 1468–1490. DOI: <https://doi.org/10.4018/978-1-5225-8060-7.ch069>
- [8] Jacopo Aleotti, Vincenzo Micelli, and Stefano Caselli. 2014. An affordance sensitive system for robot to human object handover. *International Journal of Social Robotics* 6, 4 (Nov. 2014), 653–666. DOI: <https://doi.org/10.1007/s12369-014-0241-3>
- [9] Mehdi Ammi, Virginie Demulier, Sylvain Caillou, Yoren Gaffary, Yacine Tsalamal, Jean-Claude Martin, and Adriana Tapus. 2015. Haptic human–robot affective Interaction in a handshaking social protocol. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction (HRI '15)*. ACM, New York, NY, 263–270. DOI: <https://doi.org/10.1145/2696454.2696485>
- [10] Rebecca Andreasson, Beatrice Alenljung, Erik Billing, and Robert Lowe. 2018. Affective touch in human–robot interaction: Conveying emotion to the NAO robot. *International Journal of Social Robotics* 10, 4 (Sep. 2018), 473–491. DOI: <https://doi.org/10.1007/s12369-017-0446-3>
- [11] Brenna D. Argall and Aude G. Billard. 2010. A survey of tactile human–robot Interactions. *Robotics and Autonomous Systems* 58, 10 (Oct. 2010), 1159–1176. DOI: <https://doi.org/10.1016/j.robot.2010.07.002>
- [12] Thomas Arnold and Matthias Scheutz. 2018. Observing robot touch in context: How does touch and attitude affect perceptions of a robot's social qualities?. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction (HRI '18)*. ACM, New York, NY, 352–360. DOI: <https://doi.org/10.1145/3171221.3171263>
- [13] Ali Asadi, Oliver Niebuhr, Jonas Jørgensen, and Kerstin Fischer. 2022. Inducing changes in breathing patterns using a soft robot. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction (HRI '22)*. IEEE, 683–687.
- [14] Wilma A. Bainbridge, Shunichi Nozawa, Ryohei Ueda, Kei Okada, and Masayuki Inaba. 2012. A methodological outline and utility assessment of sensor-based biosignal measurement in human–robot interaction. *International Journal of Social Robotics* 4, 3 (Aug. 2012), 303–316. DOI: <https://doi.org/10.1007/s12369-012-0146-y>
- [15] Javier A. Bargas-Avila and Kasper Hornbæk. 2011. Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, 2689–2698. DOI: <https://doi.org/10.1145/1978942.1979336>

- [16] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. DOI: <https://doi.org/10.1007/s12369-008-0001-3>
- [17] Elisabeth Baumgartner, Christiane B. Wiebel, and Karl R. Gegenfurtner. 2013. Visual and haptic representations of material properties. *Multisensory Research* 26, 5 (2013), 429–455.
- [18] Alexis E. Block, Sammy Christen, Roger Gassert, Otmar Hilliges, and Katherine J. Kuchenbecker. 2021. The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. ACM, New York, NY, 380–388. DOI: <https://doi.org/10.1145/3434073.3444656>
- [19] Alexis E. Block and Katherine J. Kuchenbecker. 2019. Softness, warmth, and responsiveness improve robot hugs. *International Journal of Social Robotics* 11, 1 (Jan. 2019), 49–64. DOI: <https://doi.org/10.1007/s12369-018-0495-2>
- [20] Alexis E. Block, Hasti Seifi, Otmar Hilliges, Roger Gassert, and Katherine J. Kuchenbecker. 2023. In the arms of a robot: Designing autonomous hugging robots with intra-hug gestures. *ACM Transactions on Human-Robot Interaction* 12, 2 (2023), 1–49.
- [21] Susanne Bødker. 2006. When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles (NordiCHI '06)*, 1–8.
- [22] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (Mar. 1994), 49–59. DOI: [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- [23] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide*. SAGE.
