A Three-Site Reproduction of the Joint Simon Effect with the NAO Robot

Megan Strait
The University of Texas Rio Grande Valley, Edinburg, TX
megan.strait@utrgv.edu

Florian Lier[1,2]
[1]Corporate Research, Robert Bosch GmbH, Germany
[2]Bielefeld University, Bielefeld, Germany
florian.lier@de.bosch.com

Jasmin Bernotat, Sven Wachsmuth, Friederike Eyssel
Bielefeld University, Bielefeld, Germany
{jbernotat,swachsmu,feyssel}@techfak.uni-bielefeld.de

Robert Goldstone, Selma Šabanović
Indiana University, Bloomington, IN
{rgoldsto,selmas}@indiana.edu

ABSTRACT

The generalizability of empirical research depends on the reproduction of findings across settings and populations. Consequently, generalizations demand resources beyond that which is typically available to any one laboratory. With collective interest in the *joint Simon effect* (JSE) – a phenomenon that suggests people work more effectively with humanlike (as opposed to mechanomorphic) robots – we pursued a multi-institutional research cooperation between robotics researchers, social scientists, and software engineers. To evaluate the robustness of the JSE in dyadic human-robot interactions, we constructed an experimental infrastructure for exact, lab-independent reproduction of robot behavior. Deployment of our infrastructure across three institutions with distinct research orientations (well-resourced versus resource-constrained) provides initial demonstration of the success of our approach and the degree to which it can alleviate technical barriers to HRI reproducibility. Moreover, with the three deployments situated in culturally distinct contexts (Germany, the U.S. Midwest, and the Mexico-U.S. Border), observation of a JSE at each site provides evidence its generalizability across settings and populations.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Human-robot interaction, joint Simon effect, reproducibility

ACM Reference Format:
Megan Strait, Florian Lier[1,2], Jasmin Bernotat, Sven Wachsmuth, Friederike Eyssel, and Robert Goldstone, Selma Šabanović. 2020. A Three-Site Reproduction of the Joint Simon Effect with the NAO Robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3319502.3374783

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.
HRI '20, March 23–26, 2020, Cambridge, United Kingdom
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374783

1 INTRODUCTION

Failures to reproduce well-accepted phenomena, particularly within the psychological sciences (e.g., [7, 14, 20]), have prompted widespread and ongoing concern regarding knowledge derived from empirical research (e.g., [2, 11, 25, 26]). Because reproduction is necessary to generalize beyond singular studies (e.g., across settings and populations distinct from that of an originating investigation), the dearth of such work has further amplified concern and called into question the reliability of extant literature (e.g., [1, 21, 24]).

Reproductions, however, are *resource-intensive.* For example, merely repeating an original experiment requires duplicate effort and resources including time, participants, and materials to execute. Moreover, as stand-alone work, reproductions garner less recognition and face greater difficulty in publishing (e.g., [16, 17, 31]).

Further challenges arise in reproducing human-robot interaction research. For example, reproduction of robot behavior relies on the accessibility of robotics hardware and software systems. A recent investigation of *code reusability* found, however, that only 16 of the 800 papers published in the 2017 Proceedings of the IEEE International Conference on Robotics and Automation referenced source code that could be successfully built and run [8]. Further highlighting the unavailability of even contemporary software systems, just one of the 16 papers referenced source code that could be run off-the-shelf; the other 15 required extensive, non-automateable work to comprehend, supplement, and execute their code.

But with applications in areas such as education [4], eldercare [6], and healthcare [34], the reproduction of HRI research is especially critical. In particular, robotics holds great promise for providing societal benefits and addressing concerns that, through the types of services robots are expected to offer, implicate a broad range of stakeholders including vulnerable users such as children and older adults (e.g., [33]). Recent work indicates, however, that people's perceptions of robot designs and uses can vary significantly between users (e.g., [27]). For example, relative to standard convenience samples (i.e., young adults), children and older adults both exhibit distinct responses to robots (e.g., [28, 32, 40]). To develop robotic technologies that are effective for all thus requires careful consideration of the *generalizability* of findings across populations.

Present Work

With collective interest in working toward generalizable theories of human-robot interaction and specific interest in the design implications of the **joint Simon effect** (JSE), we pursued a multi-institution research cooperation between roboticists, social scientists, and software engineers in order to investigate the reproducibility of the JSE in dyadic interactions with the NAO robot. Our primary contributions are as follows:

(1) Practical **application of a semi-automated approach to develop, deploy, and conduct human-robot interaction experiments** based on an existing software tool chain for reproducible robotics experiments ([30]). This includes setting up and demoing transferability of a site-independent experimental architecture, the robot's behavior and coordination, as well as stimuli presentation and data acquisition.
(2) Execution of **three conceptual reproductions of the experiment by Stenzel et al. (2012)**, which suggests that human-like robots are better suited for human-robot teaming than are mechanomorphic robots. The reproductions provide initial demonstration of the *efficacy of our technical approach* in enabling HRI reproducibility and evidence of the *robustness of the joint Simon effect* to situational and cultural variability.

Overall, we aimed to address **multiple levels of depth of both technical and experimental reproducibility** via a multi-site deployment of our infrastructure across two distinct institutional profiles (*well-resourced* vs. *resource-constrained*) and three distinct cultural contexts (*Germany*, the *U.S. Midwest*, and the *Mexico-U.S. Border*). In total, the work comprises three conceptual reproductions of [38] and three validations of our technical approach via *internal deployment* of our infrastructure at Bielefeld University, *external deployment* at Indiana University (a university with a comparable institutional profile to that of Bielefeld University), and *external deployment* at the University of Texas Rio Grande Valley (a university with limited institutional support for robotics research).

2 JOINT SIMON EFFECT

The *Simon effect* refers to the phenomenon wherein spatial congruence between a stimulus and the response facilitates task performance [37]. For example, in a visual formulation of the corresponding task paradigm (the Simon task), a person is presented with one of two stimuli (e.g., square and diamond) on either the left or right of a display. Tasked with manually responding to the stimuli *in a spatially defined manner* (e.g., press the left button if the stimulus is a square and the right button if the stimulus is a diamond), faster response times are observed when the stimulus and response are spatially congruent (e.g., a square, cuing a **left** press, is presented on the display's **left**) than when they are spatially incongruent (e.g., a square, cuing a **left** press, is presented on the display's **right**).

