“I Am Iron Man”

Priming Improves the Learnability and Memorability of User-Elicited Gestures

Abdullah X. Ali

The Information School | DUB Group, University of Washington, Seattle, WA, USA, xyleques@uw.edu

Meredith Ringel Morris

Microsoft Research, Redmond, WA, USA, merrie@microsoft.com

Jacob O. Wobbrock

The Information School | DUB Group, University of Washington, Seattle, WA, USA, wobbrock@uw.edu

ABSTRACT

Priming is used as a way of increasing the diversity of proposals in end-user elicitation studies, but priming has not been investigated thoroughly in this context. We conduct a distributed end-user elicitation study with 167 participants, which had three priming groups: a no-priming control group, sci-fi priming, and a creative mindset group. We evaluated the gestures proposed by these groups in a distributed learnability and memorability study with 18 participants. We found that the user-elicited gestures from the sci-fi group were significantly faster to learn, requiring an average of 1.22 viewings to learn compared to 1.60 viewings required to learn the control gestures, and 1.56 viewings to learn the gestures elicited from the creative mindset group. In addition, both primed gesture groups had higher memorability with 80% of the sci-fi-primed gestures and 73% of the creative mindset group gestures were recalled correctly after one week without practice compared to 43% of the control group gestures.

CCS CONCEPTS

• Human-centered computing → HCI design and evaluation methods.

KEYWORDS

Distributed Interaction Design; end-user elicitation study; end-user identification; learnability; memorability; crowdsourcing; Mechanical Turk.

1 Introduction

Including end-users in the design stage of interactive systems is a common practice in human-computer interaction (HCI). User-centered design [1] practices—*e.g.*, participatory design [34]—include potential users of a system in the design phase to inform designers of the capabilities, expectations, and preferences of the users themselves. End-user elicitation studies—as formalized by Wobbrock *et al.* [51,53]—are a popular example of this practice. The HCI literature includes more than 200 published end-user elicitation studies designing interactions with robots [36], drones [8], vehicles [10], bicycles [54], smart TVs [3,30], and involving diverse user populations like children [9] and individuals with disabilities [20] in the design of interactive systems. The method works by prompting end-user participants with the result of a computing function and asking them to propose the action or input that would trigger that function in an interactive system. Numerous published articles put forth best practices and extensions to end-user elicitation studies such as new mathematical formulae to calculate agreement among participants' proposals [12,30,47,48], methods focusing on stimulating participants' creativity [31], translating the methodology online to reach a wider, more-diverse pool of participants [4], formalizing evaluative methods [4], and utilizing online crowds and machine learning algorithms to efficiently analyze the results of elicitation studies [3].

The first contribution of this paper focuses on extending the work of Morris *et al.* [31] in which they point out a pitfall of elicitation studies they call *legacy bias*. This pitfall, as they define it, is when “participants in elicitation studies propose familiar interactions from current technologies that might not be best suited for new technologies' form factors or sensing capabilities.” Morris *et al.* [31] proposed principles to reduce legacy bias in elicitation studies. One of their principles is *priming*, a practice from the field of psychology used to enhance creative thinking. The effects of priming in elicitation studies have been explored a little in prior work. One example, Hoff *et al.* [17] found that priming results in fewer legacy gestures and quicker generation of ideas; however, their results were not statistically significant and they stated that given the typical small number of participants in (traditionally lab-based) elicitation studies, they do not recommend the use of priming.

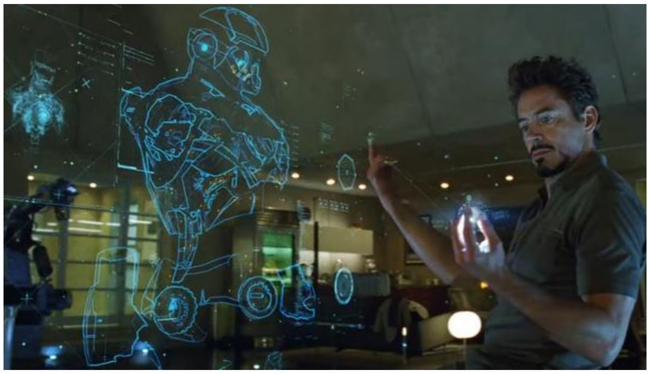

Figure 1. A still from the movie Iron Man 2 [19], showing the main character Tony Stark interacting with an augmented reality hologram interface with a hand gesture.

In this paper, we utilize on Ali *et al.'s* [4] methods and tools to run distributed elicitation studies efficiently to explore the effects of priming with a large number of participants. For our exploration, we used mixed reality (MR) environments as our use case due to their novelty as a technology for end-users. These environments are on the cusp of becoming a mainstream technology but have yet to be widely adopted by the average technology user. We used priming to push beyond legacy interactions informed by desktop or mobile computing and elicit interactions for an MR environment that are easily learnable and memorable. We employed priming in two ways: (1) By having participants view a montage of sci-fi films depicting characters interacting with technologies using gestures. For this primer, we drew inspiration from the bi-directional relationship between sci-fi and HCI [44]. (2) We looked to the field of psychology and followed Sassenberg and Moskowitz's [41] practice of using a "creative mindset" to suppress stereotyping and "think different."

We conducted a between-subjects elicitation study with priming as our independent variable. We recruited 167 participants—more than eight times the number of participants in a typical lab-based study (~20) [50]—from Amazon's Mechanical Turk (mTurk) platform. We randomly assigned participants to one of three groups: control (no priming), sci-fi priming, and creative-mindset priming. Participants were prompted with 10 functions of a media player in an MR environment and were asked to propose a mid-air gesture to trigger each of the 10 functions. As a result, we formulated three sets of user-elicited gestures. We found overlap in four gestures across all three groups. The sci-fi gesture set had only five gestures in common with each of the other two gesture sets. On the other hand, seven of the ten gestures in the creative mindset set were identical to the control. Despite the overlap in gestures across the three gesture sets, the sets had some different results when we tested their usability in terms of gesture identifiability, learnability and memorability.

Per the recommendation of Ali *et al.* [4], we followed our elicitation study with a distributed end-user identification study—the reverse of an elicitation study—with 50 new participants from mTurk. The aim of an identification study is to assess the discoverability of the gesture-function relationship. Our identification study showed that priming had no statistically significant impact on the number of correctly identified gesture-function relationships of user-elicited gestures.

The second major contribution of this paper centers around distributed interaction evaluation studies. We capitalized on Ali *et al.'s* Crowdlicit system [4] to run supervised distributed interaction-evaluation studies. Extending their work evaluating interaction-identifiability, we operationalize two new aspects of interaction usability: learnability and memorability. We recruited 18 new online participants for a two-part study to evaluate our three gesture sets. In the first part of this study, participants were randomly assigned to view one of the three user-proposed gesture sets (control, sci-fi, creative mindset). After viewing a video clip of each gesture in the set once, participants were prompted with a function and asked to perform the corresponding gesture they had just learned. After going through all 10 functions, participants were allowed to go back and view the video clips of the gestures that they got wrong, if any. We repeated this process until the participants learned and were able to correctly perform all 10 gestures in their set. Furthermore, after one week, we contacted the same participants and asked them to go through the testing protocol only once to assess the memorability of the gestures—without allowing them to view the video clips of the gestures. We found that sci-fi-primed gestures were faster to learn, as they required an average of 1.22 viewings to learn. Non-primed gestures required an average of 1.60 views and the creative mindset primed gestures required 1.56 viewings to learn. After a single viewing, 80% of the sci-fi primed gestures were learned compared to 65% gestures from the control and 58% creative mindset sets—sci-fi gestures were learned most quickly. Additionally, the primed gestures had a higher memorability accuracy compared to the control gesture set with 80% of the sci-fi gestures and 73% of the creative mindset recalled correctly compared to 43% for the control group.

This work contributes: (1) an empirical study of the effects that *priming* has in elicitation studies, (2) methods to evaluate the learnability and memorability of interaction designs in a distributed manner, and (3) a user-elicited gesture set for a media player in a mixed reality environment.

2 Related Work

Relevant prior work includes articles discussing the relationship between sci-fi and HCI, examples of end-user elicitation studies, and work on the use of priming to enhance creativity.