- [24] Rachael Bevill Burns, Hyosang Lee, Hasti Seifi, Robert Faulkner, and Katherine J. Kuchenbecker. 2022. Endowing a NAO robot with practical social-touch perception. *Frontiers in Robotics and AI* 9 (Apr. 2022). DOI: <https://www.frontiersin.org/articles/10.3389/frobt.2022.840335>
- [25] Maya Cakmak, Siddhartha S. Srinivasa, Min K. Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Using spatial and temporal contrast for fluent robot-human hand-overs. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '11)*. ACM, New York, NY, 489–496. DOI: <https://doi.org/10.1145/1957656.1957823>
- [26] Xi L. Cang, Paul Bucci, Andrew Strang, Jeff Allen, Karon MacLean, and H. Y. Sean Liu. 2015. Different strokes and different folks: Economical dynamic surface sensing and affect-related touch recognition. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, 147–154. DOI: <https://doi.org/10.1145/2818346.2820756>
- [27] Stuart K. Card (Ed.). 2017. *The Psychology of Human-Computer Interaction*. CRC Press, Boca Raton. DOI: <https://doi.org/10.1201/9780203736166>
- [28] Stuart K. Card and Thomas P. Moran. 1988. User technology: From pointing to pondering. In *Proceedings of the ACM Conference on The History of Personal Workstations (HPW '88)*. ACM, New York, NY, 489–526. DOI: <https://doi.org/10.1145/61975.66925>
- [29] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The robotic social attributes scale (RoSAS): Development and validation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, 254–262. DOI: <https://doi.org/10.1145/2909824.3020208>
- [30] Konstantinos Charalampous, Ioannis Kostavelis, and Antonios Gasteratos. 2017. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems* 93 (Jul. 2017), 85–104. DOI: <https://doi.org/10.1016/j.robot.2017.03.002>
- [31] Tiffany L. Chen and Charles C. Kemp. 2010. Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*. ACM, New York, NY, 367–374. DOI: <https://doi.org/10.1109/HRI.2010.5453162>
- [32] Tiffany L. Chen, Chih-Hung Aaron King, Andrea L. Thomaz, and Charles C. Kemp. 2014. An investigation of responses to robot-initiated touch in a nursing context. *International Journal of Social Robotics* 6, 1 (Jan. 2014), 141–161. DOI: <https://doi.org/10.1007/s12369-013-0215-x>
- [33] Josep-Arnau Claret, Gentiane Venture, and Luis Basañez. 2017. Exploiting the robot kinematic redundancy for emotion conveyance to humans as a lower priority task. *International Journal of Social Robotics* 9, 2 (Apr. 2017), 277–292. DOI: <https://doi.org/10.1007/s12369-016-0387-2>
- [34] Sandra Costa, Hagen Lehmann, Kerstin Dautenhahn, Ben Robins, and Filomena Soares. 2015. Using a humanoid robot to elicit body awareness and appropriate physical interaction in children with autism. *International Journal of Social Robotics* 7, 2 (Apr. 2015), 265–278. DOI: <https://doi.org/10.1007/s12369-014-0250-2>
- [35] Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi. 2008. An atlas of physical human–robot interaction. *Mechanism and Machine Theory* 43, 3 (Mar. 2008), 253–270. DOI: <https://doi.org/10.1016/j.mechmachtheory.2007.03.003>

- [36] Tair Faibis, Alap Kshirsagar, Guy Hoffman, and Yael Edan. 2022. Human preferences for robot eye gaze in human-to-robot handovers. *International Journal of Social Robotics* 14, 4 (2022), 995–1012.
- [37] Ylva Fernaeus, Sara Ljungblad, Mattias Jacobsson, and Alex Taylor. 2009. Where third wave HCI meets HRI: Report from a workshop on user-centred design of robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (HRI '09)*. 293–294.