If the Simon task is modified such that the spatiality of responses is eliminated by making a person responsible for responding to just one of the two stimuli (i.e., an *individual Go/No-Go Simon task*), the Simon effect disappears [22]. Yet, if the Simon task is instead distributed between two people (i.e., a *joint Go/No-Go Simon task*), people are faster to respond to stimuli aligned with themselves than to stimuli aligned with a co-actor despite being responsible for just one of the two stimuli – reflecting a facilitatory influence of spatial congruence referred to as the **joint Simon effect** ([13, 35]).

Figure 1: Joint Simon Task (from left to right): Standard (human-human) task setup; setup of the present investigation, involving Softbank Robotics' *NAO*; and setup of the original HRI study by Anna Stenzel and colleagues ([38]), involving the in-house robot *Tombatossals* [10].

2.1 Action Co-Representation

In both individual and joint versions of the Go/No-Go Simon task, a person's task is identical: operate just one button and attend to just one of the two stimuli. For example, in the *individual* version, if a person (A) is tasked with responding to squares they respond when a square is shown (*Go*) and refrain when a diamond is shown (*No-Go*). In the *joint* version, a second person (B) is tasked with responding to the other stimulus (e.g., diamonds). The presence of B, however, does not change A's task: A goes when a square is shown, while B refrains, and when a diamond is shown A refrains while B goes. Observation of a Simon effect in the joint task has thus been interpreted as implicative of **action co-representation** (i.e., both co-actors cognitively represent the actions of the other as if each person is in charge of the full, undivided task [35, 36]). Per this account, representation of the actions of the co-actor reintroduces the spatiality of responding and stimulus-response correspondence (where stimuli aligned with oneself are spatially congruent and stimuli aligned with the co-actor are spatially incongruent), despite individuals controlling just one response button.

2.2 Implications for HRI

Subsequent exploration of the joint Simon effect wherein the ontology of one of the co-actors was manipulated (as human/humanlike or not) initially suggested a restriction of the phenomenon to conspecifics (i.e., cognitive co-representation occurs only in joint action with other people). For example, replacement of one of the co-actors with a humanlike versus non-human hand was found to produce a JSE only when the task was completed alongside the humanlike hand [41]. Similarly, completion of the task with an *unseen co-actor* – described as either a person or a computer – was found to yield a Simon effect only with the "human" co-actor, despite the fact that the unseen co-actor was a computer in both conditions [42].

Preliminary investigation of the joint Simon effect in human-robot dyadic interactions, however, implied action co-representation to extend at least to humanoid robots [38]. Specifically, despite replacement of one of the two co-actors with the humanoid robot,

Tombatossals ([10]; see Figure 1: right), a JSE manifested. Additional manipulation therein of the cognitive framing of the robot (by verbally describing Tombatossals as more or less agentic) resulted in increased attributions of intentionality to the robot's behavior and amplification of the joint Simon effect [38].

These findings suggest that human-robot teaming may be facilitated by framing robot co-actors as more agentic. Given that humanlike designs foster attribution of human traits such as agency (e.g., [19]), they also imply that humanlike robots are better suited for collaborative tasks with people than are mechanomorphic robots. But absent reproduction, the findings of [38] remain preliminary with their robustness and generalizability unconfirmed.

3 METHOD

To investigate the robustness of the joint Simon effect (JSE) in human-robot dyads, three experiments were designed to reproduce the method and procedure of the originating study ([38]) as exactly as possible.[1] Based on the findings of [38], we had two predictions:

- H_1: Assuming the joint Simon effect is robust to situational and cultural variability (i.e., generalizes across settings and populations), we expected a significant main effect of *stimulus-response compatibility*, wherein participants respond faster to spatially congruent versus incongruent stimuli, across the three sites.
- H_2: Given the additional manipulation of the robot's *agency* in [38], which showed that framing the robot as more agentic produced a stronger JSE, we also examined the association between participants' attributions of agency to the NAO and the difference in response time to spatially incongruent and congruent stimuli (Δ_{RT}), anticipating a positive relationship between the two (i.e., greater agency attribution corresponds to a larger Δ_{RT}).

3.1 Sites

Three research teams (one per data acquisition site) participated in the conceptualization, implementation, and execution of the experiments. The specific institutional composition – Bielefeld University (Site 1), Indiana University (Site 2), and the University of Texas Rio Grande Valley (Site 3) – enabled investigation of technical and empirical robustness to variation across multiple factors of potential relevance (*cultural affiliation*, *research orientation*, and *robotics exposure*). While all three institutions are relatively large, public universities, each is situated in a distinct cultural context: *Germany* (Bielefeld University), the *United States Midwest* (Indiana University), and the *Mexico-U.S. Border* (the University of Texas Rio Grande Valley). The three institutions also span distinct research orientations, where Sites 1–2 (both doctoral degree-granting institutions with high levels of research activity) are relatively *well-resourced* and *research-oriented* in comparison to Site 3, which is a relatively *resource-constrained*, primarily undergraduate institution. Moreover, assessment of participants' prior experience with robotic technologies reflected greater exposure at Site 1; whereas direct exposure at Sites 2–3 remains relatively limited to date (see Table 1).

[1]Because of the unavailability of the original robot (Tombatossals), we instead used the NAO robot which was available at all three cooperating sites and offered greater comparability to existing *and* future HRI literature due to its accessibility and broader adoption. In addition, unlike [38] (wherein participants' position relative to the robot was fixed), we manipulated participants' positioning (between-subjects) so as to minimize its influence as a potential confound in the data.

Table 1: Participant demographics (sample size, mean age, and the proportion of participants who recognized the NAO and had prior *direct* exposure/s to robots) by site.

	N	$M_{age} \pm 1\ SD$	*NAO Familiarity*	*Preexposure*
Site 1	47	24.61 ± 4.01	45%	47%
Site 2	51	19.51 ± 2.39	4%	2%
Site 3	72	21.72 ± 4.41	29%	1%

3.2 Procedure

Upon arrival to the given laboratory, prospective participants were greeted by an on-site researcher and read the Site's consent form. After providing informed consent, the on-site researcher led the participant to the testing room (where the NAO and task were already set up and calibrated). In the testing room, participants were seated next to the NAO (either to the NAO's left or right depending on the session's physical setup; see Figure 1) and the NAO was introduced to the participant as their task partner with no further description or detail. Participants were informed that the monitor would display instructions as to how to proceed through the task, followed by the task itself, and once the participant was comfortable and ready to begin, the researcher triggered the start of the instructions and left the room to give the participant privacy.