2.1 Sci-fi and HCI

Science fiction movies such as *Iron Man*—the namesake of this paper—depict fascinating human-computer interactions. It is no wonder there exists a bidirectional relationship between sci-fi and HCI, where each endeavor, at its best, showcases brilliant new possibilities for people's use of technology. Schmitz *et al.* [43], in their survey of HCI in sci-fi movies, reported that there exists a collaboration between filmmakers and scientists regarding the use of HCI in film. They mentioned that director Steven Spielberg consulted with HCI researchers to develop the system and interactions shown in his movie *Minority Report*. Larson [23] stated that sci-fi depictions of technologies mirror trends in real life computing. Aaron Marcus, in his article "The History of the Future: Sci-Fi Movies and HCI" [29], stated that sci-fi movies can be a useful material to inform designers of possible future technological, social, or cultural contexts. Mubin *et al.* [33] cited many examples of devices and products that have roots that can be traced back to sci-fi movies, like mobile phones inspired by the communicators from *Star Trek*, and video conferencing similar to that depicted in *2001: A Space Odyssey*. The recurring "Future Tense" section of *Communications of the ACM* often features sci-fi writers like David Brin who, in the words of the magazine, "present stories and essays from the intersection of computational science and technological speculation" (*e.g.,* [7]). Not incidentally, David Brin also gave the ACM CHI 2002 keynote address, drawing on themes explored in his sci-fi writings to inspire an audience of HCI researchers. The work presented in this paper contributes to this body of literature by investigating the effects of using sci-fi as a primer to influence the design of future interactions.

2.2 Beyond End-User Elicitation Studies

Hundreds of end-user elicitation studies have been conducted and published since 2009. A recent survey by Villarreal-Narvaez *et al.* [49] shows that at least 216 gesture elicitation studies have been published as of 2020. End-user elicitation studies have been used to design interactions for emerging technologies like robots [36,39], drones [8,37], and virtual and augmented reality [18,38], among many other things. Going further, many elicitation studies have gone beyond reporting a set of user-elicited interactions and have actually tested the usability of these interactions. For example Morris *et al.* [32] examined users' preferences for gestures and found that gestures elicited by a large group of people were preferable to those authored by one or two designers. Ali *et al.* [4] formalized end-user identification studies as a reversed companion to elicitation studies with the aim of assessing the discoverability of the action-function relationship. Nacenta *et al.* [35] tested the memorability of two groups of gestures and found that user-elicited gestures are more memorable, easier, more fun, and less effortful. Unlike Nacenta *et al.'s* [35] work, we, in this paper, test the identifiability, learnability, and memorability of our user-elicited gestures with new populations and do not use the same participants who elicited the gestures themselves. Our work adds to the end-user elicitation study literature by conducting an elicitation study to design gestures for a media player in a mixed reality environment. We also report, to the best of our knowledge, the first distributed memorability and learnability study on user-elicited interactions.

2.3 Priming and Legacy Bias

Priming is a concept with deep roots in the field of psychology dating to 1951 [24]. Sassenberg *et al.* [42] used priming as a strategy to increase creative performance. They stated that a creative mindset undermines the copying of existing ideas. Sassenberg and Moskowitz [41] also used priming to suppress stereotyping by asking participants to recall examples of a time when they felt creative—an approach we utilize in this paper. Elicitation studies have borrowed priming from psychology. Morris *et al.* [31] proposed using priming as a way to reduce *legacy bias*, which they define as a potential pitfall in elicitation studies, saying "users' proposals are often biased by their experience with prior interfaces and technologies." That said, the elicitation literature remains divided on the usefulness of reducing legacy bias. Köpsel and Bubalo [21] presented a counterpoint to Morris *et al.*, arguing that legacy bias helps create good interactions. Citing small participant numbers, Köpsel and Bubalo argue that the non-legacy interactions will not generalize to a wide user base. They claim that legacy gestures, for example, are simpler. However, the concern with legacy interactions is their potential to limit users from taking full advantage of emerging applications, form factors, and sensing capabilities; this is the reason we focus our investigation of the usefulness of priming in an MR application. Hoff *et al.* [17] experimentally tested the effects of priming and found that it had a only a small effect on their user-elicited gestures. Hoff *et al.* reported that their study had only 30 participants—as is common in lab-based elicitation studies—and mentioned that more work with a larger pool of participants is needed to validate their findings. Beşevli *et al.* [6] tested legacy and non-legacy gestures for their memorability, situational fit, and physical ease, using self-reported Likert-type ratings. Their results showed legacy gestures to have better scores; on the other hand, users favored non-legacy gestures due to their practicality. Connell *et al.'s* [9] work with children showed legacy bias effects on proposals, as children familiar with some touchscreen interfaces proposed whole-body navigational gestures influenced by their experience with touchscreens, while children with no experience with such interfaces gave rise to a greater variety of gestures. Ruiz and Vogel [40] showed that physical priming such as constraining participants movements when eliciting whole-body gestures encouraged a wider range of proposed gestures.

The literature has not offered strong evidence to neither discard legacy-biased interactions nor to implement them confidently. We believe that in some situations a legacy interaction might be preferable, but when designing emerging interfaces with new sensing capabilities and novel interaction possibilities, reducing legacy bias remains important. In this work, we focus our efforts on investigating the effects that priming—by either viewing sci-fi movies or having a creative mindset—has on the results of an elicitation study to create gestural interactions for an MR environment. We also contribute, to the best of our knowledge, the largest (in terms of the number of participants) investigation on priming in the literature of elicitation studies.

3 The Effects of Priming on User-Elicited Gestures

To evaluate the effects of priming on the learnability and memorability of user-elicited gestures, we first conducted a distributed elicitation study accompanied by an identification study. The elicitation study resulted in three gesture sets—one for each level of priming (control, sci-fi, and creative mindset). The identification study subsequently evaluated how guessable the gestures are in each priming group.

3.1 Creating a User-Elicited Gesture Set

We conducted a between-subjects distributed elicitation study with 167 online participants using the Crowdlicit platform [4] to design gestures for a media player in an MR environment.

3.1.1 Participants

We recruited 167 participants in total using Crowdlicit [4] from Amazon Mechanical Turk (mTurk), to provide video recordings of their proposed gestures in response to 10 prompts showing functions of a futuristic media player. We followed our elicitation study with a distributed identification study, as recommended by Ali *et al.* [4]. To do so, we recruited 50 new participants from mTurk. In both studies, each participant filled out a demographic survey upon completing the study. Table 1 shows the demographic information for both studies. Participants needed to have a device with a camera (*i.e.*, a webcam or use a smart phone) to participate in the elicitation study. Participants in the elicitation study were compensated $7.50 for participation in the half-hour study. Participants in the 15-minute identification study were compensated $3.75. We based our compensation on our state's $15/hour minimum wage rate.

Demographic		*Elicitation (N=84)*	*Identification (N=33)*
Gender	Man	59 (70.24%)	20 (60.60%)
	Woman	24 (28.57%)	13 (39.39%)
	Non-binary	1 (1.19%)	0
Age	Mean (SD)	30.69 (5.61)	38.03 (10.39)
Nationality	USA	71 (84.52%)	24 (72.73%)
	India	7 (8.33 %)	4 (12.12%)
	Brazil	3 (3.57%)	3 (9.09%)
	Canada	1 (1.19%)	1 (3.03%)
	Germany	1 (1.19%)	0
	Pakistan	1 (1.19%)	0
	Italy	0	1 (3.03%)
Do you own an MR device?	Yes	15 (17.86%)	9 (27.27%)
	No	69 (82.14%)	24 (72.73%)
How often do you use an MR device?	Never	29 (34.52%)	14 (42.42%)
	Daily	2 (2.38%)	4 (12.12%)
	Monthly	20 (23.81%)	5 (15.15%)
	Once or twice ever	33 (39.29%)	10 (30.30%)
Do you use mid-air gestures?	Yes	7 (8.33%)	32 (96.97%)
	No	77 (91.66%)	1 (3.03%)
Favorite movie genre	Comedy	19 (22.62%)	7 (21.21%)
	Action & adventure	15 (17.86%)	5 (15.15%)
	Drama	5 (5.95%)	4 (12.12%)
	Horror	6 (7.14%)	1 (3.03%)
	Sci-fi	18 (21.43%)	9 (27.27%)
	Documentary	7 (8.33%)	2 (6.06%)
	Thriller	10 (11.90%)	2 (6.06%)
	Epic/ historical	2 (2.38%)	1 (3.03%)
	Musicals	2 (2.38%)	1 (3.03%)
	Western	0	1 (3.03%)
Have you seen this movie before?	Minority Report	39 (46.43%)	19 (57.57%)

Demographic		*Elicitation (N=84)*	*Identification (N=33)*
	Iron Man 2	62 (73.81%)	28 (84.85%)
	Black Mirror	43 (51.19%)	16 (48.48%)
	Gamer	11 (13.10%)	4 (12.12%)
	Enders Game	31 (36.91%)	10 (30.30%)

Table 1. Participants' demographic information from our elicitation study with 167 participants (84 filled out demographics survey) and identification study with 50 participants (33 filled out the survey).

