

Gendered or Neutral? Considering the Language of HCI

Adam Bradley*

English Language and Literature
& Systems Design Engineering
University of Waterloo

Cayley MacArthur*

English Language and Literature
& Systems Design Engineering
University of Waterloo

Mark Hancock*

Management Sciences
University of Waterloo

Sheelagh Carpendale†

Computer Science
University of Calgary

ABSTRACT

In this paper, we present a Mechanical Turk study that explores how the most common words that have been used to refer to people in recent HCI literature are received by non-experts. The top five CHI 2014 people words are: user, participant, person, designer, and researcher. We asked participants to think about one of these words for ten seconds and then to draw an image of it. After the drawing was done we asked simple demographic questions about both the participant and the created image. Our results show that while generally our participants did perceive most of these words as predominately male, there were two notable exceptions. Women appear to perceive the terms “person” and “participant” as gender neutral. That is, they were just as likely to draw a person or a participant as male or female. So while these two words are not exactly gender neutral in that men largely perceived them as male, at least women did not appear to feel excluded by these terms. We offer an increased understanding of the perception of HCI’s people words and discuss the challenges this poses to our community in striving toward gender inclusiveness.

Keywords: Users; people; language; gender issues in HCI.

Index Terms: H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

1 INTRODUCTION

At a first glance it would appear that the human-computer interaction (HCI) community is carefully gender neutral in that we fairly consistently choose gender non-specific words when referring to humans: i.e. user, participant, person, designer and researcher (CHI 2014 people words). However, like so many professions, we still face gender inequalities in our ranks, which is leading to research that considers questions of gender equality in HCI [1-10]. We join this body of research by conducting a study that explores whether the language used in our papers to describe people is actually perceived as gender neutral. Our Mechanical Turk study asked participants to think about one of the HCI people words (a user / participant / person / designer / researcher) for ten seconds and then to draw a picture of them. We found that while we may intend the language that we use within our community to be neutral, it is largely perceived as gendered.

Discussions of equality within HCI are becoming more prevalent [1, 3, 4, 9, 10-16]. Indeed, one of the core principles in human-centred design (HCD) is the consideration of the variety of needs and differences between people [17]. For example, the use of personas [18] in design intentionally includes descriptions of people’s gender, age, background, and a story to describe their unique situation, with the intention that designers visualize and

base design decisions around a variety of demographics. In some instances, these stories and scenarios find their way into academic literature in HCI, but more frequently the language used in our papers centres on the use of nouns such as “users”, “designers”, and “participants”, with the intention that these avoid specifying a gender, and can be thought of as a generic “person”.

However, it is not clear whether this use of non-gendered terms in our academic literature actually leads to non-gendered impressions. Since we, as a community, are interested in promoting gender equality [2, 12, 15, 16, 19], in encouraging young women into technological career paths in general and HCI in particular [8], we decided to apply common HCI empirical methodology to our own use of vocabulary to discover whether our choice of non-gendered terms is effective in providing gender inclusiveness.

Our study reveals that the common HCI people words are not always gender neutral. We have found that the language we use to describe people within the community carries gendered, and in some cases other demographic implications. We suggest that we, as authors, should begin to consider the implications of our word usage in publications, and we, as a community, should begin to develop strategies to deal with this likely unintentional bias.

In this paper, we provide evidence that words generally considered within the HCI community to be non-gendered, or at least gender-neutral, are instead perceived as having specific qualities by Mechanical Turkers. We recognize that Mechanical Turk has limitations, such as the tendency towards those who are aware of its existence; however, this method provides a larger and more diverse sample than, for instance, a lab study with university students. The impression of non-experts about our language use is important for a variety of reasons, including the recruitment of future women to the discipline of HCI, the tendency for people to adopt our terms (e.g., “user” is quite commonly understood as a person who uses technology), and the need for our research papers to reflect the diversity and inclusivity we aim to encourage in the use and design of technology. Our work is a first step toward understanding what these words connote, and can act as a baseline for future studies that can compare their use in various contexts.

