Meredith Ringel Morris, Microsoft Research
Andreea Danielescu, Arizona State University
Steven Drucker, Microsoft Research
Danyel Fisher, Microsoft Research
Bongshin Lee, Microsoft Research
m.c. schraefel, University of Southampton
Jacob O. Wobbrock, University of Washington

Reducing Legacy Bias in Gesture Elicitation Studies

Insights

→ Legacy bias is a limitation of current elicitation methods.

→ Participants' first proposal tends not to be their favorite, suggesting production-oriented methods.

→ There are many open challenges in updating elicitation methods to incorporate production, priming, and partner techniques.

Gesture-based systems are becoming ubiquitous. Tablets, phones, large displays, and even laptop computers are now commonly equipped with multitouch-recognizing screens. Third-party accessories like the Wii Nunchuck and the Xbox Kinect can also detect rich gestural input. To design for these increasingly prolific gesture-based systems, we need to understand how to identify and design good gestures in these contexts (a "good" gesture may be one that meets design criteria such as discoverability, ease-of-performance, memorability, or reliability). *Gesture elicitation* is one promising approach to this challenge.

Gesture elicitation (e.g., [1]) is a technique that emerges from the field of participatory design. End users are individually shown the desired effect of an action (called a *referent*) and asked to propose the gesture (called a *symbol*) that would bring that effect about. The results from all user participants are then reconciled to create a single canonical gesture set, possibly including synonyms, using metrics such as agreement [1,2], max-consensus, or consensus-distinct ratio [3]. Gesture elicitation has been applied to a wide variety of emerging interaction and sensing technologies, including touchscreens, depth cameras,

styli, foot-operated UIs, multi-display environments, mobile phones, multimodal gesture-and-speech interfaces, stroke alphabets, and above-surface interfaces.

One advantage of gesture elicitation is that the technique is not limited to current sensing technologies; it enables interaction designers to focus on end users' desires as opposed to settling for what is technically convenient at the moment. As a result, users tend to prefer gesture sets designed through elicitation studies, possibly because professionals tend to generate more physically and conceptually complex gestures [4]. End-user involvement can result in gesture sets that are more likely to be discoverable by and memorable to a large user base.

While gesture elicitation studies show great promise, a potential pitfall is that users' gesture proposals are often biased by their experience with prior interfaces and technologies, particularly the WIMP (windows, icons, menus, and pointing) interfaces that have been standard on traditional PCs for the past two decades. We refer to this problem as *legacy bias*. Users propose legacy-inspired interactions for several reasons: an explicit desire to transfer their knowledge of past systems to new ones, a desire to minimize physical and mental exertion when interacting in new modalities, and misunderstandings of the fundamental capabilities of novel sensing technologies [4]. Such biases may cause gesture elicitation methods to get caught in local minima, failing to uncover interactions that may be better suited for a given medium than those that leap readily to users' minds.

Reports from research studies employing gesture elicitation note many examples of legacy bias. For instance, Wobbrock et al. noted that, despite presenting participants with a large multitouch touchscreen without UI elements from traditional PC interfaces, most participants suggested mouse-like single-point or simple-path gestures [1]. The participants, too, acknowledged these biases: One said, "I'm a child of the mouse"; another said, "I'm falling back on the old things that I've learned." In a multimodal gesture and speech elicitation study, Morris [3] noted similar examples, including a participant who referred to his hand as the mouse, a participant who avoided bimanual gestures because they might require that a system have two cursors, and several participants who suggested speech commands based on keyboard shortcuts (e.g., saying "F5" aloud to reload a Web page).

