New Insights into User-Defined Smart Ring Gestures with Implications for Gesture Elicitation Studies

BOGDAN-FLORIN GHERAN, Ştefan cel Mare University of Suceava, Romania
RADU-DANIEL VATAVU, Ştefan cel Mare University of Suceava, Romania
JEAN VANDERDONCKT, Université catholique de Louvain, Belgium

We conduct a replication of a previously published gesture elicitation study that collected and examined user-defined gestures performed with smart rings. Our findings reveal that only 50% of the gestures elicited during the replication match those proposed by the participants of the original study, although the new sample of participants had the same characteristics as the sample of the original study, and the new participants exhibited the same amount of thinking time to propose gestures and rated the goodness of their gesture-to-function mappings at the same level as the participants of the original study. Furthermore, the new gestures present different articulation characteristics in terms of structure, complexity, symmetry, and locale. Our results have several implications for the theory and practice of gesture elicitation studies, for which we provide recommendations to researchers and practitioners.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Gestural input**; **User studies**.

Additional Key Words and Phrases: Gesture elicitation, replicability, experiment, smart rings, gesture input, RepliCHI.

ACM Reference Format:
Bogdan-Florin Gheran, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2023. New Insights into User-Defined Smart Ring Gestures with Implications for Gesture Elicitation Studies. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3544549.3585590

1 INTRODUCTION

As a distinctive type of a finger-augmentation device [23], smart rings conveniently enable several gesture input modalities [26], such as touch (on the ring surface), motion (by rolling or sliding the ring on the finger), and mid-air input (by drawing symbols in mid-air). However, the vocabulary of gesture commands for smart rings has been quite limited. For example, Magic Ring [14] employs four gestures only: pointing the finger up, down, left, and right; RingIoT [6] implements pointing, directional movements, and simple geometric shapes, such as a "circle" drawn in mid-air; and Zhang *et al.*'s [36] ring uses six gestures: four directional swipes, "click," and "circle." Despite the growing interest in rings for interactive systems [4,21,24,26,31,37], users' preferences for ring gestures have been little examined.

In this context, Gheran *et al.* [8] conducted the first study to explore users' preferences for gestures performed with smart rings. From a dataset of 672 gestures elicited from 24 participants, the authors reported an overall low consensus, but interesting insights about participants' mental models and articulation characteristics of ring gestures, e.g., 69% of the proposed gestures were performed in mid-air and 42% of the bimanual gestures were asymmetric. To arrive at these results, Gheran *et al.* implemented a "gesture elicitation study" [33,34], which is a specialized instance of participatory design [3,17], useful to unveil users' preferences for gesture input. In these studies, participants propose

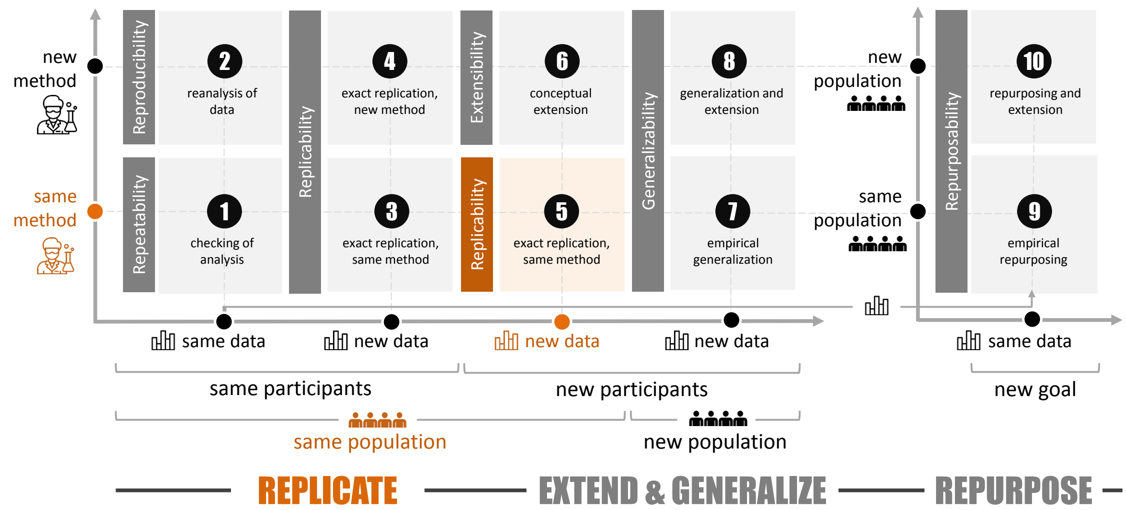

Fig. 1. Positioning of our replication in the RepliGES space [10] of gesture elicitation studies. Our study is a "type-5 replication" with the same goal and the same analysis method as the original study [8], but with new data/gestures collected from new participants.