2 RELATED WORK

There has already been significant work in HCI discussing issues of gender, including discussions of bias in design [2, 20], encouragement of participation of women in computer science and HCI [8], and feminism in HCI [1, 3, 4, 11, 12].

The feminist perspective on research and design in HCI has been gaining momentum within the HCI community, inviting methodologies to be developed [4], workshops to be held [15], and reflections to be documented [12]. However, a trend map [14] showing topics of discussion over the past 5 years of CHI demonstrates that discussions around women were more often linked to public or community life, and reached a peak about three years ago, before starting to decline.

One such contribution bridging feminism with an HCI framework is Burnett et al.’s [6] study showing that, in problem solving software, there are significant gender issues across several factors, including which features are used, people’s willingness to tinker, and general confidence. They then use this information to inform the design of problem solving software. Vasilescu et al.’s [13]

*{firstname.lastname}@uwaterloo.ca

†sheelagh@ucalgary.ca

study of women’s representation in online communities suggests moving away from the gender binary, speculating that this could be causing problems in itself.

Gender issues have also been explored in design of artifacts; for example, Significant Screwdriver [2], which is an exploration of how “designs can perpetuate and even create gender roles and divisions” [2] focused on meaning attributed to a physical object, but, as Winograd indicates, digital design choices can have the same effect [20]. The decisions involved in interface design are grounded in a series of observed interactions from which models have been abstracted and applied in practice: “theories about the nature of biological existence, about language, and about the nature of human action have a profound influence on the shape of what we build and how we use it” [20].

Wajcman [19] argues that “as with science, the very language of technology, its symbolism, is masculine.” This language adheres to neutral-sounding terms, standardized by a society that values this rationality (masculinity) over the emotional (feminine). This empiricist view of science as “(gender) neutral” has been critiqued and dismantled extensively by feminist authors throughout the 20th century [19]. In feminist theory, a “false universal” is a word that suggests a norm through unspoken connotations, even though it is supposed to encompass everyone (i.e. “mankind”) [21]. Bardzell writes “the interaction design process takes place independent of gender considerations, and even today the central concept of the whole field—the user—remains genderless” [1]. The theory of the false universal has not been formally explored in HCI, and we wish to identify the possible norms within our own field while exploring whether words like “user” have become false universals. Buttrick et al. [7] and Light [10] have brought attention to a problematic relationship between humans and technology through exploration of unconventional themes that call some norms into question. We have chosen to do this through drawing, based in part on the idea from Sontag [22] that our language contains more stereotypes than the things we draw, so examining images can help us uncover meanings that would not necessarily be conveyed through words.

The collection and analysis of drawings is a popular and well-documented methodology in psychology (e.g., [23]). Drawings can help us understand mental processes, showing us representations of perspectives imbued with history and culture [24]. One of the goals in soliciting drawings is to understand the “relationships that the artists have with the object of their representations” as well as any “gendered productions” or the discourses that would account for these drawings [25]. The collection of drawings to find out specifically about stereotypes was pioneered by Chambers’ influential Draw-A-Scientist Test (DAST) [26]. We build on this work by using this drawing method to investigate whether common “people” words used in HCI literature are gendered.

3 USE OF LANGUAGE IN HCI LITERATURE

The word most often used in CHI literature to describe people is “user” (17663 instances in 2014). This number was arrived at by creating a concordance of all papers published at CHI in 2014 and ranking them by frequency. In our study, we look at how the general public perceives the word “user” as well as other popular word choices to describe people. Feminist HCI [1] and the neurodiversity movement [27] have challenged the use of “user”, but no one has provided empirical evidence for what kinds of biases are associated with the terms. Coleman [28] suggests that the word “user” could imply “dependency, addiction, and, ultimately, objectification where one is conscripted into the logic of whatever is being used, be it a technology, drug, etc.”

Carnegie [29] deals with interface design and how in the past it has been treated as if it does not influence our interactions. Dour-

ish and Bell [30] call on the community to “acknowledge the ways in which technologies both exploit and reproduce a range of power concentrations and relationships”. As Czerwinski et al. [31] demonstrate in their study of 3D navigation that showed that wider fields of view and larger displays reduced the gender gap in performance without negatively impacting males; technologies can indeed exacerbate the gender gap.