3.2.1 Setup. The testing station was equipped with a table, monitor, computer, two side-by-side chairs (one of which served to seat the NAO and the other remained available for participants), and two keyboards (one for the robot's responses and the other for that of the participant; see Figure 1).

To enable the NAO to consistently respond to *Go* trials, it was programmed to move its left hand (if seated to the participant's right) or right hand (if seated to the participant's left) at the start of the task to hover above the keyboard. From there on, whenever the NAO received a triggering signal, it extended its arm down to depress the space bar with its fingers and then back up to the hover position. The NAO was otherwise kept still.

Like [38], we additionally programmed the robot to (1) randomly err (e.g., go on a no-go stimulus) in 1.6% of the trials and (2) mirror the variability in human response times with the NAO programmed to achieve a mean response time of $400ms$ and standard deviation of $60ms$ across responses (range: $300ms - 499ms$). Overall, this well approximated the observed response behavior of our participants ($M = 397ms$; $SD = 48ms$; range: $311ms - 723ms$).

3.2.2 Task. Following the instructions and a set of practice trials, the joint Simon task was run wherein participants were instructed to respond by pressing the spacebar of their keyboard as fast and as accurately as possible whenever a square was presented. The NAO was responsible for and programmed to respond to the stimuli that were diamonds (squares rotated 90°) by pressing the spacebar of its keyboard using its left hand (when seated on the right) or right hand (when seated on the left). The task contained 512 trials total (256 *Go* trials per agent), with breaks after every 128 trials.

3.2.3 Trial Structure. As per [38], each trial began with a fixation point for $250ms$ followed by the stimulus presentation ($250ms$).

Both the stimulus *type* (square or diamond) and spatial positioning (centered on either the left or right half of the screen) were randomly interleaved. Correct responses were defined as responses to *Go* trials given within 1800*ms* of the stimulus presentation. After either a response was provided or the response window expired, correctness was depicted via the fixation point which either turned green, indicating correctness, or red, indicating an error, for 300*ms*. A rest period of 1850*ms* then followed to separate each trial.

3.2.4 Post-Task. At the task end, participants were instructed to inform the researcher (waiting in an adjacent room). The researcher then provided access to, as well as privacy again while completing, a post-task questionnaire. At the end of the participation, participants received 8€ (Site 1) or course credit (Sites 2–3).

3.3 Software Setup

With exact reproduction of robot behavior and lab-independence as primary needs in the experimental reproductions of [38], we imposed the following requirements on the development of our technical approach: *a)* the physical setup, software deployment and experiment execution must be able to be carried out solely based on instructions provided by online documentation – similar to supplemental material of a regular publication, and *b)* the entire setup must be deployable by scientists within the multidisciplinary scope of HRI including (e.g., non-robotics researchers).

To confirm the efficacy of our technical approach in enabling independent reproductions of [38], we produced a document to serve as the primary resource in the three deployments (see Section 7). The document includes, in addition to the hardware requirements, the recommended network topology, instructions regarding the physical setup, instructions on deploying our software toolchain for reproducible robotics experiments (see Section 3.3.2) and setting up and executing the experiment on a single machine (e.g., a laptop).

3.3.1 Robot Control & Stimuli Presentation. The experiment's software stack consisted of stand-alone components to calibrate and control the NAO robot, as well as to display visual stimuli. The NAO network setup, calibration, and motion control were realized by implementing Softbank's off-the-shelf NAOqi 2.4.3 API[2] with the intent to build on a publicly available library. In particular, we implemented a calibration GUI including guided step-by-step instructions for the NAO's left and right arm with regard to its positioning relative to the keyboard. Furthermore, a NAOqi web bridge was realized providing a REST interface to the *libqi* motion API. Crucial parameters, such as joint stiffness and velocity were hardcoded in the control component.

In order to display the visual stimuli and record participant responses, we used *jsPsych* – a JavaScript framework for creating web-based behavioral experiments [12]. Functionality within jsPsych includes, for example, trial ordering, recording data, and exporting data via direct browser download to CSV files. We further augmented jsPsych using its plug-in architecture to trigger the NAO's movement (i.e., a key press during a participant's *No-Go* trial in the context of the joint Simon task) from the jsPsych timeline by invoking a REST call to the web bridge.

[2]http://doc.aldebaran.com/2-4/naoqi/index.html

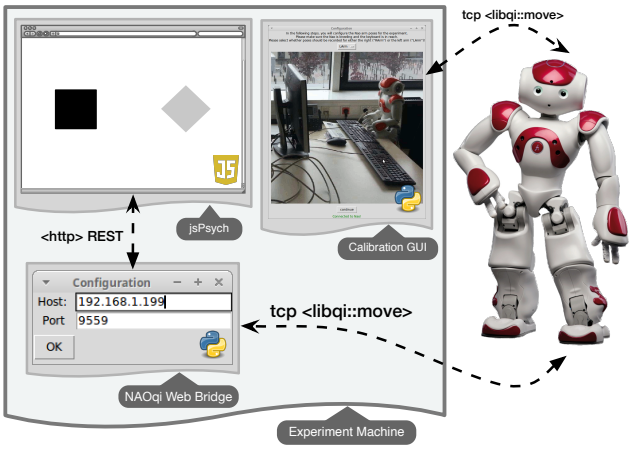

Figure 2: Simplified overview of the NAO-JSE setup.

3.3.2 Reproducible Deployment & Execution. With the aim of deploying and conducting the joint Simon experiment identically at all three sites, we used the *Cognitive Interaction Toolkit* (CITK; [30]). The CITK, which encompasses a software toolchain embedded in an iterative development process, is explicitly designed to facilitate reproducibility of software-intensive robotics and human-robot interaction experiments (see, for example, [29, 30]).