3.1.2 Apparatus

We used Ali *et al.'s* Crowdlicit system [4] to run our distributed elicitation and identification studies. We created video clips—included as in the supplementary material of this paper—of 10 functions to interact with a media player in an MR environment. Our 167 online participants viewed these video clips as prompts to propose gestures that would trigger the functions shown in the videos. Participants recorded their gesture proposals using the web interface and their personal computer's webcam or via the camera on their mobile device, then uploaded the footage to Crowdlicit. This process resulted in 15 unique gestures across the three priming groups.

We then used Crowdlicit again in our identification study with a set of 50 new online participants. The identification study showed the 15 elicited gestures as video clips as prompts. For each gesture, participants entered text descriptions of the functions they anticipated the gesture would trigger in an MR media player. The Crowdlicit system allowed us to capture self-reported Likert-type ratings after every gesture or function proposal. In addition, the system also captured the time participants needed to submit a proposal (in seconds).

To prime participants in the elicitation study who were assigned to the sci-fi priming condition, we created a short montage film from movies and TV shows like *Iron Man*, *Minority Report*, and *Black Mirror*. We created this montage film using the open catalog of gestures in sci-fi movies by Figueiredo *et al.* [11]. Their catalog[1] has tags of what task is being performed in the clip (*e.g.,* play, previous, etc.). We include these clips in the supplementary material of this paper. The clips are organized by the tag assigned to them in the Figueiredo *et al.* [11] catalog.

3.1.3 Procedure

Following Wobbrock *et al.*'s [51,53] method, we conducted a distributed elicitation study and followed it up with a distributed identification study as prescribed by Ali *et al.* [4] to produce our three user-generated gesture sets for the three priming groups.

3.1.3.1 Distributed End-User Elicitation

The 167 participants who accepted the human intelligence task (HIT) on mTurk were directed to a custom webpage that randomly assigned them to one of three priming groups (control, sci-fi, and creative mindset). Based on the assignment, participants were automatically directed to a Crowdlicit study URL. This setup allowed us to organize the elicited gestures into three groups (control, sci-fi, creative mindset). Upon navigating to the unique Crowdlicit study link, participants were presented with instructions for the study explaining that they were about to watch 10 video clips of functions for a media player in a mixed reality (MR) environment (Table 2) and that such a system responds to mid-air gestures. The participants were required to propose a gesture of their choosing to trigger each of the 10 functions depicted in the video clips. The instructions also showed a diagram instructing the participants on how to position themselves in front of the camera in such a way that would exclude their face from showing in the recording to protect their privacy.

The participants in the creative mindset group were asked to provide three examples of a time they were creative before starting the elicitation session. We borrowed this technique from Sassenberg and Moskowitz [41].

All three variations (control, sci-fi, and creative mindset) of the elicitation study on Crowdlicit were identical except for the instructions section for the sci-fi group and a pre-session task for the creative mindset group. The instructions for the sci-fi group included a section stating: "The environment should respond to gestures like, but not limited to, ones shown in this video." Below that message was a montage film of sci-fi clips[2]. We chose one clip from a tag that represents each function in our list of functions (Table 2) and compiled them into the montage film.

No.	*Function*
1.	Play a video
2.	Pause a video
3.	Fast forward
4.	Rewind
5.	Next video
6.	Previous video
7.	Close a video
8.	Pin view to a surface
9.	Bring view into field of vision

1 goo.gl/XSX5fn

2 The clips are included in the supplementary material of this paper.

10. Activate headtracking

Table 2. A list of 10 functions to control a media player in a mixed-reality (MR) environment. "View" refers to the video element. The last three functions are specific to an augmented reality environment.

3.1.3.2 Distributed End-User Identification

Whereas the elicitation study enabled us to show 10 MR functions and elicit 15 unique gestures to trigger them, an identification study enabled us to test how guessable those 15 gestures were. In essence, an identification study reverses the elicitation study. The 50 new participants who accepted the HIT on mTurk for the identification study were redirected to a custom study webpage created by the Crowdlicit system. The identification study had 15 unique gesture prompts. These gestures were the results of our elicitation study, which resulted in 15 unique gestures for the 10 functions with overlap across the three priming levels. Figure 2 shows these gestures and which ones overlap. Each prompt in the identification study showed a video of a person—the first author of this paper—performing one of the 15 gestures that resulted from our elicitation study. After viewing a video of a gesture, the participants were asked to propose the function they thought a system would trigger in the context of interacting with an MR video player.

3.1.4 Design and Analysis

The elicitation study was a single-factor between-subjects design, whose factor was *Priming*, which had three levels: no priming (control), sci-fi priming, and creative-mindset priming. In this study, we collected 1,167 gesture proposals from a total of 167 online participants. Due to a server error, 12 gestures were not recorded from the control group, leaving the control group with a total of 381 gestures; sci-fi and creative-mindset groups each had 393 gestures.

The identification study was a single-factor within-subjects design, with the same priming factor and levels as the elicitation study. In this study, we collected 750 function proposals from our 50 online participants. As a within-subjects study, the identification study showed each participant gestures from all three priming groups.

We investigated the effect of *Priming* on the results of the distributed elicitation and identification studies on three dependent variables: (1) Agreement scores for proposed gestures and functions calculated using Equation 1, below. (2) Self-reported Likert-type satisfaction ratings (ease, match, enjoyment). (3) Proposal time—*i.e.*, time it took participants to come up with a gesture or a function. In addition, we explored subjective differences across the three user-elicited gesture sets that we created from our elicitation study. We also compared identification accuracy across the three gesture sets in our identification study.

To calculate a gesture agreement score, we used Wobbrock *et al.'s* [51,53] original Equation 1:

Equation 1 $\boldsymbol{A} = \frac{\boldsymbol{1}}{|\boldsymbol{R}|} \sum_{\boldsymbol{r} \in \boldsymbol{R}} \sum_{\boldsymbol{P_i} \subseteq \boldsymbol{P_r}} \left(\frac{|\boldsymbol{P_i}|}{||\boldsymbol{P_r}|}\right)^{\boldsymbol{2}}$

In Equation 1, r is a prompt in the set of all prompts R. P_r is the set of all proposals for a given prompt r. P_i is the subset of similar proposals in P_r. Wobbrock *et al.'s* [51,53] Equation 1 was updated in some subsequent publications [12,47,48]; however, because we collected a single gesture proposal per prompt in the same manner as Wobbrock *et al.'s* original paper, we opted to use the original equation in our analysis. Furthermore, Ali *et. al* [4] used that equation to calculate function agreement in identification studies. To provide a sense of uniformity for our analysis throughout this paper, we decided that Equation 1 was the best fit for our analysis. Agreement scores have an upper limit of 1.0. The upper limit represents total agreement in which all the proposals collected in response to a prompt match each other.

Because agreement scores are bounded, we used a non-parametric Kruskal-Wallis test [22] to assess the differences in agreement scores among the three levels of *Priming*. We followed a significant omnibus test with *post hoc* comparisons using a pairwise Mann-Whitney *U* test [28], corrected with a Tukey HSD test [5]. To investigate differences in the ordinal Likert-type self-reported ease, match, and enjoyment ratings, we used mixed ordinal logistic regression [2,15] [cite]. Then, we conducted *post hoc* analyses on any significant omnibus tests using a Tukey HSD test [5]. To investigate the effect of *Priming* on the time it took participants to propose a gesture or a function, we used a linear mixed model analysis of variance [13,26]. We log-transformed the time response prior to analysis, as is common [25], to comply with the assumption of conditional normality. We followed up any significant omnibus test with a *post hoc* Tukey HSD test [5].