In our work, we investigate whether using conventional “people” words from HCI literature could be contributing to a gender gap by studying whether these words have gendered connotations.

3.1 Vocabulary usage motivation

Psychological research has already set a precedent for linguistic awareness and change within the HCI community that can be found in the discussion about the word “subject” and the word “participant”. In 1995, The British Psychological society “[a]fter noting that psychologists owe a debt to those who agree to take part in their studies... deserve to be treated with the highest standards of consideration and respect”, the society recommended that the term “subject” be abandoned and replaced by “participant” [32]. In 1998, Boynton [33] published an article in the British Medical Journal entitled “People should participate in, not be subjects of, research” in which he called for terminology reform in psychological research studies to shift from the generally accepted word “subject” to that of “participant”. This action was spawned by a belief that the words that psychologists used held power and had an influence over how they were conducting their research studies. Figure 1 shows an n-gram analysis of the use of the words “subject” and “participant” in CHI papers published since the first SIGCHI conference in 1982. It is clear that in or around 1998 the use of the word “subject” was replaced by the word “participant”.

What the psychologists were in fact doing was realizing that by labeling their “participants” as “subjects”, the pragmatics of the word “subjects” was negatively impacting their work. It goes without saying that psychology is a large part of HCI, and our community has already shown that they support this kind of thinking by changing their terminology to match that used in psychology. Following Boynton’s suggestion, we have also chosen to refer to our Mechanical Turk workers as “participants” throughout this paper. What we suggest is that there are other words in our lexicon, specifically the word “user”, that act in negative ways like the word “subject”. By considering how these words operate and how their history has affected their meaning, we can gain a larger picture of exactly how their use may be affecting our work.

4 STUDY: DRAWING PEOPLE WORDS

We designed a study to specifically investigate whether words we use in our research papers are perceived as gendered. We were particularly interested in the gender, age, race, and other demo-

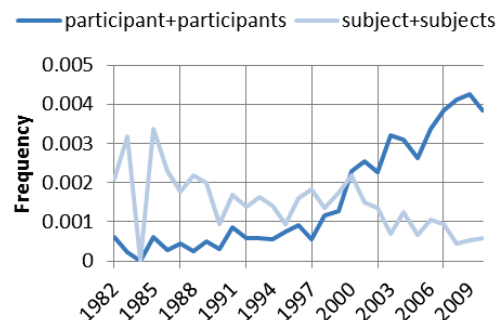


Figure 1: Frequency of word usage for the words “participant” and “subject” from the dataset of all 1982-2011 CHI papers.

graphic information people associate with these words, and whether this relates to one's own demographic information (e.g., do men think of men, and women think of women). We therefore designed a study in which we asked participants to draw what they imagined for people-themed words from CHI 2014 literature.

Our study design was modelled after work by McMaster [34], who explored the feasibility of similar image-based research using Mechanical Turk. In this work, he suggests that a globalizing effect on imagery may be taking place, accounting for consistency in visual themes collected despite the diversity of respondents [34]. As Cohn points out [35], the images people draw cannot directly be translated into specific motivations, but can provide higher-level themes showing how the imagery originates from a cultural and temporal context. In collecting images we do not intend to identify individuals with biases, but to uncover broader themes.

We also took inspiration from Schmettow [36], who sought to illustrate individual differences in what people associate with computers by using priming to understand implicit associations; in our study, participants are making associations with the people words. Rather than asking for word associations after seeing a picture, we wanted to inspire visual representations of common words to see whether participants identified with the universal terms used in HCI. Given the prevalence and success of drawing studies in psychology and the work done by McMaster, we were interested in applying this approach to HCI people words.