The core of the CITK is a file-based and machine-interpretable software system composition approach to alleviate systematic aggregation and deployment of experiment-related *artifacts*, such as software components (i.e., programs), datasets, documentation, configuration/parameter files, and executable experiment workflows. In the CITK ecosystem, these artifacts are represented by text files (YAML), so-called *recipes*. For example, a recipe constituting a software component merely includes a reference to the corresponding source code repository (see Listing 1, line 5), available versions (lines 6-8), and the nature of the project in the form of a corresponding template (e.g, *cmake-catkin*; lines 1-3).

Thus, recipes represent distinct research artifacts and their properties such as type and nature. Templates, in contrast, implement *how* to process the associated recipe (e.g, how to download, compile, install, or patch it). For instance: a software component recipe includes the cmake-catkin template. The template, in turn, implements the necessary information how to compile, test, and install software that makes use of the catkin build system *in general*.

```
1  templates:
2  - cmake-catkin
3  - base
4  variables:
5    repository: https://opensource.cit-ec.de
6    branches:
7    - master
8    - v1.0
```

Listing 1: Example of a CITK recipe's YAML file. The approach follows minimalist standards which serve to reducing the effort needed to integrate existing software.

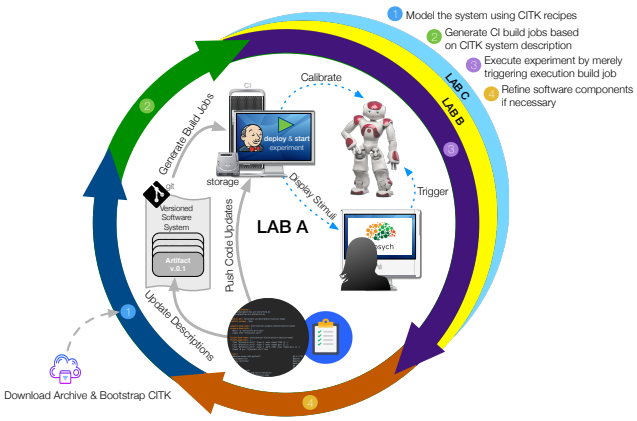

Figure 3: CITK development process: from artifact aggregation ❶ to deployment ❷, execution ❸ and refinement ❹

Application of a template-based approach introduces the following benefits with respect to reproducible software: *(i)* templates are usually written by experts. Hence, in the above case for instance, a researcher does not need to become an expert in software engineering to make use of/integrate catkin-built software. Consequently: *(ii)* artifact-specific best practices are transparently implemented and applied by default; *(iii)* error prone redundancy (e.g, for artifact configuration) is mitigated; *(iv)* templates are highly reusable and can be added, extended, replaced, and combined immediately; and *(v)* recipe code, usually written by researchers, is minimal which reduces effort and provides a lower level to entry.

Moreover, recipes of diverse type are usually added to a composite representation, called *distribution*. Like recipes, distributions are YAML files; they represent a software system and must comprise all artifacts that are **required** to execute an associated experiment. More precisely, all recipes incorporated in a distribution file must reference a **version**, i.e, `id@version` (see: figure 2). Finally, recipes, templates, and distributions are stored in a public git repository which permits versioning of the aforementioned files, as well as public access, participation, and contribution (see Section 7).

3.3.3 JSE-NAO Software Stack. Initially, all components (see Section 3.3.1) were integrated into the CITK by creating one recipe per artifact and a corresponding distribution file (see Figure 3, step ❶).

In the subsequent step (❷), we utilized two additional tools bundled with the CITK (which is downloaded in the form of a single archive). The *build job generator* (a) was applied to automatically add build jobs to a *Continuous Integration (CI) server* (b; see [15] for further details). In particular, the build job generator analyzes all recipes, and thus, all associated templates of the target distribution. The following information is assembled per artifact: *(i)* how the artifact must be deployed; *(ii)* what version of the artifact is requested; and *(iii)* what are its dependencies (e.g., location in the calculated build job dependency tree and required OS packages). Subsequently, the generator applies this information to automatically instantiate and configure one job per artifact.

```
[...]
versions:
[...]
- aldebaran-naoqi-sdk-python27@2.4.3.28
- jsp-nao-calibrate@master
- jspsych@master
- runnable-remotelab-jsp-nao-calibration@master
```

Listing 2: Abbreviated CITK distribution file. It references multiple research artifact versions that constitute the software system aspect of the JSE-NAO experiment.

In the CITK ecosystem, the CI server acts as an easy-to-use interface for *interdisciplinary* researchers since it provides a web-based GUI where all jobs, applicable actions (e.g., run/stop job), and statuses (e.g., success, failed, aborted) are itemized (see Figure 4). Specifically, expert knowledge (e.g., of software compilation and deployment) is abstracted away in the build jobs' configuration. Jobs are readily triggered by activating them in the GUI and since an experiment may comprise tens, or even hundreds, of distinct artifacts, a special orchestration and experiment execution job are created. Once the orchestration job is activated, all associated jobs are triggered in the **correct order** (i.e., adhering the dependency graph) to ensure accurate deployment of the target software system.

In order to run the experiment, a researcher activates the *execution job* in step ❸ which, in turn, invokes an experiment workflow. The experiment workflow implements a state machine-based approach to configure and start all required software programs that comprise the software-centric aspects of the experiment. It handles process spawning, error reporting, and safe shutdown of incorporated programs. In the development phase, this step is used to test and refine the software components driving the experimental procedure. Therefore, if necessary, changes to the source code can be pushed directly to the origin repository and immediately applied (i.e., redeployment using the CI server) in step ❹. Moreover, if required, new recipes can be added by iterating on steps ❶+❷.

3.4 Participants

In recruiting participants, effort was made to match the sample size of each of the three sites to that of the originating study (N = 48; [38]). To confirm this would be sufficiently sensitive to the magnitude of the joint Simon effect, we ran a high-powered power analysis ($1-\beta = .95$) assuming an effect size directly computed from that reported in [38] ($\eta_p^2 = .31 \implies f = .67$) and with $\alpha_{crit} = .05$. Calculated in g*Power [18], with F tests as the test family and two conditions in the experimental design (responses to spatially

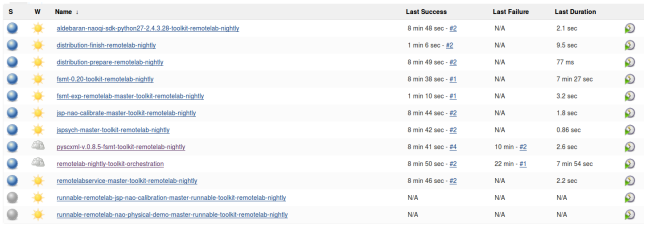

Figure 4: CI server front-end. From left to right: status and trend, name, last runs, and execution. Calibration and the experiment are started by triggering the last two items.

congruent vs. *incongruent* stimuli), the analysis produced a lower bound of 32 thus confirming $N = 48$ to be sufficiently powered.