Legacy bias does have some benefits. Because they draw upon culturally shared metaphors, participants tend to propose similar legacy-inspired interactions, resulting in high agreement scores in elicitation studies [1]. This agreement indicates that legacy-inspired interactions are easily guessable and learnable [2], and perhaps appropriate for systems intended to be walk-up-and-use, such as touch-based kiosks in public venues. Legacy-inspired interactions also tend to be relatively simple to execute: Mouse-inspired gestures typically require only a single finger, which may reduce fatigue for frequent interactions [4] and/or increase accessibility for users with physical or situational impairments. In general, however, legacy bias limits the potential of user-elicitation methodologies for producing interactions that take full advantage of the possibilities and requirements of emerging application domains, form factors, and sensing capabilities. Here, we propose modifications to Wobbrock et al.'s basic elicitation methodology [1,2], aimed to counteract users' legacy tendencies. We report on initial findings that point toward the potential of our proposed approach and discuss open research challenges related to these proposals.

Legacy bias limits the potential of user-elicitation methodologies for producing interactions that take full advantage of emerging application domains, form factors, and sensing capabilities.

BEYOND WIMP: IMPROVING ELICITATION

We propose three techniques for improving end-user elicitation studies: *production*, *priming*, and *partners*. These techniques are aimed specifically at reducing legacy bias and increasing the novelty of gestures produced. They could be used alone or in combination. Our suggestions are rooted in research findings from other domains that

we believe have much to offer for improving elicitation studies.

Production. Requiring users to produce multiple interaction proposals for each referent may force them to move beyond simple, legacy-inspired techniques to ones that require more reflection. Production has been shown to increase variety and creativity in output in other domains. For example, the ESP image-labeling game lists "taboo" words to prevent users from always proposing obvious tags for images. Research on the design process, such as that by Dow et al. [5], finds that forcing designers to generate a large set of initial ideas results in better final designs.

These concepts could be applied to interaction elicitation methodologies by encouraging participants to generate many different symbols. For example, users could be required to produce some minimum number of different symbols for each referent. Alternatively, participants could be instructed to continue producing symbols until the point at which they propose a novel interaction that has not yet been proposed by prior participants in the elicitation exercise. Many other creative variations on production strategies are also possible.

Priming. Priming users to think about the capabilities of a new form factor or sensing technology is another approach that may reduce the impact of legacy bias. Priming has been shown to have a wide range of applications in psychology studies, including enhancing creative thinking. Priming techniques have also been used effectively in HCI; for example, North et al. found that users who performed a task with physical objects before using

a multitouch table were less likely to use only pointing-based interactions than users who had performed the task first with a mouse [6].

This lesson can be applied to interaction elicitation. For example, participants could watch demonstrations, either videos or the experimenter's own actions, of a variety of possible ways of using the target technology. This might prompt users to think more generally about what gestures could be used to accomplish a given task, as well as correct any misconceptions about the capabilities of new technologies. Perhaps participants could be shown gestures created by HCI professionals, to be inspired by the more complex designs that professionals tend to create [4]. Indeed, rather than merely viewing these options, participants could be asked to mimic the priming examples they see, in order to more fully immerse themselves in the creative process, in much the way that improvisational actors enhance their creativity through physical warm-up exercises.

Partners. Inviting users to participate in elicitation studies in groups, rather than individually, can be another approach to overcoming legacy bias. Borrowing again from the field of design, where group brainstorming is a common practice for leveraging others' ideas, we find much evidence that users can fruitfully build upon one another's ideas (e.g., [5]). Morris conducted an elicitation study with pairs of participants and noted that participants would often improvise based on their partner's suggestions [3]. In addition to facilitating creativity, group-based approaches can also increase ecological validity for eliciting interactions for multi-user systems and scenarios.

Partner-based methods might be as simple as Morris's approach of having pairs (or small groups) of participants engage in the elicitation exercise jointly [3]. Alternatively, more complex methods of establishing rules for the multi-user engagement might facilitate achieving certain design outcomes. For example, a premise based on popular games like Charades, in which a participant knows that their partner will need to be able to guess the meaning of their gesture, could be used to encourage participants to develop interactions that are more readily guessable. A premise based on the classic Telephone game, in which partners know that others will have to accurately mimic their gestures, might be used to encourage the generation of highly memorable and/or reproducible gesture candidates.