gestures they would like to use to effect specific system functions, such as turning off the TV with a smart ring. The practitioner analyzes the elicited gestures to identify the most preferred [28] and to understand which gesture commands are perceived intuitive [33], low effort [34], and memorable [18], likely to reflect the preferences of the larger user population. With time, this understanding should turn into consolidated knowledge [15,29,30] via repeated validation and replication, on which the community can capitalize to inform gesture input design. However, a recent paper [10] revealed very few replications of gesture elicitation studies despite their large popularity in the HCI community [29].

In this context represented by a growing scientific interest in smart rings for diverse application domains and contexts of use [26], but overall few replications of gesture elicitation studies [10], we contribute a replication of Gheran *et al.* [8]. Our results show that only 50% of the gestures discovered during our replication match those proposed by the participants of the original study, although we used a similar sample and the same analysis method; see Figure 1. Based on our results, we stress the importance of replications to confirm and consolidate—or contradict and refute—previous findings about user-defined gestures, and draw implications for the theory and practice of gesture elicitation studies.

2 RELATED WORK

Relevant related work is represented by the scientific literature on gesture elicitation studies, for which we refer readers to Wobbrock *et al.*'s [33,34] seminal papers, Villarreal *et al.*'s [29] systematic literature review, and Vatavu and Wobbrock's [28] critical overview and clarification of methodological aspects. According to the operational model for general end-user elicitation studies in HCI [28], the outcomes of such studies are represented by consensus sets of signs (e.g., consensus gesture sets for gesture elicitation studies) and/or characterizations of traits to understand user behavior with interactive systems. However, the outcomes of any study need to be confirmed towards consolidating knowledge, but replications have been overall limited in HCI [13]. To foster replication of gesture elicitation studies, Gheran *et al.* [10] introduced RepliGES, which we illustrate in Figure 1 to position our work as a "type-5 replication."

Despite the popularity of gesture elicitation studies [29], examination of user-defined ring gestures has been scarce. Gheran *et al.* [8] conducted the first elicitation study about ring gestures motivated by the potential of ring-based input for a variety of application domains and contexts of use. We performed a forward snowballing procedure[1] [35] starting from [8] and identified 87 references, of which only two [2,9] were about gesture elicitation: Gheran *et al.* [9], a "type-9 empirical repurposing," according to RepliGES terminology, where data from [8] were analyzed for a different goal, and Bardot *et al.* [2], who conducted a gesture elicitation study focused on a specific ring designed for input while grasping objects. In this context, design knowledge about ring gestures reflective of users' preferences remains very limited.

3 EXPERIMENT

We performed a replication of the first study on users' preferences for ring gestures [8]. We chose a "type-5 replication," according to RepliGES terminology (Figure 1), since it is the only replication type not yet conducted for gesture elicitation studies; see [10, p. 5:5]. Moreover, a type-5 replication enables verification of prior findings by being both *exact* [25] and *strict* [13], i.e., the same method is used as in the original study with gestures elicited from new participants.

3.1 Participants

We recruited 24 participants (15 male, 9 female) aged between 22 and 45 years (M=26.3, SD=5.7). The size and gender structure of our sample were identical to [8, p. 625], while an independent-samples *t*-test did not detect any significant difference between the mean age of the original participants and our sample (27.5 vs. 26.3 years, $t_{(46)}$=0.639, p=.526).

3.2 Design

Our experiment was a within-subjects design with two independent variables, replicating the design of [8]:

(1) Referent, nominal variable with five conditions representing various system functions common to a smart home environment: *turn TV on/off, start music player, turn volume up, turn lights on/off,* and *go to the next item in a list.* These referents are a subset of those used in [8]; see next for more details about our choice.
(2) Num-Rings, ordinal variable with two conditions, *one-ring* and *two-ring* gestures, specifying two contexts of use corresponding to unimanual and bimanual gesture input, respectively, as examined in [8].