Participants were recruited via convenience sampling at each site (i.e., advertisement to the corresponding student body).[3] In total, 170 students participated across Sites 1-3 ($N_1 = 47$, $N_2 = 51$, and $N_3 = 72$). Ages ranged from 18 to 41 years old ($M_{age} = 21.84$, $SD = 4.24$) and gender identification reflected relatively equal participation by men and women ($N_{men} = 74$, $N_{women} = 81$, $N_{nonbinary} = 1$, and $N_{NR} = 14$). Participants' cultural identification mirrored that of the institutions' geographic situation, with consistent monocultural identification across Sites 1–2 (87% of Site 1 participants identified as German and 82% of Site 2 participants identified as American[4]) and bicultural Mexican-American identification at Site 3 (42% of participants identified as American, 29% as bicultural Mexican-American, and 18% as Mexican). Except for Site 3, wherein participants largely identified as Hispanic (80%) and people of color (87%), ethnic and racial identification were not expressly recorded for Sites 1–2. It is assumed, however, that the Site 1–2 demographics reflected those of the corresponding institutions (i.e., predominately non-Hispanic and White).

3.5 Measures

During the joint Simon task, we recorded participants' **response time** (RT) – defined as the time it took to respond since the start of the stimulus presentation on each *Go* trial (i.e., trials wherein participants were prompted to respond). Correspondingly, only *Go* trials – excluding those wherein participants failed to respond – were retained for analysis. The retained RT data were then preprocessed and analyzed exactly in accordance with the procedures described by Stenzel and colleagues ([38]). For each participant:

- We calculated the *median* RT for trials wherein the spatial positioning of the stimulus was *congruent* to the participant's position and for trials wherein the stimulus was spatially *incongruent*.
- A difference measure (Δ_{RT}) was computed as the difference between these two medians (i.e., $RT_{incongruous} - RT_{congruent}$).

Following the task, we recorded participants' responses to a questionnaire containing standard demographic items (e.g., age, gender, cultural identification) and two Likert-type items designed by Stenzel and colleagues [38] to measure participants' attribution of *intentionality* to the NAO ("*the robot acted intentionally*" and "*the robot decided actively when to respond to a stimulus*"). Responses to the items were solicited with a 5-point scale (from "disagree" to "agree") and averaged to produce a composite score (of *intentionality* attributed to the NAO). This measure served in testing H_2 (whether the JSE is associated with intentionality ascribed to the co-actor).[5]

[3]Study sessions were held until: the original N (48) was *thought* to be met (Site 1; $N = 47$): the original N was met, *excluding* 3 sessions in which recording errors occurred (Site 2; $N = 51$); and all interested participants had been run (Site 3; $N = 72$).
[4]Although "American" encompasses all American identities (e.g., Canadian, Mexican, United Statesian), given its widespread colloquial usage to imply "United Statesian" (cultural identification of being of/from the US), we have similarly adopted it here.
[5]Given the opportunity in data collection created by the three-site endeavor, several additional constructs were included within the questionnaire to measure participants' experiences in the study toward subsequent exploration of normative values, internal consistencies, and implications for HRI. The full set of items, sources, and clustering by constructs are reported in Appendix A. As these items were exploratory and disjoint from the present work, however, they were not included in the reporting here.

4 RESULTS

We analyzed the data, by site, in relation to our two predictions derived from the findings in [38]. Assuming the JSE is robust to situational and cultural variation, we expected participants' response times to spatially congruent stimuli to be significantly faster than those to spatially incongruent stimuli (H_1). Moreover, assuming the JSE is tuned to *social others* such that greater agency facilitates action co-representation thus amplifying the JSE, we expected to find a significant, positive correlation between participants' Δ_{RT} and their ratings of the NAO's intentionality (H_2). All analyses and results are publicly available and reproducible via CodeOcean.

4.1 Joint Simon Effect

During the initial deployment of our system infrastructure (internally at Site 1) and at the start of the first external deployment (Site 2), errors arose in recording data such that response time data from 10 participants ($N_1 = 7$ and $N_2 = 3$) were not saved. In addition, as per [38], participants with a Δ_{RT} that fell ±2.5 standard deviations from the group mean were excluded from further analysis. This resulted in the exclusion of one participant from Site 1 and one participant from Site 2. Thus the final, adjusted sample sizes of Sites 1–3 in hypothesis testing were: $N = 39$, 47, and 72 respectively.

To evaluate the effect of **spatial congruence** between a stimulus and a participant's position, we conducted an analysis of variance (ANOVA) on response time data. With three independent datasets, we conducted a total of three ANOVAs in testing H_1 with *spatial congruence* (*congruent* vs. *incongruent*; within-subjects) and *participant position* (*left vs. right*; between-subjects) included as independent variables in the statistical model (see Table 2).

Consistent with [38] and the broader JSE literature, the data from all three sites showed a significant main effect of spatial congruence on response time ($ps < .02$). As expected, participants were faster to respond to spatially congruent stimuli than they were to respond to spatially incongruent stimuli ($ds > .37$). Thus, $\boldsymbol{H}_1$ **was supported**. Moreover, we did not observe a significant main effect of participants' *position* relative to the robot ($ps > .10$), nor a significant *congruence* × *position* interaction ($ps > .07$).

To explore whether any differences manifested due to the setting or cultural context in which data were acquired, we ran a further ANOVA with **site** (1, 2, or 3) included as an independent variable (in addition to *spatial congruence* and *participant position*); however, no significant main or interaction effects emerged ($ps > .07$). While the absence of significance is not in itself evidence of absence, these findings lend further support to the robustness of the joint Simon effect in the context of dyadic human-robot interactions – across three distinct settings and participant populations.