3.2 Results

We identified the gesture with highest agreement for each one of the 10 functions in each of the priming groups (control, sci-fi, and creative mindset). If all gestures were unique, this would result in 10 × 3 = 30 gestures, but due to substantial overlap among the groups, there were 15 unique gestures in all. These gestures are shown in**Error! Reference source not found.**, and served as prompts in our subsequent identification study.

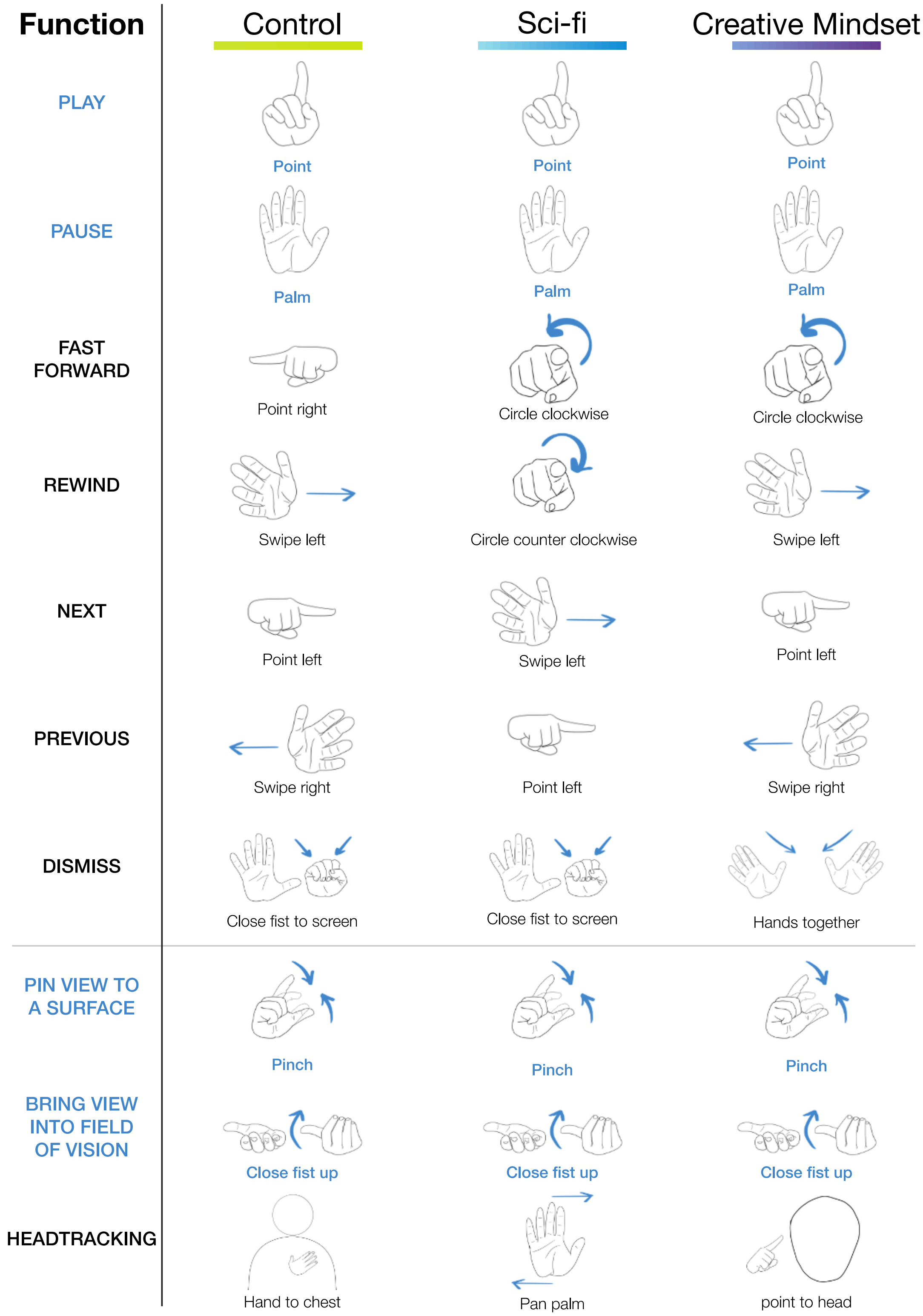

Figure 2. Three gesture sets for 10 functions containing a total of 15 unique gestures. Functions in blue show that the same gesture among the three sets was proposed by the majority of participants in that group. The last three functions are specific to an augmented reality environment

3.2.1.1 User-Elicited Gesture Sets

We compared gestures for each function across the three sets for their similarity and found that for four functions (play, pause, pin view to a surface, and bring view into field of vision) had the exact same gestures across the three gesture sets (point, palm, pinch, close fist up, respectively). The sci-fi gesture set had five gestures in common with the control set that triggered the same function. The creative-mindset set had seven in common with the control set, leaving only three functions triggered by a different gesture (fast forward, dismiss and headtracking). The sci-fi and creative mindset gesture sets had five gestures in common (the four gestures found across all three gesture sets and the gesture circle clockwise to trigger the fast forward function).

We compared the gestures in the sci-fi group against the gestures appearing in our sci-fi montage video that we used as a primer to see if any of the gestures in the video were replicated by participants. The circle finger clockwise and counterclockwise gestures were present in the montage clip (*Black Mirror* clip under the rewind tag) and so was the "pan palm to screen" gesture (*Enders Game* clip under the play tag). Other gestures from the montage like close fist up, and swipe from the clip were present throughout all three gesture sets and not limited to the sci-fi group.

3.2.1.2 Agreement Scores

For the control level of *Priming,* the mean agreement score—the degree to which the participants agreed on a gesture to trigger a function—was A = .182 (SD = .077); for sci-fi, it was A = .178 (SD = .088); and for creative mindset it was A = .186 (SD = .084). A Kruskal-Wallis test found no statistical significance among these scores ($\chi^2(2, N{=}30) = 0.267$, *n.s.*).

For the identification study, the function agreement scores were also similar across *Priming* levels. For control the mean agreement was A = .170 (SD = .094); for sci-fi, it was A = .191 (SD = .086); for creative mindset it was A = .173 (SD = .067). A Kruskal-Wallis test found no statistical significance in the differences among these scores ($\chi^2(2, N{=}30) = 0.937$, *n.s.*).

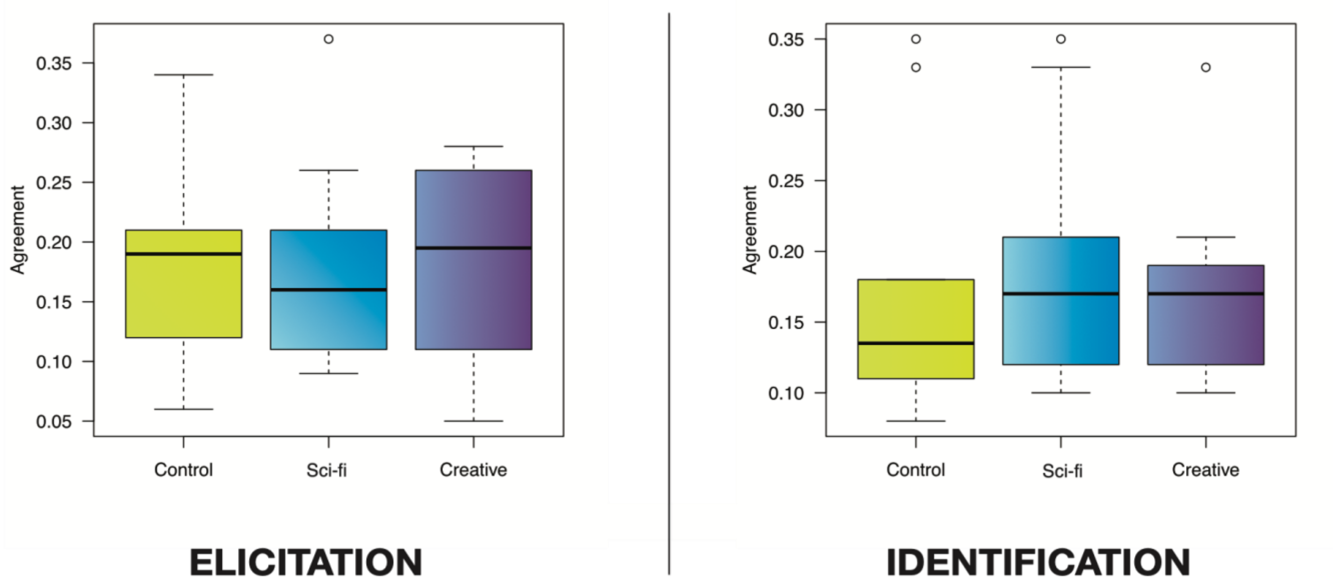

Figure 3. Agreement scores for three gesture sets created under three *Priming* levels (control, sci-fi, and creative mindset)