3.3 Procedure and Measures

We did our best to replicate all of the implementation details of the original study as closely as possible. Namely, we recruited a sample of participants of equal size and similar age and gender profiles, used the same apparatus (two Ring Zero[2] devices), and presented the same instructions to our participants as in [8, p. 626]; see Table 1. Furthermore, an author of the original study[3] agreed to analyze on our behalf the videos of the elicited gestures using the same procedure employed in the original study [8, pp. 626-627]. However, to avoid taking too much of their volunteered time, we selected just five of the fourteen referents used in the original study as the conditions of our Referent independent variable. We specifically chose these five referents because of their contrasting agreement rates, spanning a wide range in the original study, from the highest agreement (.225 for *next*) to the lowest (.040 for *lights on/off*); see [8, p. 627]. Furthermore, following [8, pp. 627-628], the elicited gestures were classified along four dimensions,[4] as follows:

- *Structure* characterizes the importance of hand movement and hand poses during ring gesture articulation with five categories: *buttons only*, *hand poses only*, *hand motion*, *hand poses & motion*, and *mixed locales.*

[1] Using Google Scholar, query ran on October 15, 2022.
[2] https://www.buysmartrings.com/product/ring-zero-smart-ring-white
[3] The gesture elicitation study [8] had three authors, but only one performed the gesture clustering procedure to calculate agreement rates.
[4] Gheran *et al.* [8] also employed the *Nature* dimension, according to which gestures were classified as *symbolic*, *metaphorical*, and *abstract*. Since this dimension is less important for the implementation of a gesture recognizer, we did not pursue it in our replication; see more discussion in [19].

Table 1. Characteristics of the original gesture elicitation study and our replication.

Characteristic	Original study [8]	Replication (this paper)
Participants	N=24 (15 male and 9 female) between 21 and 45 years old (M=27.5, SD=7.9 years)	N=24 (15 male and 9 female) between 22 and 45 years old (M=26.3, SD=5.7 years)
Apparatus	two Ring Zero devices, video recording	two Ring Zero devices, video recording
Design	within-subjects with two independent variables: REFERENT and NUM-RINGS	within-subjects with two independent variables: REFERENT and NUM-RINGS
REFERENT	fourteen referents, of which T*V on/off, start player, volume up, lights on/off, go to the next item in a list*	five referents: *TV on/off, start player, volume up, lights on/off, go to the next item in a list*
NUM-RINGS	*one-ring* and *two-ring*, randomized across participants	*one-ring* and *two-ring*, randomized across participants
Gesture set	672 elicited gestures, of which 240 for the five referents used in the replication	240 elicited gestures corresponding to the five referents used in the replication
Measures	AR [27], THINKING-TIME (seconds), GESTURE-GOODNESS (10-point Likert scale rating)	AR [27], THINKING-TIME (seconds), GESTURE-GOODNESS (10-point Likert scale rating)

- *Complexity* differentiates between *simple* and *compound* ring gestures, where the latter can be decomposed into individually meaningful gestures.
- *Symmetry* describes how the two hands synchronize to produce ring gestures, with four categories: *dominant unimanual*, *nondominant unimanual*, *symmetric bimanual*, and *asymmetric bimanual*.
- *Locale* specifies where gestures are performed: *on the ring*, *on other surface*, *in the air*, and *mixed locales*.

We also used the following measures from [8]:

- The Agreement Rate (AR) of a referent r is the proportion of pairs of participants in agreement over their gesture proposals for referent r out of the total number of pairs that could be in agreement [27]:

$$AR(r) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \delta_{i,j} \tag{1}$$

  where N is the number of participants and $\delta_{i,j}$ evaluates to 1 if the gestures elicited from the i-th and j-th participants are equivalent or substantially similar and 0 otherwise. AR computes in the interval $[0, 1]$.
- THINKING-TIME represents the time, in seconds, taken by a participant to propose a gesture for referent r.
- GESTURE-GOODNESS represents participants' subjective assessments, as a rating between 1 and 10, of how well the proposed gestures fit the referents (10 indicates the best fit). This measure was introduced by Wobbrock *et al.* [34] with a 7-point Likert scale but, for purposes of replication fidelity, we used the 10-point scale from [8].

We also employed several new measures to compare the results of our replication against the original study, as follows:

- DISTINCT-GESTURES represents the number of unique gestures, either in the original study or the replication.
- SHARED-PROPORTION-DISTINCT represents the proportion of distinct gestures found in both the original study and the replication. For example, if the original study identified 42 distinct gestures and the replication 49, of which 40 were also found in the original study, then SHARED-PROPORTION-DISTINCT is 40/49=81.6%.
- SHARED-PROPORTION-ALL represents the proportion of all gestures found in both the original study and the replication. Unlike for SHARED-PROPORTION-DISTINCT, repetitions are considered. For example, if the 40 distinct

Table 2. Results obtained in the original gesture elicitation study and our replication.