PILOT STUDY

We conducted a pilot study to explore whether modifications to gesture elicitation methodology aimed at reducing legacy biases could result in better outcomes; this initial study included both production and priming components but did not explore partner-based techniques.

We were designing free-space gestures for a hypothetical depth-camera-based data navigation system. We asked 17 participants with non-technical backgrounds to individually perform an elicitation exercise. Each participant was asked to produce a sequence of possible gestures for each referent, on the presumption that participants might produce more creative suggestions once they had "used up" their legacy-inspired ideas. Though we targeted five gestures, we did not communicate this threshold to participants; rather, we repeatedly prompted them to produce additional possibilities until they had suggested at least five and at most nine interactions. It took participants between 45 and 90 minutes to complete this exercise for a set of 14

referents.

Some participants were primed with a video that showed them gestures used outside computing scenarios: For example, it showed a sports referee, an aircraft carrier signaler, and friends waving to each other. A subset of these participants were also *kinesthetically primed*: They were asked to carry out physical actions, such as touching their toes, doing jumping jacks, and pointing at the corner of the room.

Though all participants proposed legacy interactions (such as pointing at the screen to select items), they also proposed a wide variety of gestures that took advantage of the depth camera's vision-based sensing capabilities, using not only hands and arms but also their legs, heads, and full bodies (moving in the space, leaning, turning, and twisting). Participants who were kinesthetically primed tended to produce more gestures that involved moving about the room than others, although this trend fell short of statistical significance; replication with more participants is likely necessary to verify the impact of priming on gesture style. We asked participants to identify which of their proposed gestures for each referent was their favorite. The median position of the favorite gesture was the third one, indicating that participants' first suggestions were generally not optimal, and that production of more than one symbol has value. However, we noticed that some participants' gestures diminished in variety: Having pointed with their hand up, for example, they might then point with their hand down, and then point again with a fist. This suggests there is room for improvement in the basic production method.

OPEN CHALLENGES

Our pilot study indicates that these modifications hold promise for increasing the variety, novelty, and quality of user-elicited interactions. However, there is much more work to be done in fine-tuning this design methodology. In this section, we consider some of the latent variables in these designs.

Consider production. Does it truly increase the variety of proposed interactions, or are downstream proposals from the same participant simply minor variants on earlier

We noticed that some participants' gestures diminished in variety: Having pointed with their hand up, for example, they might then point with their hand down, and then point again with a fist.

ones? How should such variety be measured? What is the minimum number of symbols that participants must be asked to produce in order to move beyond legacy bias? Is there a number of symbols beyond which a participant's proposals drop in quality, as our pilot suggests? Would knowing the number of symbols they will be asked to produce impact participants' creativity? What is the relationship of a symbol's position within a production stream to the level of agreement with other users' proposals? How does a symbol's position within a production stream relate to various quality metrics for an interaction, such as learnability, guessability, memorability, ease of performance, and user preference? Are symbols at later positions within a production stream more similar to the types of interactions proposed by professional designers?

Turning to priming, we wonder what types (videos, professional exemplars, kinesthetics, etc.) best counteract users' legacy biases. What would be the ideal duration of priming? How much does priming expand or restrict the range of symbols produced by participants? Does priming introduce its own biases into elicitation methodologies? Are these beneficial or detrimental? Is it possible to create a standard set of priming materials?

For partner techniques, open questions include how the number of group members relates to the reduction in legacy bias or the quality of the final interactions produced. Further, how does group composition influence bias reduction or interaction quality? Possible compositions include groups of strangers or of close ties, groups with similar or diverse backgrounds, groups that involve a mixture of novice end users and HCI professionals, or even groups that include a confederate of the experimenter. Another key open question is how partner-oriented methodologies could be designed so as to minimize potential inhibitory effects, such as social conformance. It also remains to be seen whether certain gamification strategies, such as our examples aimed at optimizing guessability (Charades)

or memorability (Telephone) for a partner, risk amplifying legacy bias, and whether group members should be positioned as competitors or collaborators for best results.