3.2.1.3 Priming Effect on Self-Reported Ratings

We collected Likert-type ratings on a scale of 1–7 (1. Strongly disagree, 7. Strongly agree). The scales assessed the following statements: 1. **Ease**, *my proposal is easy to perform*. 2. **Match**, *my proposal is a good match for its intended purpose.* 3. **Pleasure**, *my proposal is enjoyable to perform.* Admittedly, Likert-type ratings are ordinal in nature and their numeric markings (1–7) cannot be taken as scalar responses. That said, for illustrative purposes and for the reader's appreciation, we report the means and standard deviations of the Likert-type ratings in Table 3. The self-reported ease, match, and pleasure scores from the 167 participants in the elicitation study and 50 participants in the identification study. Higher numbers mean "easier," "better matches," and "more enjoyable," respectively. *Bold font indicates statistically significant differencein addition to the median scores and interquartile ranges.

	Priming	Rating	Median (IQR)	Mean (SD)	Priming	Rating	Median (IQR)	Mean (SD)	
ELICITATION	Control	Ease	6 (3)	5.12 (2.47)	Control	Ease	7 (1)	6.07 (1.25)	IDENTIFICATION
	Sci-fi		6 (1)	6.06 (1.19)	Sci-fi		6 (2)	6.00 (1.27)	
	Creative mindset		7 (1)	5.69 (2.21)	Creative mindset		6 (2)	6.02 (1.27)	
	Control	Match	6 (3)	4.97 (2.43)	Control	Match	6 (2)	5.30 (2.15)	
	Sci-fi		6 (2)	5.71 (1.29)	Sci-fi		6 (2)	5.30 (2.13)	
	Creative mindset		6 (2)	5.48 (2.20	**Creative mindset***		**6 (2)**	**5.87 (1.27)**	
	Control	Pleasure	5 (3)	4.62 (2.34)	Control	Pleasure	5 (3)	4.84 (2.11)	
	Sci-fi		6 (3)	5.48 (1.31).	Sci-fi		5 (3)	4.86 (2.09)	
	Creative mindset		6 (3)	5.08 (2.16)	**Creative mindset***		**6 (3)**	**5.36 (1.40)**	

Table 3. The self-reported ease, match, and pleasure scores from the 167 participants in the elicitation study and 50 participants in the identification study. Higher numbers mean "easier," "better matches," and "more enjoyable," respectively. *Bold font indicates statistically significant differences using mixed ordinal logistic regression [2,15] ($p < .05$).

The numeric **ease** ratings were higher, on average, for sci-fi (6.06) compared to control (5.12) and creative mindset (5.69) as reported by the 167 participants in our elicitation study. On the other hand, the ease ratings were nearly identical across all three levels as reported by the 50 participants in our identification study. An analysis of variance based on mixed ordinal logistic regression indicated no statistically significant effect on ease ratings of *Priming* in either the elicitation study ($\chi^2(2, N{=}1167) = 4.41$, *n.s.*) or identification study ($\chi^2(2, N{=}750) = 3.22$, *n.s.*).

The numeric **match** ratings for sci-fi (5.71) were also higher, on average, than control (4.97) and creative mindset (5.48), according to the participants who proposed the gestures in the elicitation study. However, sci-fi (5.30) and control (5.30) match scores were similar and lower than creative mindset (5.87) as rated by the participants who attempted to identify the function associated with the gesture in our identification study. These differences were not detectably different in the elicitation study ($\chi^2(2, N{=}1167) = 1.98$, *n.s.*). However, the creative mindset match ratings in the identification study were significantly higher than the other priming groups ($\chi^2(2, N{=}750) = 8.60$, $p < .05$). *Post hoc* pairwise comparisons using Tukey's HSD correction indicated that creative mindset *vs.* control ($Z = 2.44$, $p < .05$) and creative mindset *vs.* sci-fi ($Z = -2.60$, $p < .05$) were statistically significant. The sci-fi *vs.* control match ratings were not detectably different ($Z = -0.14$, *n.s.*).

Finally, the numeric **pleasure** ratings had a similar outcome to the match ratings with sci-fi gestures (5.48) rated higher than control (4.62) and creative mindset (5.08) gestures in the elicitation study, showing no detectable differences ($\chi^2(2, N{=}1167) = 3.42$, *n.s.*). In the identification study, the creative mindset gestures (5.36) had significantly higher pleasure ratings ($\chi^2(2, N{=}750) = 8.67$, $p < .05$) than the almost identical ratings of sci-fi (4.86) ($Z = -2.457$, $p < .05$) and control gestures (4.84) ($Z = 2.60$, $p < .05$).

3.2.1.4 Priming Effect on Elicitation Time

The mean time needed by participants to provide a gesture in the control group was the highest at 58.3 seconds ($SD = 59.2$). The creative-mindset group was second at an average 49.3 seconds ($SD = 41.8$), and the sci-fi group had the fastest elicitation time with an average 46.5 seconds ($SD = 46.1$). Despite the effect of priming on lowering the mean elicitation time, a linear mixed-effects model analysis of variance indicated no statistical significance in time differences ($F(2, 123.81) = 1.617$, *n.s.*).

As for the time needed to identify the function associated with a gesture in the identification study, the results were very close across all three priming groups—control: 38.0 seconds ($SD = 34.6$); sci-fi: 38.7 ($SD = 41.7$); creative mindset: 39.4 ($SD = 39.0$). There were no statistically significant differences among these results ($F(2, 14.973) = 0.70$, *n.s.*).

3.2.1.5 Priming Effect on Identifiability

The percentage of correctly identified gestures did not differ much across the three levels of *Priming*. The control gesture set had 22.2% ($SD = 41.6\%$) of its gestures identified correctly. The sci-fi gesture set had 24.2% ($SD = 42.8\%$) identified correctly, and the creative-mindset gesture set had 22.2% ($SD = 41.47\%$) of its gestures identified correctly by the 50 participants in the distributed identification study. Priming had no statistically significant effect on the count of correctly identified gestures across the three levels, according to an analysis of variance based on mixed logistic regression [14] ($\chi^2(2, N{=}750) = 3.417$, *n.s.*).

4 Learnability and Memorability of Elicited Gestures

Priming had some effect on the elicited gestures and their identifiability, but of crucial importance to system designers is how *learnable* a set of gestures is—and once learned, how *memorable* those gestures are. In this phase of our investigation, we sought to see whether priming affects gesture learnability and memorability. Accordingly, we conducted a two-part supervised distributed study again using the Crowdlicit platform [4] to evaluate the learnability and memorability of the three gesture sets we created as a result of our distributed elicitation study.

4.1 Participants

We recruited 18 new participants using convivence and snowball sampling by advertising our study on our university's Slack channels, and on social media platforms. Two of our participants failed to complete the demographics survey. Of our 16 participants who did complete the demographics survey, nine were female, six were male, and one non-binary. The mean age was 27.3 years (SD = 4.84). The participants' nationalities were mostly from the United Stated (10/16), other nationalities included India, China, and Kazakhstan. Seven participants had never used an MR device, and five had only used one once or twice. Two participants used an MR device on a monthly basis and two others used one on a daily basis. As for participants' use of mid-air gestures to interact with technologies, only two participants reported having used mid-air gestures to interact with a desktop music app and an Xbox Kinect.

4.2 Apparatus

We used the Crowdlicit system [4], once again utilizing its web-based video recording capabilities to collect, store, and organize the data in our study. We used Google Meet to video-call our participants and guide them through the procedures of the studies. For the first part of the study, which was devoted to learnability, we created a custom learning website for participants, which came in three versions corresponding to the three gesture sets we created as a result of our elicitation study (control, sci-fi, and creative mindset). In each version, the website displayed 10 videos, each depicting a gesture from the corresponding set.