- [38] Kerstin Fischer, Franziska Kirstein, Lars C. Jensen, Norbert Krüger, Kamil Kukliński, Maria Vanessa aus der Wieschen, and Thiusius R. Savarimuthu. 2016. A comparison of types of robot control for programming by demonstration. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '16)*. ACM, New York, NY, 213–220. DOI: <https://doi.org/10.1109/HRI.2016.7451754>
- [39] Naomi T. Fitter and Katherine J. Kuchenbecker. 2020. How does it feel to clap hands with a robot? *International Journal of Social Robotics* 12, 1 (Jan. 2020), 113–127. DOI: <https://doi.org/10.1007/s12369-019-00542-x>
- [40] Roger Gassert and Volker Dietz. 2018. Rehabilitation robots for the treatment of sensorimotor deficits: A neurophysiological perspective. *Journal of NeuroEngineering and Rehabilitation* 15, 1 (Jun. 2018), 46. DOI: <https://doi.org/10.1186/s12984-018-0383-x>
- [41] Manuel Giuliani and Alois Knoll. 2013. Using embodied multimodal fusion to perform supportive and instructive robot roles in human-robot interaction. *International Journal of Social Robotics* 5, 3 (Aug. 2013), 345–356. DOI: <https://doi.org/10.1007/s12369-013-0194-y>
- [42] Brian Gleeson, Katelyn Currie, Karon MacLean, and Elizabeth Croft. 2015. Tap and push: Assessing the value of direct physical control in human-robot collaborative tasks. *Journal of Human-Robot Interaction* 4, 1 (Jul. 2015), 95–113. DOI: <https://doi.org/10.5898/JHRI.4.1.Gleeson>
- [43] Sami Haddadin and Elizabeth Croft. 2016. Physical human–robot interaction. In *Springer Handbook of Robotics*. Bruno Siciliano and Oussama Khatib (Eds.), Springer International Publishing, 1835–1874. DOI: https://doi.org/10.1007/978-3-319-32552-1_69
- [44] Mahmoud Hamandi, Mike D'Arcy, and Pooyan Fazli. 2019. DeepMoTion: Learning to navigate like humans. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN '19)*. 1–7. DOI: <https://doi.org/10.1109/RO-MAN46459.2019.8956408>
- [45] Zhao Han and Holly Yanco. 2019. The effects of proactive release behaviors during human-robot handovers. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. ACM, New York, NY, 440–448. DOI: <https://doi.org/10.1109/HRI.2019.8673085>
- [46] Gunnar Harboe and Elaine M. Huang. 2015. Real-world affinity diagramming practices: Bridging the paper-digital gap. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, 95–104. DOI: <https://doi.org/10.1145/2702123.2702561>
- [47] Steve Harrison, Deborah Tatar, and Phoebe Sengers. 2007. The three paradigms of HCI. In *Proceedings of the Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems*. 1–18.
- [48] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*. Peter A. Hancock and Najmedin Meshkati (Eds.), Human Mental Workload, Vol. 52, 139–183. DOI: [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [49] Marc Hassenzahl. 2013. User experience and experience design. *The Encyclopedia of Human-Computer Interaction* 2 (2013), 1–14.
- [50] Marc Hassenzahl and Noam Tractinsky. 2006. User experience—A research agenda. *Behaviour & Information Technology* 25, 2 (Mar. 2006), 91–97. DOI: <https://doi.org/10.1080/01449290500330331>
- [51] Sabrina Hauser, Melinda J. Suto, Liisa Holsti, Manon Ranger, and Karon E. MacLean. 2020. Designing and evaluating calmer, a device for simulating maternal skin-to-skin holding for premature infants. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–15. DOI: <https://doi.org/10.1145/3313831.3376539>
- [52] Erin Hedlund, Michael Johnson, and Matthew Gombolay. 2021. The effects of a robot's performance on human teachers for learning from demonstration tasks. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. ACM, New York, NY, 207–215. DOI: <https://doi.org/10.1145/3434073.3444664>
- [53] Morten Hertzum and Kasper Hornbæk. 2023. Frustration: Still a common user experience. *ACM Transactions on Computer-Human Interaction* 30, 3 (2023), 1–26.
- [54] Takahiro Hirano, Masahiro Shiomi, Takamasa Iio, Mitsuhiro Kimoto, Ivan Tanev, Katsunori Shimohara, and Norihiro Hagita. 2018. How do communication cues change impressions of human–robot touch interaction? *International Journal of Social Robotics* 10, 1 (Jan. 2018), 21–31. DOI: <https://doi.org/10.1007/s12369-017-0425-8>

- [55] Kasper Hornbæk. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies* 64, 2 (Feb. 2006), 79–102. DOI: <https://doi.org/10.1016/j.ijhcs.2005.06.002>
- [56] Andrew Howes, Geoffrey B. Duggan, Kiran Kalidindi, Yuan-Chi Tseng, and Richard L. Lewis. 2016. Predicting short-term remembering as boundedly optimal strategy choice. *Cognitive Science* 40, 5 (Jul. 2016), 1192–1223. DOI: <https://doi.org/10.1111/cogs.12271>
- [57] Yuhan Hu and Guy Hoffman. 2019. Using skin texture change to design emotion expression in social robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. 2–10.