Additionally, for all three of our proposed approaches, it is important to consider how specific methodologies can be designed to minimize possible side effects such as participant fatigue and stress, and what tools or tool features are necessary to support elicitation studies that use each of these techniques.

CONCLUSION

Our aim here has been to begin a discussion within the HCI community about a potential shortcoming of end-user interaction elicitation, a popular methodological strategy, and particularly for gesture elicitation. We propose that legacy bias, though not without its peripheral benefits, results in suboptimal outcomes for elicitation studies, and that modification of elicitation techniques to incorporate strategies based on production, priming, and/or partners is a promising avenue of research. Our own initial experiment suggests that such approaches hold potential. Several opportunities and challenges left to the HCI community include rigorously evaluating the efficacy of these suggestions; quantifying their impact on bias reduction, agreement, and interaction quality; and identifying the associated methodological modifications necessary to optimize these techniques.

Endnotes
1. Wobbrock, J.O., Morris, M.R. and Wilson, A.D. User-defined gestures for surface computing. *Proc. of CHI 2009*.
2. Wobbrock, J.O., Aung, H.H., Rothrock, B., and Myers, B.A. Maximizing the guessability of symbolic input. *Extended Abstracts of CHI 2005*.
3. Morris, M.R. Web on the wall: Insights from a multimodal interaction elicitation study. *Proc. of ITS 2012*.
4. Morris, M.R., Wobbrock, J.O., and Wilson, A.D. Understanding users' preferences for surface gestures. *Proc. of GI 2010*.
5. Dow, S.P., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D.L., and Klemmer, S.R. Prototyping dynamics: Sharing multiple designs improves exploration, group rapport, and results. *Proc. of CHI 2011*.
6. North, C., Dwyer, T., Lee, B., Fisher, D.,

Isenberg, P., Robertson, G., and Inkpen, K. Understanding multi-touch manipulation for surface computing. *Proc. of Interact 2009*.

🔸 **Meredith Ringel Morris** is a senior researcher at Microsoft Research. Her research primarily focuses on collaborative and social technologies. Her work on interaction techniques for large touchscreen displays led to her interest in gesture elicitation. She received her Ph.D. in computer science from Stanford University.
→ merrie@microsoft.com

🔸 **Andreea Danielescu** is a doctoral student in computer science with a concentration in arts, media, and engineering at Arizona State University. Her research focuses on designing intuitive and discoverable gestures for walk-up-and-use interfaces. Her broader research interests include data visualization, image processing, and human cognition.
→ lavinia.danielescu@asu.edu

🔸 **Steven Drucker** is a principal researcher at Microsoft Research focusing on human-computer interaction for organizing and communicating large amounts of information. Prior to his 18 years at MSR, he received his Ph.D. from the MIT Media Lab and his master's from the MIT AI Lab.
→ sdrucker@microsoft.com

🔸 **Danyel Fisher** is a researcher at Microsoft Research; his work centers on information and data visualization. He is interested in how users can employ visualization to better make sense of their data. He received his M.S. from UC Berkeley and his Ph.D. from UC Irvine.
→ danyelf@microsoft.com

🔸 **Bongshin Lee** is a researcher at Microsoft Research, currently focusing on developing new ways for people to create information visualizations and interact with their data through natural user interfaces (NUIs), including pen and touch. She received her M.S. and Ph.D. from the University of Maryland at College Park.
→ bongshin@microsoft.com

🔸 **m.c. schraefel** is a professor of computer science and human performance at the University of Southampton; her research focus is designing information systems that enhance innovation, creativity, and discovery.
→ mc@ecs.soton.ac.uk

🔸 **Jacob O. Wobbrock** is an associate professor in the Information School and an adjunct associate professor in computer science and engineering at the University of Washington. His research focuses on input and interaction, notably the EdgeWrite alphabet, the first gesture set created via a user-defined gesture methodology.
→ wobbrock@uw.edu