		One-ring gestures			Two-ring gestures		
Criteria	**Category**	**Orig.**	**Replication**	**Orig. *vs.* Replication**	**Orig.**	**Replication**	**Orig. *vs.* Replication**
Total number of elicited gestures		120	120	–	120	120	–
DISTINCT-GESTURES		47	57	$\chi^2_{(1)}$=1.374, p=.241, $n.s.$	63	67	$\chi^2_{(1)}$=0.151, p=.698, $n.s.$
SHARED-PROPORTION-DISTINCT		–	13/57=22.8%	–	–	15/67=22.3%	–
SHARED-PROPORTION-ALL		–	66/120=55.0%	–	–	60/120=50.0%	–
Average Agreement Rate (AR)		.118	.090	-23.7%	.085	.065	-23.5%
	AR (*TV on/off*)	.120	.069	-42.5%	.065	.036	-44.6%
	AR (*start player*)	.076	.025	-67.1%	.018	.025	+38.9%
	AR (*volume up*)	.130	.083	-36.2%	.062	.054	-12.9%
	AR (*lights on/off*)	.040	.029	-27.5%	.134	.054	-59.7%
	AR (*next*)	.225	.243	+8.0%	.145	.156	+7.6%
THINKING-TIME (s)		4.9	3.4	$t_{(46)}$=1.678, p=.100, $n.s.$	3.1	3.0	$t_{(46)}$=0.121, p=.904, $n.s.$
GESTURE-GOODNESS		7.9	8.0	$t_{(46)}$=0, p=1, $n.s.$	8.1	8.4	$t_{(46)}$=1.288, p=.204, $n.s.$
Structure†	*buttons only*	14	15	p=1, $n.s.$	12	13	p=1, $n.s.$
	hand motion	13	78	p<.001	48	82	p<.001
	hand motion & poses	67	18	p<.001	41	17	p<.001
	hand poses	8	3	p=.217, $n.s.$	5	4	p=1, $n.s.$
	mixed locales	18	6	p=.018, $n.s.$	14	4	p=.027, $n.s.$
Complexity†	*simple*	88	92	p=.655, $n.s.$	100	80	p=.005
	compound	32	28	p=.655, $n.s.$	20	40	p=.005
Symmetry†	*dom. unimanual*	117	94	p<.001	52	82	p<.001
	nondom. unimanual	0	2	p=.478, $n.s.$	7	0	p=.021, $n.s.$
	symmetric bimanual	3	13	p=.020, $n.s.$	44	19	p<.001
	asymmetric bimanual	0	11	p=.002	17	19	p=.857, $n.s.$
Locale†	*on the ring*	19	16	p=.715, $n.s.$	14	13	p=1, $n.s.$
	in the air	88	85	p=.774, $n.s.$	82	87	p=.572, $n.s.$
	on other surface	0	8	p<.0125	13	7	p=.243, $n.s.$
	mixed locales	13	11	p=.830, $n.s.$	11	13	p=.830, $n.s.$

†Bonferroni corrections were applied at the level of significance α=.05/n, where n is the number of comparisons. For example, α was .05/5=.01 for the tests of proportions conducted for the *Structure* dimension, .05/4=.0125 for the *Locale* dimension, etc.

gestures shared between the two studies were proposed for 111 times (repetitions included) of the 120 gestures elicited during the replication, then SHARED-PROPORTION-ALL is 111/120=92.5%.

4 RESULTS

The average agreement rates reported for the five referents in the original study [8] were AR=.118 and .085 for the *one-ring* and *two-ring* conditions. In our replication, the average ARs were .090 and .065, representing low agreement as in the original study, but nevertheless 24% smaller. To understand these results, we analyze the number of distinct gestures elicited in each study and the proportion of shared gestures between the two studies. But first, we examine the thinking times and gesture goodness ratings to understand similarities in participants' behavior and perceptions.

4.1 Thinking Time and Gesture Goodness

The average THINKING-TIME needed by the participants of the original study to propose gestures was 4.9s for the *one-ring* condition and 3.1s for *two-ring* [8, p. 629]. The average GESTURE-GOODNESS was 7.9 and 8.1 [8, p. 629], showing high