- [58] Chien-Ming Huang, Maya Cakmak, and Bilge Mutlu. 2015. Adaptive coordination strategies for human-robot handovers. In *Proceedings of the Robotics: Science and Systems*. 1–10. DOI: <https://doi.org/10.15607/RSS.2015.XI.031>
- [59] Claire A. G. J. Huijnen, Hanneke A. M. D. Verreussel-Willen, Monique A. S. Lexis, and Luc P. de Witte. 2021. Robot KASPAR as mediator in making contact with children with autism: A pilot study. *International Journal of Social Robotics* 13, 2 (Apr. 2021), 237–249. DOI: <https://doi.org/10.1007/s12369-020-00633-0>
- [60] Francesco Iori, Gojko Perovic, Francesca Cini, Angela Mazzeo, Egidio Falotico, and Marco Controzzi. 2023. DMP-based reactive robot-to-human handover in perturbed scenarios. *International Journal of Social Robotics* 15, 2 (2023), 233–248.
- [61] Naoki Ise, Yoshihiro Nakata, Yutaka Nakamura, and Hiroshi Ishiguro. 2022. Gaze motion and subjective workload assessment while performing a task walking hand in hand with a mobile robot. *International Journal of Social Robotics* 14, 8 (2022), 1875–1882.
- [62] J ISO 9241–210. 2010. Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems. *Isotopenpraxis* 2010 (2010), 1–19. Retrieved from <https://www.iso.org/standard/77520.html>
- [63] Serena Ivaldi, Sébastien Lefort, Jan Peters, Mohamed Chetouani, Joelle Provost, and Elisabetta Zibetti. 2017. Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human–robot assembly task. *International Journal of Social Robotics* 9, 1 (Jan. 2017), 63–86. DOI: <https://doi.org/10.1007/s12369-016-0357-8>
- [64] Mitsuhiro Kamezaki, Ayano Kobayashi, Yuta Yokoyama, Hayato Yanagawa, Moondeep Shrestha, and Shigeki Sugano. 2020. A preliminary study of interactive navigation framework with situation-adaptive multimodal inducement: Pass-by scenario. *International Journal of Social Robotics* 12, 2 (May 2020), 567–588. DOI: <https://doi.org/10.1007/s12369-019-00574-3>
- [65] Christian Keysers and Valeria Gazzola. 2009. Expanding the mirror: Vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology* 19, 6 (2009), 666–671.
- [66] David E. Kieras and Anthony J. Hornof. 2014. Towards accurate and practical predictive models of active-vision-based visual search. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, 3875–3884. DOI: <https://doi.org/10.1145/2556288.2557324>
- [67] Erin Kim and Oliver Schneider. 2020. Defining haptic experience: Foundations for understanding, communicating, and evaluating HX. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–13. DOI: <https://doi.org/10.1145/3313831.3376280>
- [68] Lawrence H. Kim and Sean Follmer. 2019. SwarmHaptics: Haptic display with swarm robots. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, 1–13. DOI: <https://doi.org/10.1145/3290605.3300918>
- [69] Kheng L. Koay, Dag S. Syrdal, Mohammadreza Ashgari-Oskoei, Michael L. Walters, and Kerstin Dautenhahn. 2014. Social roles and baseline proxemic preferences for a domestic service robot. *International Journal of Social Robotics* 6, 4 (Nov. 2014), 469–488. DOI: <https://doi.org/10.1007/s12369-014-0232-4>
- [70] Laura Kunold. 2022. Seeing is not feeling the touch from a robot. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN '22)*. IEEE, 1562–1569.