4.3 Procedure

Once the first author connected with a participant over Google Meet, he asked them to share their screen and directed them to the custom learning site. The page displayed all 10 videos of one of the three gesture sets shown in Figure 2 in a random order. Participants were asked to view each video once and then navigate via a Crowdlicit-generated link to perform all 10 gestures. We used the Crowdlicit system to prompt participants with a function and asked them to record a video of themselves performing the gesture corresponding to that function, which they had just learned on the page with 10 videos. The first author monitored the video recordings as they were being uploaded to the Crowdlicit system and assessed their correctness. The first author then refreshed the custom learning page, removing the videos of the correctly performed gestures, leaving only videos of the gestures that the participant missed. Participants repeated the learning and performing process until they learned all 10 gestures successfully. How many times the videos had to be viewed before all 10 gestures were learned successfully was taken as a measure of gesture learnability.

To assess memorability, all 18 participants from the learnability study were informed they would be contacted via Google Meet one week after the learnability session to perform the gestures again using the Crowdlicit system. Participants did not have access to the learning page to practice the videos in the interim period.

4.4 Design and Analysis

Both the learnability and memorability studies used single-factor between-subjects designs, with a factor for *Priming* having three levels: no priming (control), sci-fi priming, and creative mindset priming. We collected a total of 263 gesture trials from our 18 online participants. The learnability study had the following dependent variables: (1) initial learnability—the number of gestures learned after a single viewing of all 10 gesture videos; (2) overall learnability—the total video views needed to learn a gesture; and (3) learned-gesture performance time.

We used mixed logistic regression [14] to analyze the dichotomous results of the first trial, *i.e.*, whether a gesture was performed correctly or incorrectly after viewing each of the 10 gesture videos once. We carried out *post hoc* testing using Tukey's HSD test [45,46] for multiple comparisons. For the total number of views required to learn a gesture, we used an interaction contrast [16,52], corrected with Tukey's HSD correction [45,46]. Again, *post hoc* pairwise comparisons were conducted using Tukey's HSD correction. For analyzing gesture performance time, we used the same analysis approach as in the elicitation study.

In the memorability study, we collected $18 \times 10 = 180$ gesture trials from our 18 online participants. We used the same analysis approach from the learnability study to evaluate two dependent variables: (1) the number of correctly recalled gestures, analyzed in the same manner as initial learnability; (2) the time to recall and perform a gesture analyzed in the same manner as the learned-gesture performance time.

4.5 Results

We evaluated the effects *Priming* had on the learnability of user-elicited gestures in terms of initial learnability, overall learnability, and learned-gesture performance time. For the effects of priming on memorability, we evaluated the gestures based on the number of correctly remembered gestures and gesture-recall time.

4.5.1 Initial Learnability

Initial learnability is the number of gestures performed correctly (out of 10) after one viewing each of the 10 gesture videos in each priming set. The number of learned gestures after a single viewing for the control set was 39 gestures (out of 60 total gestures, or 65%); for the sci-fi set it was 48 gestures (80%); and for the creative mindset set it was 35 (58%). These differences were marginally significant for *Priming*'s overall effect on initial learnability ($\chi^2(1, N=180) = 5.75$, $p = .056$). *Post hoc* pairwise comparisons using Tukey's HSD correction indicated that creative mindset gestures were significantly less learnable initially than sci-fi gestures (Z = -2.36, $p < .05$).

4.5.2 Overall Learnability

Overall learnability is the total number of gesture video views required to learn all 10 gestures from a giving priming group. For the control gesture set, the mean count of viewings required to learn a gesture was 1.60 (*SD* = 1.06); for the sci-fi gesture set it was 1.22 (*SD* = 0.45); and for the creative mindset gesture set it was 1.56 (*SD* = 0.87). *Priming* had a statistically significant effect on overall learnability ($F(2, 38.5)= 3.35$, $p < .05$). Post hoc pairwise comparisons using Tukey's HSD correction showed that the difference between sci-fi and creative mindset was marginally significance ($t(267) = -2.40$, $p = .062$). Specifically, sci-fi gestures seemed easier to learn than creative mindset gestures.

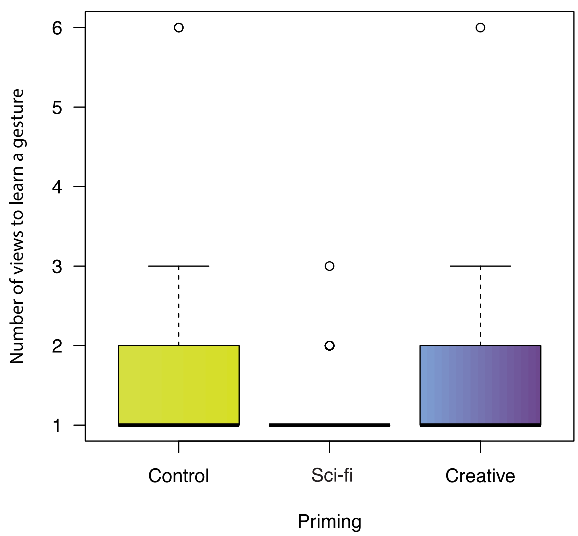

Figure 4 A boxplot of the number of viewings required to learn a gesture by Priming.

4.5.3 Memorability

For each level of *Priming* (control, sci-fi, and creative mindset), we assessed the memorability of the gesture set—*i.e.*, whether or not a gesture was recalled and performed correctly one week after the learning session, without continued exposure between sessions. Each priming condition had 6 participants (3 conditions × 6 participants) who were asked to recall 10 gestures one week after learning them. Out of the 60 gestures for the control group (6 participants × 10 gestures), only 26 gestures (43%) were recalled correctly. For the primed groups, the percentage of correctly recalled gestures increased significantly, with 48 of the 60 sci-fi gestures (80%) recalled correctly, and 44 of the 60 (73%) creative mindset gestures recalled correctly. Indeed, *Priming* had a statistically significant effect on memorability ($\chi^2(1, N=180) = 11.54$, $p < .01$). *Post hoc* pairwise comparisons using Tukey's HSD correction indicated that the control group gestures were recalled significantly less than either the sci-fi gestures (Z = 3.17, $p < .01$) or the creative mindset gestures (Z = 2.57, $p < .05$).

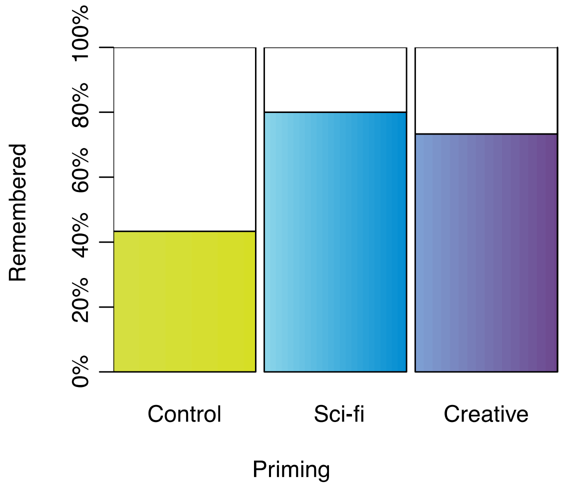

Figure 5. A bar chart of the percentage of correctly remembered gestures by level of *Priming*.

4.5.4 Gesture Performance and Recall Time

In the learnability study, the mean time to perform a gesture successfully after learning it by watching a video clip was 17.3 seconds (*SD* = 10.3) for the control gesture set; for the sci-fi gesture set it was 16.6 seconds (*SD* = 11.9); and for the creative mindset gesture set it was 16.8 seconds (*SD* = 12.5). These differences were not statistically significant ($F(2, 14.97) = 0.25$, *n.s.*).

In the memorability study, the mean time to recall and perform a gesture from the control gesture set was 18.8 seconds (*SD* = 10.6); for the sci-fi gesture set, it was 14.0 seconds (*SD* = 7.3); and for the creative mindset gesture set, it was 13.7 seconds (*SD* = 6.7). These differences were not detectably different ($F(2, 14.99) = 2.19$, *n.s.*).

5 Discussion

In this section, we discuss our findings about the effects of priming across our various studies (elicitation, identification, learnability, and memorability), and some insights on running distributed interaction design studies.