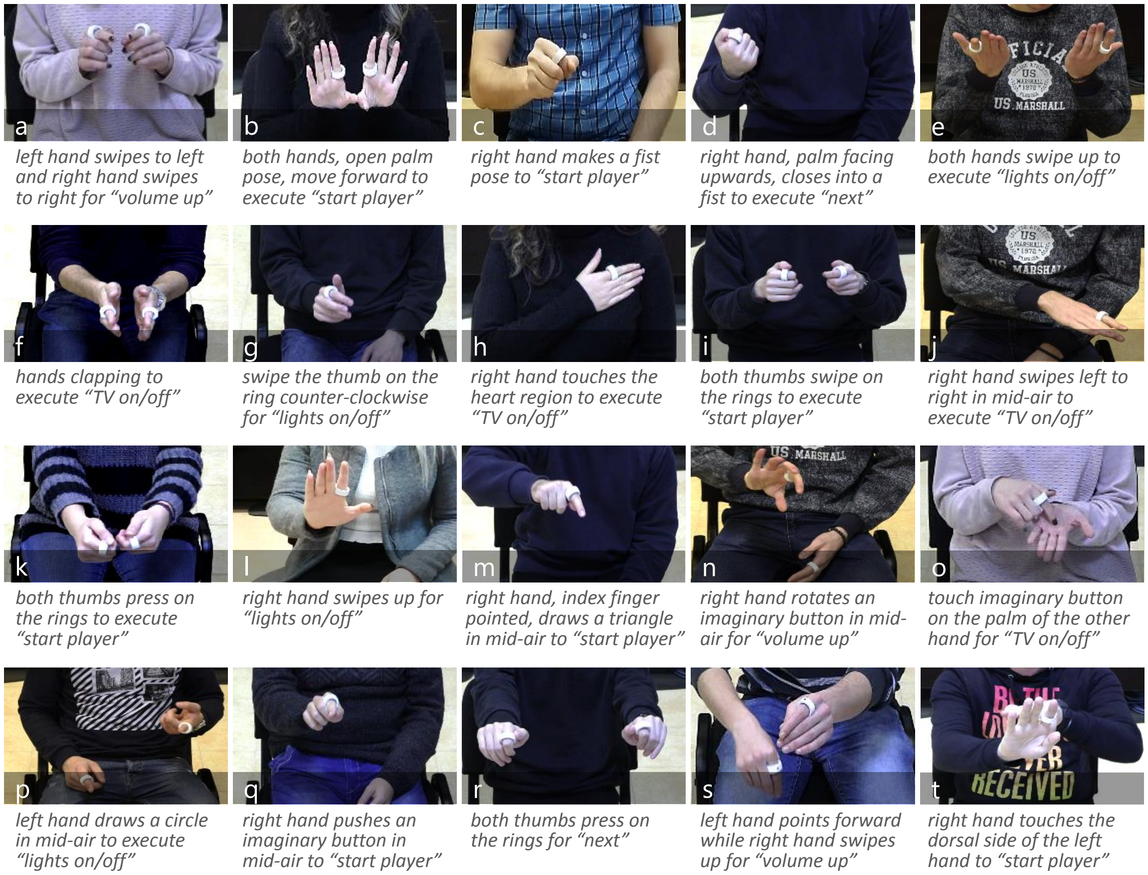

Fig. 2. A selection of user-defined gestures from our replication, not discovered in the original study [8].

confidence of the original study participants in their gesture-to-referent mappings. The results of our replication were remarkably similar (Table 2, rows 11-12), and independent-samples t-tests showed no significant differences between the mean Thinking-Time (3.4s vs. 3.0s) and Gesture-Goodness (8.0 vs. 8.4) measured in our replication compared to the original study. These results confirm the findings of [8] and highlight similarities between [8] and our replication in terms of participant behavior (Thinking-Time) and perception of their proposed gestures (Gesture-Goodness).

4.2 Unique and Shared Gestures

The original study identified 47 distinct gestures of the 120 elicited in the *one-ring* condition and 63 distinct gestures of the 120 elicited in the *two-ring* condition. With new participants and the same gesture analysis procedure, we found 57 distinct gestures for *one-ring* and 67 for *two-ring*, respectively (Table 2, 2nd row). A Pearson chi-square test for the equality of proportions (Bonferroni corrections applied) did not find any significant difference between the Distinct-Gestures of the original study and our replication. The numbers of shared distinct gestures were only 13 and 15, respectively, resulting in a Shared-Proportion-Distinct of approximately 22.0% for both Num-Rings conditions. When counting gesture repetitions, Shared-Proportion-All were 66/120=55% and 60/120=50%, respectively.

These results show that about half of the gestures elicited in our replication are new compared to the original study, and complement the dataset reported in [8]. Figure 2 illustrates a selection of these gestures. For example, one of our participants proposed a bimanual gesture, where the left hand swipes to the left and the right hand to the right, to *turn the volume up* (Figure 2a). The same participant proposed a *one-ring* gesture, where the finger touches an imaginary button on the opposite hand (Figure 2o), to *turn the TV on/off*. Another participant pushed both hands forward to *start the music player* (Figure 2b) and moved the hand wearing the ring close to the heart to *turn the TV on/off* (Figure 2h), which are two examples of body-referenced ring gestures not discovered in the original study. Furthermore, gestures similar to those proposed by the original participants were attributed to different referents by the participants from the replication, such as moving the finger up and down to *turn the lights on/off* (Figure 2l). For further examination, we provide the full list of elicited gestures as a companion dataset to our paper; see Subsection 5.2. The important takeaway is that our replication confirmed only 50% of the ring gestures found in the original study of Gheran *et al.* [8].