- [71] Matthijs Kwak, Kasper Hornbæk, Panos Markopoulos, and Miguel B. Alonso. 2014. The design space of shape-changing interfaces: A repertory grid study. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, 181–190. DOI: <https://doi.org/10.1145/2598510.2598573>
- [72] Mikko Lauri, David Hsu, and Joni Pajarinen. 2023. Partially observable Markov decision processes in robotics: A survey. *IEEE Transactions on Robotics* 39, 1 (Feb. 2023), 21–40. DOI: <https://doi.org/10.1109/TRO.2022.3200138>
- [73] Theresa Law, Bertram F. Malle, and Matthias Scheutz. 2021. A touching connection: How observing robotic touch can affect human trust in a robot. *International Journal of Social Robotics* 13, 8 (Dec. 2021), 2003–2019. DOI: <https://doi.org/10.1007/s12369-020-00729-7>
- [74] Bruno Leme, Masakazu Hirokawa, Hideki Kadone, and Kenji Suzuki. 2021. A socially assistive mobile platform for weight-support in gait training. *International Journal of Social Robotics* 13, 3 (Jun. 2021), 459–468. 1875–4805 <https://doi.org/10.1007/s12369-019-00550-x>

- [75] Richard L. Lewis, Andrew Howes, and Satinder Singh. 2014. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science* 6, 2 (Apr. 2014), 279–311. DOI: <https://doi.org/10.1111/tops.12086>
- [76] Jamy J. Li, Wendy Ju, and Byron Reeves. 2017. Touching a mechanical body: Tactile contact with body parts of a humanoid robot is physiologically arousing. *Journal of Human-Robot Interaction* 6, 3 (Dec. 2017), 118–130. DOI: <https://doi.org/10.5898/JHRI.6.3.Li>
- [77] Ying Siu Liang, Damien Pellier, Humbert Fiorino, and Sylvie Pesty. 2022. iRoPro: An interactive robot programming framework. *International Journal of Social Robotics* 14, (2022), 177–191.
- [78] Jessica Lindblom and Rebecca Andreasson. 2016. Current challenges for UX evaluation of human-robot interaction. In *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future (Advances in Intelligent Systems and Computing)*. C. Schlick and S. Trzcieliński (Eds.), Springer International Publishing, 267–277. DOI: https://doi.org/10.1007/978-3-319-41697-7_24
- [79] Marta Lopez Infante and Ville Kyrtti. 2011. Usability of force-based controllers in physical human-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '11)*. ACM, New York, NY, 355–362. DOI: <https://doi.org/10.1145/1957656.1957790>
- [80] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O’Malley. 2018. A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. *Applied Mechanics Reviews* 70, 1 (2018), 010804.
- [81] Paweł Maciejasz, Jörg Eschweiler, Kurt Gerlach-Hahn, Arne Jansen-Troy, and Steffen Leonhardt. 2014. A survey on robotic devices for upper limb rehabilitation. *Journal of Neuroengineering and Rehabilitation* 11 (Jan. 2014), 3. DOI: <https://doi.org/10.1186/1743-0003-11-3>
- [82] Ian. Scott MacKenzie. 2018. Fitts’ law. In *The Wiley Handbook of Human Computer Interaction*. Kent L. Norman and Jurek Kirakowski (Eds.), John Wiley & Sons, Ltd., 347–370. DOI: <https://doi.org/10.1002/9781118976005.ch17>
- [83] Karon E. MacLean, Oliver S. Schneider, and Hasti Seifi. 2017. Multisensory haptic interactions: Understanding the sense and designing for it. In Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, Antonio Krüger (Eds.) *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations-Volume 1*. 97–142.
- [84] Arlene Mannion, Sarah Summerville, Eva Barrett, Megan Burke, Adam Santorelli, Cheryl Kruschke, Heike Felzmann, Tanja Kovacic, Kathy Murphy, Dympna Casey, and Sally Whelan. 2020. Introducing the social robot MARIO to people living with dementia in long term residential care: Reflections. *International Journal of Social Robotics*. 12, 2 (May 2020), 535–547. DOI: <https://doi.org/10.1007/s12369-019-00568-1>
- [85] Sachiko Matsumoto, Auriel Washburn, and Laurel D. Riek. 2022. A framework to explore proximate human-robot coordination. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–34.