5.1 Effects of Priming on User-Proposed Gestures

Sci-fi priming had half of its gestures (5 out of 10) unique only to it. By contrast, creative mindset priming and no priming (control) had 7 out of 10 gestures in common. The sci-fi gesture set showed more cohesion, despite the gestures for each function being elicited independently. This cohesion was visible in functions that have inverse associations like fast-forward and rewind. The sci-fi gesture set contained the gestures "circle counterclockwise" and "circle clockwise." These gestures were present in our sci-fi movie montage clip (Black Mirror clip under the Rewind tag) that we showed our participants as a primer, and they mapped to the same functions. The control gesture set had "point right" and "swipe right" for fast forward and rewind, respectively. We noticed this lack of cohesion during our learnability and memorability studies, as these gestures were the hardest to learn—*i.e.*, they required the largest number of viewings.

On average, priming seemed to increase the match and pleasure ratings of gestures, both in the elicitation study and in the identification study, although only creative mindset gestures were shown to have statistically significantly higher match and pleasure scores in the identification study. Although the results from the Likert-type ratings were largely non-significant, taken together, there is a clear trend suggesting that priming produces gestures that are perceived to be a better fit to their functions and more pleasurable to perform. This is a limitation of our results and further research is needed, perhaps with a different set of participants or different types of priming, to confirm this trend. This paper offers a large-scale investigation into the contested usefulness of priming in end-user elicitation studies and highlights multiple directions for future research to build upon the results shown here.

5.2 Gestures' Learnability and Memorability

Sci-fi priming significantly increased the initial learnability of gestures, as 80% of the sci-fi gestures were performed correctly after a single viewing, compared to only 58% of creative mindset gestures and 65% of the control gesture set. Sci-fi priming also significantly increased the overall learnability of gestures, with an average of 1.22 views required to learn each sci-fi gesture, compared to 1.60 views for control gestures and 1.56 views for creative mindset gestures. Furthermore, priming in general resulted in gestures that were more memorable than control group gestures, with 80% of sci-fi gestures and 73% of creative mindset gestures recalled correctly, but only 43%

of control group gestures recalled correctly. Thus, it seems again that priming, has advantages in gesture learnability and memorability. Perhaps Sci-fi gestures provided a sense of familiarity that led to their higher learnability and memorability. The enhanced performance of gestures elicited under the sci-fi condition could perhaps be attributed to the high-budget production of the movies that made gestures depicted in them more appeal and easier to learn and remember, or the fact that some of these gestures were created to be memorable by consulting with HCI researchers—as the case is with Minority Report whose director collaborated with researcher from MIT to create these gestures. In any case, interaction designers working on utilizing gestures in their systems could benefit from using movies as inspiration for creating learnable and memorable gestures for easy enjoyable experiences for their systems. In addition, given the similarity in performance from both primed gesture sets (sci-fi and creative mindset) the work we present in this paper shows that it is perhaps the mindset of primed participants that lead to improved learnability and memorability performance not just the type of priming. These results offer an argument for using priming in future elicitation studies at least in the use case presented in this paper.

5.3 Distributed Design Studies

In this work, we demonstrated multiple methods of conducting distributed user-centered design studies typically carried out in a lab. We replicated the distributed elicitation and identification methods presented by Ali *et al.* [4]. Further, we added two more usability metrics to their distributed interaction-evaluation approach: learnability and memorability. Due to the requirement of providing feedback in learnability studies to participants—so participants would know what gestures they learned and which ones they needed to review and attempt to perform again—we had to conduct our distributed learnability study in a supervised manner. A limitation of supervised distributed user studies is they are slower than unsupervised distributed studies like the elicitation and identification studies. It took us a few hours to recruit and collect data for our unsupervised distributed studies compared to our supervised studies that required multiple days to conduct—plus the one week that separated the learnability and memorability studies. Supervised studies, like our learnability and memorability studies, are hard to run in parallel like the unsupervised elicitation and identification studies. This limitation is the reason why our learnability and memorability studies had only 18 participants like an in-lab study—a number that was sufficient for our investigation in this paper. Having multiple researchers conducting a supervised study could increase the number of participants. Distributed learnability and memorability studies enjoy other benefits of distributed studies like increased diversity of participants—in terms of both geographical distribution and physical abilities—and discarding physical requirements such as testing labs. In addition, Ali *et al.* [4] reported that participants are more willing to participate in online studies than lab-based ones. In this work, we were able to capitalize on Ali *et al.'s* [4] Crowdlicit platform to facilitate all of our studies, collecting, organizing and storing study data, validating this platform's versatility. The platform also provided the participants with an easy-to-use interface to participate in the studies.

5.4 Limitations and Future Work

These studies aim to inform the design of future devices with users' preferences, cutting down on the resources required to build, deploy, test, and adjust interaction designs. A limitation of our approach is we were relying on our participants' imaginations to interact with a system. Interacting with an actual system would be a different experience that takes into account gesture-recognition error, among other limitations. Our study only tested two types of priming: the viewing of a sci-fi montage and recalling times of creativity to be in a creative mindset. Priming with other sci-fi clips might produce different results and have a different impact on use cases other than a gesture-controlled mixed reality video player—especially given that some clips depicted characters interacting with a media player. Another limitation to this work with sci-fi priming is that participants could have mimicked gestures popularized by the movies which led to the creation of a more learnable and memorable gesture set. Given the overlap in gesture sets and ambiguity of some of the results, it is hard to tell from the results of this study whether the impact on the sci-fi gestures was because participants were primed to think creatively or mimicked the gestures from the primer. Further investigation is needed to solidify these results. Given the advances the field of elicitation studies has seen recently with systems such as Crowdlicit [4] and Gelicit [27] that enable the collection of gestures online at a large scale, it would be easy to replicate or extend this study to gain a deeper understanding of the effects of priming in elicitation studies in general, and perhaps the effects of sci-fi priming more specifically.

Future work would be to take the gesture sets recommended in this paper and test them either in a usability study that mimics an actual system, like a Wizard of Oz type of study, or invest the resources to build an interactive prototype of a system to test those interactions. Other future work might test the effects of the sci-fi clips we include in this paper on other modes of user-elicited interactions like voice commands of graphical interface elements like icons, or use the video clips of the MR functions we include here to replicate this elicitation study under different conditions. For example, using different priming approaches like having participants do physical activities before an elicitation session could have different effects on the mid-air gestures proposed to trigger the functions listed in this paper to control an MR video player.

6 Conclusion

We conducted the largest investigation, to our knowledge, into the effects of priming on user-elicited gesture in a distributed end-user elicitation study. We evaluated the effects of priming by viewing science fiction clips and having a creative mindset on user-elicited gestures with a novel approach of running supervised distributed learnability and memorability studies. We showed that priming produces user-elicited gestures that are significantly faster and easier to learn and remember. Besides the empirical investigation into the effects of priming, the distributed learnability and memorability methodological contributions, this work contributes a user-elicited gesture set for a media player in a mixed reality environment. We recommend using priming in elicitation studies to unlock participants creativity and elicit interactions that are guessable, learnable, and memorable.

ACKNOWLEDGMENTS

This work was funded in part by Microsoft Research. Any opinions, findings, conclusions or recommendations expressed in our work are those of the authors and do not necessarily reflect those of any supporter.