4.2 Association with Agency

In the original investigation, the framing of the robot co-actor (Tombatossals) was manipulated by describing it to participants as being more or less agentic. Participants to whom Tombatossals was introduced as more agentic attributed greater intentionality to the robot. Importantly, a larger joint Simon effect was observed in their response times to the stimuli as well [38]. Based on this finding, we hypothesized that the joint Simon Effect (Δ_{RT}) would positively correlate with the intentionality attributed to the NAO

Table 2: Results of hypothesis testing (H_1). Inferential and descriptive statistics (from left to right): Main effect of *congruence* on reaction time (F), degrees of freedom of the denominator (DFd), significance (p), effect size (*partial* η^2); mean difference in RT between incongruent and congruent trials (i.e., *joint Simon effect*; $M_{\Delta_{RT}}$), standard deviation (SD), and magnitude of the paired difference (Cohen's d_z).

	F	DFd	p	η_p^2	$M_{\Delta_{RT}}$	SD	d_z
Site 1	25.27	37	$< .01$	.40	9.74 *ms*	12.48 *ms*	.78
Site 2	6.36	45	$< .02$	.12	5.40 *ms*	14.59 *ms*	.37
Site 3	22.20	70	$< .01$	.24	6.72 *ms*	12.04 *ms*	.55

by participants here (H_2). Due to researcher error, the two questionnaire items indexing participants' attributions of *intentionality* to the NAO were absent from those used at Sites 1–2. Consequently, data from Site 3 only were available for testing of H_2. Analysis of that data revealed that while participants, on average, ascribed intentionality to the NAO ($M = 1.43$, $SD = .75$) on a scale from -2 = "disagree" to 2 = "agree", Pearson's Product-Moment Correlation test of the relationship between attributions and their Δ_{RT}, however, did not show a significant association: $t(70) = -1.30$ ($p = .19$). Thus, support was not found for H_2.

5 DISCUSSION

The aim of the present work was to investigate the *joint Simon effect* in dyadic human-robot interactions. To that end, we constructed an experimental infrastructure to enable *exact, lab-independent* reproduction of robot behavior and we deployed it across three sites comprising two distinct institutional profiles and three distinct cultural contexts, thereby addressing multiple levels of both empirical and technical reproducibility. In total, we carried out **three *conceptual reproductions* of the experiment by Stenzel et al. (2012)**. The reproductions also served as **three demonstrations of our technical approach**: *internal deployment* of our infrastructure (Site 1), *external deployment* at a university with a similar institutional profile to that of Site 1 (Site 2), and *ecological validation* via deployment at a university with limited institutional resources for supporting robotics research (Site 3).

5.1 Summary of Findings

Across all three sites, we observed a significant main effect of *spatial congruence* – i.e., a **joint Simon effect** (H_1). Simultaneously, we did not find participants' *positioning* relative to the robot or the *site* of data acquisition explained the difference in response times to congruent (vs. incongruent) stimuli. To summarize, we find that: (i) the JSE reliably manifests in people's completion of the joint Simon task with the NAO robot; (ii) the JSE is robust to the variability encapsulated by the composition of sites at which data was collected; and (iii) alternative factors (i.e., positioning, site) do not suffice to account for the effect. With the three sites representing distinct settings, culturally distinct populations, and different degrees of exposure to robotic platforms, these findings, in turn, suggest that **the joint Simon effect generalizes across both settings and populations**.

We did not, however, observe a significant association between participants' ascriptions of *intentionality* to the NAO and the differences in response time to congruent versus incongruent stimuli (Δ_{RT}). Thus, support was not found for H_2. Because the data from Sites 1–2 were not available for this analysis, we do not know whether significant associations between intentionality ascriptions and Δ_{RT} for those participants exist. Nevertheless, **we did not find evidence that the joint Simon effect is amplified by perceptions of agency in one's co-actor on the task**.

There are two possible interpretations of this finding: either no association exists or our method here was insufficient for detecting the association. Lacking additional data, we are unable to determine whether our observation at Site 3 mirrors that which we might find at the other sites (suggesting the finding to be a *true negative*), or whether the other sites differ (suggesting the finding to be a *false negative*). It may also or alternatively be that the Site 3 data are constrained by a ceiling effect, as participants' responses reflect a high degree of intentionality attributed ($M_{ratings} = 1.43$ out of 2.00 maximum). Moreover, there is limited variance in participants' responses ($SD = .75$). Thus, explicit manipulation of the robot's intentionality may be necessary (e.g., to induce sufficiently different attributions) for testing the existence of an association.

5.2 Contributions & Implications for HRI

While reproducibility has been vividly discussed in relation to the psychological, social, and medical sciences, its criticality is broadly evident. For example, the findings of a 2016 Nature survey called into question the general reliability of research across Biology, Physics, Engineering, and Medicine, where more than 60% of the researchers surveyed ($N = 1,576$) reported that they failed to reproduce experiments by other researchers and over 40% reported failures to reproduce their *own* work [2]. Although HRI research has not been similarly surveyed, the analysis of the 2017 ICRA Proceedings ([8]) – wherein only 2% of the publications referenced build-able and run-able source code, with only one that could be run off-the-shelf – taken together with *failures to replicate* prior HRI findings (e.g., [23, 39]), the *homogeneity of sampling* in HRI research [3], and indications that *perceptions of robots are demographically variable* (e.g., [27, 28, 32]) underscore a need for attention to the reproducibility of both systems and theories.

To motivate reproducible research, mechanisms have emerged such as dedicated publication venues with requirements of traceable and executable source code to ensure future access to contemporary robotics software (e.g., the Robotics and Automation Magazine; [5]). Still, there remain substantial barriers to carrying out HRI reproductions including the availability of robot platforms and software for exactly reproducing their behavior.

The aim of the present work was to explore these barriers, with respect to the joint Simon effect in a human-robot teaming scenario (the joint Simon task), thereby addressing multiple levels of both technical and experimental reproducibility. Because the initial findings by Stenzel and colleagues suggest human-robot teaming can be facilitated by human-like embodiments [38], the JSE – if robust – carries a strong design implication (i.e., increase the humanlikeness of robots intended for joint action with human partners). By pursuing a multi-institution research cooperation, we were able to

evaluate the robustness of the joint Simon effect across three geographically distinct sites with three culturally distinct populations. Exact, lab-independent reproduction of robot behavior, however, was necessary to ensure the observations made at the three sites were indeed comparable (i.e., were collected using the same methods). With a research cooperation that was also multidisciplinary, we were able to construct and deploy an experimental infrastructure for exactly reproducing the experimental methodology – including the NAO's behavior – across the three sites.