- [86] Conor McGinn and Dylan Dooley. 2020. What should robots feel like?. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. ACM, New York, NY, 281–288. DOI: <https://doi.org/10.1145/3319502.3374835>
- [87] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14, 4 (Dec. 1996), 261–292. DOI: <https://doi.org/10.1007/BF02686918>
- [88] Takashi Minato and Hiroshi Ishiguro. 2008. Construction and evaluation of a model of natural human motion based on motion diversity. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '08)*. ACM, New York, NY, 65–72. DOI: <https://doi.org/10.1145/1349822.1349832>
- [89] Thomas M. Moerland, Joost Broekens, and Catholijn M. Jonker. 2018. Emotion in reinforcement learning agents and robots: A survey. *Machine Learning* 107, 2 (Feb. 2018), 443–480. DOI: <https://doi.org/10.1007/s10994-017-5666-0>
- [90] Youssef Mohamed, Giulia Ballardini, Maria Teresa Parreira, Séverin Lemaignan, and Iolanda Leite. 2022. Automatic frustration detection using thermal imaging. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE, 451–459.
- [91] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. 2006a. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & Society* 20, 2 (Feb. 2006), 138–150. DOI: <https://doi.org/10.1007/s00146-005-0012-7>
- [92] Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kensuke Kato. 2008. Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics* 24, 2 (2008), 442–451. DOI: <https://doi.org/10.1109/TRO.2007.914004>
- [93] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006b. Measurement of negative attitudes toward robots. *Interaction Studies* 7, 3 (Jan. 2006), 437–454. DOI: <https://doi.org/10.1075/is.7.3.14nom>
- [94] Heather L. O’Brien and Elaine G. Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69. DOI: <https://doi.org/10.1002/asi.21229>

- [95] Valerio Ortenzi, Akansel Cosgun, Tommaso Pardi, Wesley P. Chan, Elizabeth Croft, and Dana Kulic. 2021. Object handovers: A review for robotics. *IEEE Transactions on Robotics* 37, 6 (Dec. 2021), 1855–1873. DOI: <https://doi.org/10.1109/TRO.2021.3075365>
- [96] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational rationality as a theory of interaction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, 1–14. DOI: <https://doi.org/10.1145/3491102.3517739>
- [97] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372 (2021). DOI: <https://doi.org/10.1136/bmj.n71>. Retrieved from <https://www.bmjjournals.org/content/372/bmj.n71.full.pdf>
- [98] Matthew K. X. J. Pan, Elizabeth A. Croft, and Günter Niemeyer. 2018. Evaluating social perception of human-to-robot handovers using the robot social attributes scale (RoSAS). In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, 443–451. DOI: <https://doi.org/10.1145/3171221.3171257>
- [99] Daehyung Park, Yuuna Hoshi, and Charles C. Kemp. 2018. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. *IEEE Robotics and Automation Letters* 3, 3 (Jul. 2018), 1544–1551. DOI: <https://doi.org/10.1109/LRA.2018.2801475>
- [100] Akanksha Prakash, Jenay M. Beer, Travis Deyle, Cory-Ann Smarr, Tiffany L. Chen, Tracy L. Mitzner, Charles C. Kemp, and Wendy A. Rogers. 2013. Older adults' medication management in the home: How can robots help? In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '13)*. IEEE Press, 283–290. DOI: <https://doi.org/10.1109/HRI.2013.6483600>
- [101] Vinita Ranganeni, Mike Sinclair, Eyal Ofek, Amos Miller, Jonathan Campbell, Andrey Kolobov, and Edward Cutrell. 2023. Exploring levels of control for a navigation assistant for blind travelers. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. 4–12.
- [102] Nur L. A. Rashid, Leow Yihong, Piyanee Klainin-Yobas, Itoh Sakiko, and Wu V. Xi. 2023. The effectiveness of a therapeutic robot, 'Paro', on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults with dementia: A systematic review and meta-analysis. *International Journal of Nursing Studies* 145 (2023), 104530.