4.3 Properties of User-Defined Gestures

We used the *Structure, Complexity, Symmetry,* and *Locale* dimensions to analyze the gestures proposed by the new participants. Table 2 shows that most of the comparisons conducted between the original study and our replication resulted in non-significant differences (Bonferroni corrections applied) in terms of the proportion of the elicited gestures falling in the categories of these dimensions. For instance, the original study reported a number of 19 gestures performed *on the ring surface* in the *one-ring* condition and our replication found 16 (p=.715>.05/4=.0125, *n.s.*). However, we also detected several statistically significant differences. These included the *hand motion* and *hand motion & poses* categories of the *Structure* dimension in both the *one-ring* and *two-ring* conditions (p<.001), and the *simple* and *compound* categories of the *Complexity* dimension in the *two-ring* condition (p=.005); see Table 2 for details. These results show that not only did our replication uncover new gestures compared to the original study, but those gestures also have different characteristics in terms of *Structure*, *Complexity*, *Symmetry*, and *Locale* compared to the original ones.

5 DISCUSSION

We use our empirical findings to outline several implications for future work on replicating gesture elicitation studies, but also on the theory and practice of such studies. We also present limitations of our experiment.

5.1 Implications for Gesture Elicitation Studies

The most important finding of our replication experiment is that we confirmed only half of the gestures found in the original study of Gheran *et al.* [8], while the other half represented new gestures, not proposed by the original participants, with different articulation characteristics. One key implication of this finding is that conducting gesture elicitation studies with around 20-25 participants[5] most likely limits the discovery of a representative set of user-defined gestures reflective of the preferences of the larger end-user population. On a more general level of the whole HCI field, a statement from one of the RepliCHI panels/workshops [32] corroborates on this aspect: “As a community, therefore, we do not value, facilitate, or reward replication in research, and often take the significant results of a single user study on 20 users to be true” (p. 463). Two subsequent implications immediately follow: (i) replications of previously published gesture elicitation studies that employed around 20-25 participants need to be conducted in order to confirm (or contradict) their findings, and consolidate (or discover new) user-defined gestures, and (ii) future work is needed to

[5]Villarreal *et al.* [29] reported, in their systematic literature review of 216 gesture elicitation studies, that the most frequent choice of sample size was 20 participants (36% of the published studies) and the average number of participants was 25 (p. 859).

understand the minimum number of participants to involve in a gesture elicitation study to be able to compile with high confidence a consensus gesture set representative of the preferences of the larger end-user population.

The first implication strengthens Hornbæk *et al.*'s [13] recommendation, following the results of their literature review, to conduct more replications in the field of HCI in order to confirm, expand, or generalize the findings reported by earlier studies. In the specific context of gesture elicitation, the replication types of the RepliGES space [10], illustrated in Figure 1, can be used to inform new studies according to the goal, method, data, and population criteria. The RepliGES space was built on the replication literature [11,13,25,32], from which it adopted methods and provided terminology adaptations to speak directly to HCI practitioners and help identify replication types relevant for gesture elicitation studies. For example, repurposability (type-9 and type-10 in Figure 1) and reproducibility (type-2) can be implemented to discover more about users' preferences for gesture input in contexts of use [5] other than those addressed in the original studies. Unfortunately, many gesture elicitation studies do not fully specify the context of use in their experimental setup, which makes replications difficult. Nevertheless, such replications are recommended, since space and place are known to influence how participants engage in a gesture elicitation study [1] and location and audience influence the social acceptability of gestures performed in public places [22]. Moreover, different user categories may have different gesture preferences according to their cultural background [7,16]. In this context, instead of focusing exclusively on originality, i.e., publish gesture elicitation studies for the newest application domains or interactive systems, incremental research in the form of replicating previous findings—e.g., agreement rates, gesture sets, gesture articulation characteristics, etc.—is needed to consolidate our current knowledge of user-defined gestures. This recommendation aligns with Hornbaek's [12] opinion and recommendation of being "more wrong" in HCI research.