5.2.1 Joint Simon Effect. Consistent with the work by Stenzel et al. (2012), we found the joint Simon effect reliably manifested across all three sites. This provides further evidence, contrary to early research on the JSE (e.g., [42]), the phenomenon applies beyond joint actions with other humans. Moreover, observation of the JSE across the three sites indicates it to be robust to situational and cultural variation, implying it to generalize across settings and populations. We did not, however, find evidence that the effect is amplified by one's perception of the NAO's agency. Thus, further investigation of the design implications suggested by [38] is warranted.

5.2.2 HRI Reproducibility. To enable the above reproductions, **we integrated the functional components of the experiment into an existing toolchain which provided the means to version, test, deploy, and execute the software-driven aspects of our experiment semi-automatically, by scientists within the multidisciplinary scope of HRI, in a robust and consistent fashion across all three sites**. As a result, just steps ❷+❸ of Figure 3 were required for Sites 2–3. In our approach, we utilized the CITK to validate its applicability – which, as evidenced by the deployments and reproduction of [38] at Sites 1–3, was a success. Of particular note is that, with Sites 2–3, we have preliminarily demonstrated the success of our approach beyond internal deployment. Moreover, evidenced by successful deployment in an education-focused (as opposed to research-focused) institution, we suggest that the approach can facilitate broader participation in reproducibility.

5.3 Limitations & Avenues for Future Research

Although the present methodology was well-suited to address the robustness of the joint Simon effect in HRI, we nevertheless encountered challenges and limitations in the execution thereof which highlight avenues for future research. First and foremost, despite our best efforts we made mistakes. For example, a gap in the technical documentation (specifying compatible browser *versions*) lead to use of an untested browser release at Site 3. This, in turn, unearthed a regression in our jsPsych plug-in implementation – which required coordination with the developers to remedy. In addition, despite our focus on reproducibility, questionnaire items for evaluating H_2 were missed in translation between the three sites. Therefore we suggest that, even with an eye on best practices, translating artifacts across laboratories creates additional room for error. Thus, beyond the mere reproduction of prior findings, careful reflection on the needs of *multi-site* reproductions may be fruitful toward HRI reproducibility more broadly.

The above items also highlight issues that arose with the aimed independence of the research teams. Specifically, while we aimed for the deployments of our system to execute independently of the development team, consultation was necessary in several instances. For example, unanticipated elevated network security prevented communication with the robot at Site 2. Even though this was resolved independently by local staff (i.e., a local network via router was established), a possible malfunction was discussed with the developers first. Overall, none of the items fell outside the scope of what can arise in setting up an HRI experiment; however, some cannot be taken care of in the technical infrastructure (e.g., local network security level such as firewall configuration). That is, the pipeline (toolchain) worked as expected, but problems outside of the CITK scope nevertheless had to be addressed. Therefore, we suggest that future applications of this approach will likely need a dedicated "point person" for managing deployment of the infrastructure.

Finally, while we first planned to provide "just" a docker or virtual machine (VM) image instead of bootstrapping the CITK natively, their configurations introduce additional setup complexity (e.g, regarding the network setup of a VM). Furthermore, docker requires non-trivial setup just to run GUI applications on a host machine. However, simplification of docker configuration (e.g., reusing [9]) may be a valuable future avenue, since the CITK supports docker-based deployment and the combination of the two technology stacks could reduce limitations while still providing a well-defined process.

6 CONCLUSIONS

We pursued a multi-disciplinary and multi-institutional research cooperation to investigate the robustness of the joint Simon effect in human-robot teaming. Construction of an experimental infrastructure enabled exact, lab-independent reproduction of robot behavior and successful deployment across three sites suggests it can lower barriers to reproducibility in HRI. Moreover, with the three deployments situated in distinct contexts, reproduction of the JSE at each site suggests the phenomenon generalizes across both settings and populations. However, even with a well-developed technical approach, there were nevertheless challenges encountered in carrying out the external deployments – suggesting that at least some involvement by developers is necessary in HRI reproductions.

7 SUPPLEMENTAL MATERIAL

- VirtualBox image of CITK infrastructure: **techfak.de/clf-distro/archive/JSE-NAO.ova**
- Document as provided to repro. sites: **perma.cc/QZ2A-H38N**
- CITK deployment process (video): **perma.cc/4HBY-LJFD**
- CITK git repository: **perma.cc/T57S-WLWV**
- Statistical analyses: **codeocean.com/capsule/8848932/**

ACKNOWLEDGMENTS

Thanks to Martin Strob for helping to implement the NAO web bridge, Hannah Koppenrade for helping to execute the experiment in Bielefeld, Yiyao Wei for helping to translate the survey to English and execute the experiment at Indiana University, and the UTRGV research assistants – Jaime Alvarez Perez, Virginia Contreras, Hideki Garcia, Noemi Garcia, and Ana Sánchez Ramos – for helping to execute the experiment in Texas. Finally, thanks to Jan Moringen for the excellent work on the CITK project at large.

REFERENCES

[1] Jeffrey J Arnett. 2008. The neglected 95%: why American psychology needs to become less American. *American Psychologist* 63, 7 (2008), 602.

[2] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970. [Online; accessed 16-September-2019].

[3] Paul Baxter, James Kennedy, Emmanuel Senft, Severin Lemaignan, and Tony Belpaeme. 2016. From characterising three years of HRI to methodology and reporting recommendations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 391–398.

[4] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science robotics* 3, 21 (2018), eaat5954.

[5] Fabio Bonsignorio. 2017. A new kind of article for reproducible research in intelligent robotics [from the field]. *IEEE Robotics & Automation Magazine* 24, 3 (2017), 178–182.

[6] Joost Broekens, Marcel Heerink, Henk Rosendal, et al. 2009. Assistive social robots in elderly care: a review. *Gerontechnology* 8, 2 (2009), 94–103.