- [103] Byron Reeves, Jeff Hancock, and Xun Sunny Liu. 2020. Social robots are like real people: First impressions, attributes, and stereotyping of social robots. *Technology, Mind, and Behavior*, 1, (1) (2020). DOI: <https://doi.org/10.1037/tmb0000018>
- [104] Laurel D. Riek. 2012. Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (July 2012), 119–136. DOI: <https://doi.org/10.5898/JHRI.1.1.Riek>
- [105] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 1161–1178. DOI: <https://doi.org/10.1037/h0077714>
- [106] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. 2021. Measuring Haptic Experience: Elaborating the HX model with scale development. In *Proceedings of the IEEE World Haptics Conference (WHC '21)*. 979–984. DOI: <https://doi.org/10.1109/WHC49131.2021.9517220>
- [107] Samuel B. Schorr, Zhan Fan Quek, William R. Provancher, and Allison M. Okamura. 2015. Environment perception in the presence of kinesthetic or tactile guidance virtual fixtures. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, 287–294. DOI: <https://doi.org/10.1145/2696454.2696488>
- [108] Hasti Seifi, Steven A. Vasquez, Hyunyoung Kim, and Pooyan Fazli. 2023. First-hand impressions: Charting and predicting user impressions of robot hands. *ACM Transactions on Human-Robot Interaction* 12, 3 (Apr. 2023), 351–35:25. DOI: <https://doi.org/10.1145/3580592>
- [109] Masahiro Shiomi and Norihiro Hagita. 2021. Audio-visual stimuli change not only robot's hug impressions but also its stress-buffering effects. *International Journal of Social Robotics* 13, 3 (Jun. 2021), 469–476. DOI: <https://doi.org/10.1007/s12369-019-00530-1>
- [110] Masahiro Shiomi, Kayako Nakagawa, Kazuhiko Shinozawa, Reo Matsumura, Hiroshi Ishiguro, and Norihiro Hagita. 2017. Does a robot's touch encourage human effort? *International Journal of Social Robotics* 9, 1 (Jan. 2017), 5–15. DOI: <https://doi.org/10.1007/s12369-016-0339-x>
- [111] Masahiro Shiomi, Aya Nakata, Masayuki Kanbara, and Norihiro Hagita. 2021. Robot reciprocation of hugs increases both interacting times and self-disclosures. *International Journal of Social Robotics* 13, 2 (April 2021), 353–361. DOI: <https://doi.org/10.1007/s12369-020-00644-x>
- [112] Michael Sugitan and Guy Hoffman. 2019. Blossom: A handcrafted open-source robot. *Journal of Human-Robot Interaction* 8, 1 (Mar. 2019), 27 pages. DOI: <https://doi.org/10.1145/3310356>

- [113] Katherine Tsui, Holly Yanco, David Kontak, and Linda Beliveau. 2008. Development and evaluation of a flexible interface for a wheelchair mounted robotic arm. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '08)*. ACM, New York, NY, 105–112. DOI: <https://doi.org/10.1145/1349822.1349837>
- [114] Yasemin Vardar, Christian Wallraven, and Katherine J. Kuchenbecker. 2019. Fingertip interaction metrics correlate with visual and haptic perception of real surfaces. In *Proceedings of the IEEE World Haptics Conference (WHC '19)*. 395–400.
- [115] Amber M. Walker, David P. Miller, and Chen Ling. 2015. User-centered design of an attitude-aware controller for ground reconnaissance robots. *Journal of Human-Robot Interaction* 4, 1 (Jul. 2015), 30–59. DOI: <https://doi.org/10.5898/JHRI.4.1.Walker>
- [116] Dangxiao Wang, Yuan Guo, Shiyi Liu, Yuru Zhang, Weiliang Xu, and Jing Xiao. 2019. Haptic display for virtual reality: Progress and challenges. *Virtual Reality & Intelligent Hardware* 1, 2 (Apr. 2019), 136–162. DOI: <https://doi.org/10.3724/SP.J.2096-5796.2019.0008>
- [117] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54 (Jun. 1988), 1063–1070. DOI: <https://doi.org/10.1037/0022-3514.54.6.1063>
- [118] Alexandra Weidemann and Nele Rußwinkel. 2021. The role of frustration in human–robot interaction—What is needed for a successful collaboration? *Frontiers in Psychology* 12 (2021), 707.