The second implication recommends further work on the theoretical foundations of end-user elicitation studies. Besides formalizing agreement calculation and analysis [28], the impact of the sample of participants on the results of a gesture elicitation study should be equally considered. A desirable outcome in this regard is a mathematical model to estimate the sample size needed to confidently discover gestures representative for the larger end-user population. To this end, Nielsen and Landauer's [20] approach to finding usability problems can represent a good starting point for informing such a model, while Villarreal *et al.*'s [29] systematic literature review of gesture elicitation studies represents a rich source of information about the characteristics of samples employed in previous studies.

5.2 Limitations

There are two limitations to our work. First, we elicited gestures for only a subset of the referents of the original study. Although this subset was enough to discover new gestures, not proposed in the original study, running a full replication with all of the referents is recommended towards ring gesture sets that better reflect users' preferences. For such examinations, we provide the full list of the elicited gestures as companion data to our paper at http://anonymous.web.address. Second, we propose implications for gesture elicitation studies informed by our specific findings on ring gestures. Given that smart rings represent new technology with which users are little or not familiarized, the low agreement rates observed in the original study and our replication may be the direct result of technology novelty. For other application domains or interactive devices, e.g., multitouch surfaces, replications could confirm more of the original results. Nevertheless, future work is needed to replicate the findings of gesture elicitation studies.

6 CONCLUSION

In this late-breaking work, we conducted a replication experiment about user-defined ring gestures in the context of a growing interest for smart rings and ring-based input. We found that only half of the gestures of the original study were confirmed as the new participants proposed new gestures. Our findings have implications for the replication of gesture elicitation studies in order to confirm and consolidate—or contradict and refute—their findings, and also for the theoretical foundations and best practices of running such studies. We hope that our results and corresponding implications will be inspiring for the community, and we look forward to future work in this direction.

REFERENCES

[1] Imeh Akpan, Paul Marshall, Jon Bird, and Daniel Harrison. 2013. Exploring the Effects of Space and Place on Engagement with an Interactive Installation. In *Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2213–2222. https://doi.org/10.1145/2470654.2481306

[2] Sandra Bardot, Surya Rawat, Duy Thai Nguyen, Sawyer Rempel, Huizhe Zheng, Bradley Rey, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2021. ARO: Exploring the Design of Smart-Ring Interactions for Encumbered Hands. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. ACM, New York, NY, USA, Article 12, 11 pages. https://doi.org/10.1145/3447526.3472037

[3] Susanne Bødker and Morten Kyng. 2018. Participatory Design That Matters - Facing the Big Issues. *ACM Transactions on Computer-Human Interaction* 25, 1 (2018), 31 pages. https://doi.org/10.1145/3152421

[4] Victor Cheung and Audrey Girouard. 2019. Tangible Around-Device Interaction Using Rotatory Gestures with a Magnetic Ring. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. ACM, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3338286.3340137

[5] Joëlle Coutaz, James L. Crowley, Simon Dobson, and David Garlan. 2005. Context is Key. *Commun. ACM* 48, 3 (2005), 49–53. https://doi.org/10.1145/1047671.1047703

[6] Rajkumar Darbar, Mainak Choudhury, and Vikalp Mullick. 2017. *RingIoT: A Smart Ring Controlling Things in Physical Spaces.* Technical Report EBSE-2007-01. Samsung R&D Institute, Bangalore, India. https://doi.org/10.13140/RG.2.2.14245.76007

[7] Jane L. E, Ilene L. E, James A. Landay, and Jessica R. Cauchard. 2017. Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6794–6799. https://doi.org/10.1145/3025453.3025755

[8] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 623–635. https://doi.org/10.1145/3196709.3196741

[9] Bogdan-Florin Gheran, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2018. Ring X2: Designing Gestures for Smart Rings Using Temporal Calculus. In *Proceedings of the ACM Conference Companion Publication on Designing Interactive Systems (DIS '18 Companion)*. ACM, New York, NY, USA, 117–122. https://doi.org/10.1145/3197391.3205422

[10] Bogdan-Florin Gheran, Santiago Villarreal-Narvaez, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2022. RepliGES and GEStory: Visual Tools for Systematizing and Consolidating Knowledge on User-Defined Gestures. In *Proceedings of the ACM International Conference on Advanced Visual Interfaces (AVI '22)*. ACM, New York, NY, USA, Article 5, 9 pages. https://doi.org/10.1145/3531073.3531112

[11] Steven N. Goodman, Daniele Fanelli, and John P.A. Ioannidis. 2016. What Does Research Reproducibility Mean? *Science Translational Medicine* 8, 341 (June 2016), 341ps12. http://dx.doi.org/10.1126/scitranslmed.aaf5027

[12] Kasper Hornbæk. 2015. We Must Be More Wrong in HCI Research. *Interactions* 22, 6 (oct 2015), 20–21. https://doi.org/10.1145/2833093