[7] Evan C Carter, Lilly M Kofler, Daniel E Forster, and Michael E McCullough. 2015. A series of meta-analytic tests of the depletion effect: self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General* 144, 4 (2015), 796.

[8] E. Cervera. 2019. Try to Start It! The Challenge of Reusing Code in Robotics Research. *IEEE Robotics and Automation Letters* 4, 1 (Jan 2019), 49–56. https://doi.org/10.1109/LRA.2018.2878604

[9] Enric Cervera and Angel Del Pobil. 2019. ROSLab: Sharing ROS Code Interactively with Docker and JupyterLab. *IEEE Robotics & Automation Magazine* (2019).

[10] Eris Chinellato, Marco Antonelli, Beata J Grzyb, and Angel P Del Pobil. 2011. Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Transactions on Autonomous Mental Development* 3, 1 (2011), 43–53.

[11] Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716.

[12] Joshua R De Leeuw. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods* 47, 1 (2015), 1–12.

[13] Thomas Dolk, Bernhard Hommel, Lorenza S Colzato, Simone Schütz-Bosbach, Wolfgang Prinz, and Roman Liepelt. 2014. The joint Simon effect: a review and theoretical integration. *Frontiers in Psychology* 5 (2014), 974.

[14] Stéphane Doyen, Olivier Klein, Cora-Lise Pichon, and Axel Cleeremans. 2012. Behavioral priming: it's all in the mind, but whose mind? *PloS one* 7, 1 (2012), e29081.

[15] Paul M Duvall, Steve Matyas, and Andrew Glover. 2007. *Continuous integration: improving software quality and reducing risk*. Pearson Education.

[16] Brian D Earp and David Trafimow. 2015. Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology* 6 (2015), 621.

[17] Jim Albert Charlton Everett and Brian D Earp. 2015. A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *Frontiers in Psychology* 6 (2015), 1152.

[18] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[19] Kurt Gray and Daniel M Wegner. 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* 125, 1 (2012), 125–130.

[20] Martin S Hagger, Nikos LD Chatzisarantis, Hugo Alberts, Calvin O Anggono, Cedric Batailler, Angela R Birt, Ralf Brand, Mark J Brandt, Gene Brewer, Sabrina Bruyneel, et al. 2016. A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science* 11, 4 (2016), 546–573.

[21] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33, 2-3 (2010), 61–83.

[22] Bernhard Hommel. 1996. SR compatibility effects without response uncertainty. *The Quarterly Journal of Experimental Psychology Section A* 49, 3 (1996), 546–571.

[23] Bahar Irfan, James Kennedy, Séverin Lemaignan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme. 2018. Social psychology and human-robot interaction: An uneasy marriage. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 13–20.

[24] Sven Kepes and Michael A McDaniel. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology* 6, 3 (2013), 252–268.

[25] Richard A Klein, Kate A Ratliff, Michelangelo Vianello, Reginald B Adams Jr, Štěpán Bahník, Michael J Bernstein, Konrad Bocian, Mark J Brandt, Beach Brooks, Claudia Chloe Brumbaugh, et al. 2014. Investigating variation in replicability. *Social Psychology* (2014).

[26] Richard A Klein, Michelangelo Vianello, Fred Hasselman, Byron G Adams, Reginald B Adams Jr, Sinan Alper, Mark Aveyard, Jordan R Axt, Mayowa T Babalola, Štěpán Bahník, et al. 2018. Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science* 1, 4 (2018), 443–490.

[27] Hee Rin Lee and Selma Sabanović. 2014. Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 17–24.

[28] Hee Rin Lee, Haodan Tan, and Selma Šabanović. 2016. That robot is not for me: Addressing stereotypes of aging in assistive robot design. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 312–317.

[29] Florian Lier, Marc Hanheide, Lorenzo Natale, Simon Schulz, Jonathan Weisz, Sven Wachsmuth, and Sebastian Wrede. 2016. Towards automated system and experiment reproduction in robotics. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3298–3305.

[30] Florian Lier, Johannes Wienke, Arne Nordmann, Sven Wachsmuth, and Sebastian Wrede. 2014. The cognitive interaction toolkit–improving reproducibility of robotic systems experiments. In *International Conference on Simulation, Modeling, and Programming for Autonomous Robots*. Springer, 400–411.

[31] Michael Mulkay and G Nigel Gilbert. 1986. Replication and mere replication. *Philosophy of the Social Sciences* 16, 1 (1986), 21–37.

[32] André Pereira, Catharine Oertel, Leonor Fermoselle, Joe Mendelson, and Joakim Gustafson. 2020. Effects of Different Interaction Contexts when Evaluating Gaze Models in HRI. In *Proceedings of the 2020 ACM/IEEE international conference on Human-robot interaction*. ACM.

[33] Laurel Riek and Don Howard. 2014. A code of ethics for the human-robot interaction profession. *Proceedings of We Robot* (2014).

[34] Laurel D Riek. 2017. Healthcare robotics. *Commun. ACM* 60, 11 (2017), 68–78.

[35] Natalie Sebanz, Günther Knoblich, and Wolfgang Prinz. 2003. Representing others' actions: just like one's own? *Cognition* 88, 3 (2003), B11–B21.

[36] Natalie Sebanz, Günther Knoblich, and Wolfgang Prinz. 2005. How two share a task: corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance* 31, 6 (2005), 1234.

[37] J Richard Simon. 1969. Reactions toward the source of stimulation. *Journal of experimental psychology* 81, 1 (1969), 174.

[38] Anna Stenzel, Eris Chinellato, Maria A Tirado Bou, Ángel P del Pobil, Markus Lappe, and Roman Liepelt. 2012. When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance* 38, 5 (2012), 1073.

[39] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Reliability of NIRS-Based BCIs: a placebo-controlled replication and reanalysis of Brainput. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 619–630.

[40] Megan Strait, Heather L Urry, and Paul Muentener. 2019. Children's Responding to Humanlike Agents Reflects an Uncanny Valley. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 506–515.

[41] Chia-Chin Tsai and Marcel Brass. 2007. Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science* 18, 12 (2007), 1058–1062.

[42] Chia-Chin Tsai, Wen-Jui Kuo, Daisy L Hung, and Ovid JL Tzeng. 2008. Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience* 20, 11 (2008), 2015–2024.