- [119] Astrid Weiss, Regina Bernhaupt, and Manfred Tscheligi. 2011. The USUS evaluation framework for user-centered HRI. *New Frontiers in Human–Robot Interaction* 2 (Dec. 2011), 89–110. DOI: <https://doi.org/10.1075/ais.2.07wei>
- [120] Christian J. A. M. Willemse, Gijs Huisman, Merel M. Jung, Jan B. F. van Erp, and Dirk K. J. Heylen. 2016. Observing touch from video: The influence of social cues on pleasantness perceptions. In *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications (EuroHaptics '16)*. 196–205.
- [121] Christian J. A. M. Willemse and Jan B. F. van Erp. 2019. Social touch in human–robot interaction: Robot-initiated touches can induce positive responses without extensive prior bonding. *International Journal of Social Robotics* 11, 2 (Apr. 2019), 285–304. DOI: <https://doi.org/10.1007/s12369-018-0500-9>
- [122] Nialah J. Wilson-Small, David Goedcke, Kirstin Petersen, and Shiri Azenkot. 2023. A drone teacher: Designing physical human-drone interactions for movement instruction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. 311–320.
- [123] Sebastian Wrede, Christian Emmerich, Ricarda Grünberg, Arne Nordmann, Agnes Swadzba, and Jochen Steil. 2013. A user study on kinesthetic teaching of redundant robots in task and configuration space. *Journal of Human-Robot Interaction* 2, 1 (Mar. 2013), 56–81. DOI: <https://doi.org/10.5898/JHRI.2.1.Wrede>
- [124] Yuki Yamashita, Hisashi Ishihara, Takashi Ikeda, and Minoru Asada. 2019. Investigation of causal relationship between touch sensations of robots and personality impressions by path analysis. *International Journal of Social Robotics* 11, 1 (Jan. 2019), 141–150. DOI: <https://doi.org/10.1007/s12369-018-0483-6>
- [125] Steve Yohanan and Karon E. MacLean. 2011. Design and assessment of the haptic creature's affect display. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '11)*. ACM, New York, NY, 473–480. DOI: <https://doi.org/10.1145/1957656.1957820>
- [126] Steve Yohanan and Karon E. MacLean. 2012. The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *International Journal of Social Robotics* 4, 2 (Apr. 2012), 163–180. DOI: <https://doi.org/10.1007/s12369-011-0126-7>
- [127] Yongjae Yoo, Taekbeom Yoo, Jihyun Kong, and Seungmoon Choi. 2015. Emotional responses of tactile icons: Effects of amplitude, frequency, duration, and envelope. In *Proceedings of the IEEE World Haptics Conference (WHC '15)*. 235–240. DOI: <https://doi.org/10.1109/WHC.2015.7177719>
- [128] Naoya Yoshimura, Yushi Sato, Yuta Kageyama, Jun Murao, Satoshi Yagi, and Parinya Punpongsanon. 2022. Hugmon: Exploration of affective movements for hug interaction using tensegrity robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE, 1105–1109.
- [129] James E. Young, JaYoung Sung, Amy Voda, Ehud Sharlin, Takeo Igarashi, Henrik I. Christensen, and Rebecca E. Grinter. 2011. Evaluating human-robot interaction. *International Journal of Social Robotics* 3, 1 (Jan. 2011), 53–67. DOI: <https://doi.org/10.1007/s12369-010-0081-8>
- [130] Yuefang Zhou, Tristan Kornher, Janett Mohnke, and Martin H. Fischer. 2021. Tactile interaction with a humanoid robot: Effects on physiology and subjective impressions. *International Journal of Social Robotics* 13, 7 (Nov. 2021), 1657–1677. DOI: <https://doi.org/10.1007/s12369-021-00749-x>
- [131] Ayberk Özgür, Séverin Lemaignan, Wafa Johal, Maria Beltran, Manon Briad, Léa Pereyre, Francesco Mondada, and Pierre Dillenbourg. 2017. Cellulo: Versatile handheld robots for education. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 119–127. DOI: <https://doi.org/10.1145/2909824.3020247>

Received 11 May 2023; revised 10 November 2023; accepted 7 March 2024