[13] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough? On the Extent and Content of Replications. In *CHI '14*. ACM, NY, USA, 3523–3532. https://doi.org/10.1145/2556288.2557004

[14] Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. 2013. Magic Ring: A Self-Contained Gesture Input Device on Finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 39, 4 pages. https://doi.org/10.1145/2541831.2541875

[15] Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical Research in Mid-Air Interaction: A Systematic Review. *Int. Journal of Human-Computer Interaction* 35, 18 (2019), 1747–1768. https://doi.org/10.1080/10447318.2019.1572352

[16] Dan Mauney, Jonathan Howarth, Andrew Wirtanen, and Miranda Capra. 2010. Cultural Similarities and Differences in User-defined Gestures for Touchscreen User Interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 4015–4020. https://doi.org/10.1145/1753846.1754095

[17] Michael J. Muller and Sarah Kuhn. 1993. Participatory Design. *Commun. ACM* 36, 6 (1993), 24–28. https://doi.org/10.1145/153571.255960

[18] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-Designed and User-Defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142

[19] Michael Nebeling, Alexander Huber, David Ott, and Moira C. Norrie. 2014. Web on the Wall Reloaded: Implementation, Replication and Refinement of User-Defined Interaction Sets. In *ITS '14*. ACM, New York, NY, USA, 15–24.

[20] Jakob Nielsen and Thomas K. Landauer. 1993. A Mathematical Model of the Finding of Usability Problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 206–213. https://doi.org/10.1145/169059.169166

[21] Keunwoo Park, Daehwa Kim, Seongkook Heo, and Geehyuk Lee. 2020. MagTouch: Robust Finger Identification for a Smartwatch Using a Magnet Ring and a Built-in Magnetometer. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376234

[22] Julie Rico and Stephen Brewster. 2010. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 887–896. https://doi.org/10.1145/1753326.1753458

[23] Roy Shilkrot, Jochen Huber, Jürgen Steimle, Suranga Nanayakkara, and Pattie Maes. 2015. Digital Digits: A Comprehensive Survey of Finger Augmentation Devices. *Comput. Surveys* 48, 2, Article 30 (2015), 29 pages. https://doi.org/10.1145/2828993

[24] Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4, Article 164 (Jan. 2018), 23 pages. https://doi.org/10.1145/3161416

[25] Eric W.K. Tsang and Kai-Man Kwan. 1999. Replication and Theory Development in Organizational Science: A Critical Realist Perspective. *The Academy of Management Review* 24, 4 (1999), 759–780. http://dx.doi.org/10.2307/259353

[26] Radu-Daniel Vatavu and Laura-Bianca Bilius. 2021. GestuRING: A Web-Based Tool for Designing Gesture Input with Rings, Ring-Like, and Ring-Ready Devices. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. ACM, New York, NY, USA, 710–723. https://doi.org/10.1145/3472749.3474780

[27] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[28] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2022. Clarifying Agreement Calculations and Analysis for End-User Elicitation Studies. *ACM Trans. Comput.-Hum. Interact.* 29, 1, Article 5 (2022), 70 pages. https://doi.org/10.1145/3476101

[29] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the ACM Designing Interactive Systems Conference (DIS '20)*. ACM, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[30] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2018. Gesture Elicitation Studies for Mid-Air Interaction: A Review. *Multimodal Technol. Interact.* 2, 4 (2018), 1–21. https://doi.org/10.3390/mti2040065

[31] Mathias Wilhelm, Jan-Peter Lechler, Daniel Krakowczyk, and Sahin Albayrak. 2020. Ring-Based Finger Tracking Using Capacitive Sensors and Long Short-Term Memory. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20)*. ACM, New York, NY, USA, 551–555. https://doi.org/10.1145/3377325.3377535

[32] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI Should Be Replicating and Validating Results More: Discuss. In *Proceedings of CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 463–466. https://doi.org/10.1145/1979742.1979491

[33] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *Proceedings of CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[34] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[35] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*. ACM, New York, NY, USA, Article 38, 10 pages. https://doi.org/10.1145/2601248.2601268

[36] Boning Zhang, Yiqiang Chen, Yueliang Qian, and Xiangdong Wang. 2011. A Ring-Shaped Interactive Device for Large Remote Display and Mobile Device Control. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 473–474. https://doi.org/10.1145/2030112.2030177

[37] Tengxiang Zhang, Xin Zeng, Yinshuai Zhang, Ke Sun, Yuntao Wang, and Yiqiang Chen. 2020. ThermalRing: Gesture and Tag Inputs Enabled by a Thermal Imaging Smart Ring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376323