REFERENCES

1. Chadia Abras, Diane Maloney-Krichmar, and Jenny Preece. 2004. User-Centered Design. *Encyclopedia of Human-Computer Interaction*: 14.
2. Alan Agresti. 2010. *Analysis of Ordinal Categorical Data*. John Wiley & Sons, Inc.
3. Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2018. Crowdsourcing Similarity Judgments for Agreement Analysis in End-User Elicitation Studies. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18* (UIST '18), 177–188. https://doi.org/10.1145/3242587.3242621
4. Abdullah X Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2019. Crowdlicit: A System for Conducting Distributed End-User Elicitation and Identification Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 255. https://doi.org/10.1145/3290605.3300485
5. Yoav Benjamini and Henry Braun. 2002. John W. Tukey's Contributions to Multiple Comparisons. *The Annals of Statistics* 30, 6: 1576–1594.
6. Ceylan Beşevli, Oğuz Turan Buruk, Merve Erkaya, and Oğuzhan Özcan. 2018. Investigating the Effects of Legacy Bias: User Elicited Gestures from the End Users Perspective. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - DIS '18*, 277–281. https://doi.org/10.1145/3197391.3205449
7. David Brin. 2010. Future Tense: How the Net ensures our cosmic survival. *Communications of the ACM* 53, 6: 120–ff. https://doi.org/10.1145/1743546.1743576
8. Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, 361–365. https://doi.org/10.1145/2750858.2805823
9. Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 277–280. https://doi.org/10.1145/2485760.2485823
10. Hessam Jahani Fariman, Hasan J. Alyamani, Manolya Kavakli, and Len Hamey. 2016. Designing a User-defined Gesture Vocabulary for an In-vehicle Climate Control System. In (OzCHI '16), 391–395. https://doi.org/10.1145/3010915.3010955
11. Lucas S. Figueiredo, Mariana G.M. Gonçalves Maciel Pinheiro, Edvar X.C. Vilar Neto, and Veronica Teichrieb. 2015. An Open Catalog of Hand Gestures from Sci-Fi Movies. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15), 1319–1324. https://doi.org/10.1145/2702613.2732888
12. Leah Findlater, Ben Lee, and Jacob Wobbrock. 2012. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2679–2682. https://doi.org/10.1145/2207676.2208660
13. Brigitte N. Frederick. 1999. *Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results*. Retrieved July 1, 2020 from https://eric.ed.gov/?id=ED426098
14. A. R. Gilmour, R. D. Anderson, and A. L. Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 3: 593–599. https://doi.org/10.1093/biomet/72.3.593
15. Donald Hedeker and Robert D. Gibbons. 1994. A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics* 50, 4: 933–944. https://doi.org/10.2307/2533433
16. James J. Higgins and Suleiman Tashtoush. 1994. An aligned rank transform test for interaction. *Nonlinear World* 1, 2: 201–211.
17. Lynn Hoff, Eva Hornecker, and Sven Bertel. 2016. Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias? In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '16*, 86–91. https://doi.org/10.1145/2839462.2839472
18. Wen-jun Hou, Kai-xiang Chen, Hao Li, and Hu Zhou. 2018. User Defined Eye Movement-Based Interaction for Virtual Reality. In *Cross-Cultural Design. Methods, Tools, and Users* (Lecture Notes in Computer Science), 18–30.
19. Jon Favreau. 2010. *Iron Man II*. Paramount Pictures.
20. Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. In (CHI '11), 413–422. https://doi.org/10.1145/1978942.1979001
21. Anne Köpsel and Nikola Bubalo. 2015. Benefiting from legacy bias. *interactions* 22, 5: 44–47. https://doi.org/10.1145/2803169
22. William H. Kruskal and W. Allen Wallis. 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* 47, 260: 583–621. https://doi.org/10.2307/2280779

23. Jerrod Larson. 2008. Limited imagination: Depictions of computers in science fiction film. *Futures* 40, 3: 293–299. https://doi.org/10.1016/j.futures.2007.08.015
24. K. S. Lashley. 1951. The Problem of Serial Order in Behavior. *New York: Wiley.*: PP. 112-131.
25. Eckhard Limpert, Werner A. Stahel, and Markus Abbt. 2001. Log-normal Distributions across the Sciences: Keys and CluesOn the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience* 51, 5: 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2
26. R. C. Littell, P. R. Henry, and C. B. Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science* 76, 4: 1216–1231. https://doi.org/10.2527/1998.7641216x
27. Nathan Magrofuoco and Jean Vanderdonckt. 2019. Gelicit: A Cloud Platform for Distributed Gesture Elicitation Studies. *Proceedings of the ACM on Human-Computer Interaction.* 3, EICS: 6:1–6:41. https://doi.org/10.1145/3331148
28. H. B. Mann and D. R. Whitney. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 18, 1: 50–60.
29. Aaron Marcus. 2013. The History of the Future: Sci-fi Movies and HCI. *interactions* 20, 4: 64–67. https://doi.org/10.1145/2486227.2486240
30. Meredith Ringel Morris. 2012. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (ITS '12), 95–104. https://doi.org/10.1145/2396636.2396651
31. Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3: 40–45. https://doi.org/10.1145/2591689
32. Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding Users' Preferences for Surface Gestures. In *Proceedings of graphics interface 2010* (GI '10), 261–268. Retrieved February 25, 2018 from http://dl.acm.org/citation.cfm?id=1839214.1839260
33. Omar Mubin, Mohammad Obaid, Philipp Jordan, Patricia Alves-Oliveria, Thommy Eriksson, Wolmet Barendregt, Daniel Sjolle, Morten Fjeld, Simeon Simoff, and Mark Billinghurst. 2016. Towards an Agenda for Sci-Fi Inspired HCI Research. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology* (ACE '16), 10:1–10:6. https://doi.org/10.1145/3001773.3001786
34. Michael J. Muller and Sarah Kuhn. 1993. Participatory design. *Communications of the ACM* 36, 6: 24–28. https://doi.org/10.1145/153571.255960
35. 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*, 1099. https://doi.org/10.1145/2470654.2466142
36. Mohammad Obaid, Markus Häring, Felix Kistler, René Bühling, and Elisabeth André. 2012. User-Defined Body Gestures for Navigational Control of a Humanoid Robot. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons and Mary-Anne Williams (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 367–377. https://doi.org/10.1007/978-3-642-34103-8_37
37. Mohammad Obaid, Felix Kistler, Gabriel\.e Kasparavičiūt\.e, Asim Evren Yantaç, and Morten Fjeld. 2016. How Would You Gesture Navigate a Drone?: A User-centered Approach to Control a Drone. In (AcademicMindtrek '16), 113–121. https://doi.org/10.1145/2994310.2994348
38. Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *Human-Computer Interaction – INTERACT 2013* (Lecture Notes in Computer Science), 282–299. https://doi.org/10.1007/978-3-642-40480-1_18
39. Patricia Pons and Javier Jaen. 2019. Interactive spaces for children: gesture elicitation for controlling ground mini-robots. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-019-01290-6
40. Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In (CHI '15), 3347–3350. https://doi.org/10.1145/2702123.2702583
41. Kai Sassenberg and Gordon B. Moskowitz. 2005. Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology* 41, 5: 506–514. https://doi.org/10.1016/j.jesp.2004.10.002
42. Kai Sassenberg, Gordon B. Moskowitz, Adam Fetterman, and Thomas Kessler. 2017. Priming creativity as a strategy to increase creative performance by facilitating the activation and use of remote associations. *Journal of Experimental Social Psychology* 68: 128–138. https://doi.org/10.1016/j.jesp.2016.06.010
43. Michael Schmitz, Christoph Endres, and Andreas Butz. 2007. A Survey of Human-computer Interaction Design in Science Fiction Movies. In *Proceedings of the 2Nd International Conference on INtelligent TEchnologies for Interactive enterTAINment* (INTETAIN '08), 7:1–7:10. Retrieved February 20, 2019 from http://dl.acm.org/citation.cfm?id=1363200.1363210
44. Nathan Shedroff and Chris Noessel. 2012. Make It So: Learning from Sci-fi Interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI '12), 7–8. https://doi.org/10.1145/2254556.2254561
45. John W. Tukey. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics* 5, 2: 99–114. https://doi.org/10.2307/3001913

46. John W. Tukey. 1953. The Problem of Multiple Comparisons. *Princeton, NJ: Princeton University*.
47. 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*, 1325–1334. https://doi.org/10.1145/2702123.2702223
48. Radu-Daniel Vatavu and Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 3390–3402. https://doi.org/10.1145/2858036.2858228
49. Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O Wobbrock. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies? 26.
50. Tijana Vuletic, Alex Duffy, Laura Hay, Chris McTeague, Gerard Campbell, and Madeleine Grealy. 2019. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies* 129: 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011
51. Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the guessability of symbolic input. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, 1869. https://doi.org/10.1145/1056808.1057043
52. Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 143–146. https://doi.org/10.1145/1978942.1978963
53. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (CHI '09), 1083–1092. https://doi.org/10.1145/1518701.1518866
54. Yiqi Xiao and Renke He. 2018. The Handlebar as an Input Field: Evaluating Finger Gestures Designed for Bicycle Riders. In *Advances in Human Aspects of Transportation* (Advances in Intelligent Systems and Computing), 648–659. https://doi.org/10.1007/978-3-319-93885-1_59