4.1 Participants

Participants were recruited from the Mechanical Turk marketplace. We posted 5 unique Human Intelligence Tasks (HITs), one per word with 150 workers per HIT, for a total of 750 instances; however, some participants completed multiple HITs leaving only 433 unique respondents (58 completed all 5, 9 completed 4, 12 completed 3, and 34 completed 2). We received 766 responses, which is greater than the number of requested HITs, perhaps due to participants restarting the study or revisiting the URL after problems such as system failure. Thus, in our data cleaning process, we used only complete responses, with an image and corresponding questionnaire data. This gave us 757 unique responses. Therefore, our analysis is based on the unit of responses, rather than participants, affecting sample size, degrees of freedom, etc.

As to English proficiency, 328 participants indicated English as their primary language, and, of the 316 that indicated another language as primary, 226 (72%) indicated being fluent, 73 (23%) indicated functional knowledge, and 17 (5%) indicated limited or no knowledge of how to communicate in English. Demographic information, such as gender, ethnicity, and age are primary factors in our analysis, and so are discussed in the Results section.

4.2 Task & Conditions

Each worker was directed to our server via Mechanical Turk and asked to think for 10 seconds about one of five words: "user", "participant", "person", "designer", and "researcher" using the exact phrase "for the next 10 seconds think about a ...". Our software prevented them from continuing the survey until this 10 seconds had elapsed. They were then asked to draw that word with the phrase: "in this box, draw a ... sitting down". We specifically recommended that participants switch to a device that provided touch or pen input, if they had one available. Once they uploaded a screenshot of this drawing, they were given a survey designed to gather data about the drawing (its gender, age, ethnicity, etc.), and then asked to describe themselves (again, gender, age, ethnicity, etc.) via a set of predefined questions.

Because our study was specifically targeted at collecting data about gender issues, we took some care in how we asked participants about gender. Specifically, we asked:

"Does the subject of your drawing have a gender, and if so, what is it?"

With the options: Male/Female/Other, and a follow-up question:

"If you answered 'other', you may elaborate here".

Similarly, for ethnicity, we asked:

"What is the race/ethnicity of the person in your drawing? (i.e., peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, colour, language or religion)"

With options: Caucasian, Latino/Hispanic, Middle Eastern, African, Caribbean, South Asian, East Asian, Mixed, Other (clarify).

The study was implemented through Mechanical Turk, but participants were directed to a local web server that hosted a web-based application that allowed participants to draw in an HTML5 canvas, and then respond to survey questions.

4.3 Word selection

The list of five words was derived by gathering all of the papers published in the CHI 2014 proceedings, performing a frequency analysis of individual words, and making a list of the top five words used to reference people within this dataset. To come up with this final list we combined words that had the same lemmas. For example we combined the words "participant" and "participants" into a single list item, as well as "person", "persons", and "people". In this literature there were 17663 instances of "user", 14523 of "participant", 2337 of "person", 1092 of "designer", and 1044 of "researcher". Note that the word "individual" was sometimes used to refer to people and actually has 1795 instances in CHI 2014 literature. We discussed including this word in our list; however, many of these instances were not referring to people (e.g., the phrase "each individual <object>" was often used to described artifacts, instances of data, etc.), and so we did not consider this to be in the top five of words referring to people.

4.4 Hypothesis

We began thinking about this work because we thought that there was a possibility that the word "user" was a false universal representing a norm, and would be predominantly thought of as male. We also expected that this bias would exist for participants of all genders. We thus hypothesized:

H1. Participants would draw "a user" and describe the drawing more frequently as "male".

After an initial pilot, with the conditions "user" and "person", we observed the trend that both were thought of as male, and decided to include the five conditions described above. We thought that this might indicate that all of the words we use to describe or identify people within CHI papers are thought of as male. We thus can describe these same hypotheses for any of the words as follows:

H1 (revised). Participants would draw any of these words and describe the drawing more frequently as "male".

We expected a similar bias about ethnicity descriptions, and thus had the following hypothesis:

H2. Participants would draw any of these words and describe the drawing more frequently as "Caucasian".

5 RESULTS

Participants drew a wide variety of images, which ranged significantly in detail and showed a clear attention to thoughtful interpretations of the words. While we hypothesized that any gender effects in the data would be population wide, what we found was that they were divided by the gender of the participants. For this

reason we have divided our analysis into separate conditions based on gender of the participants. Figure 2 shows a sample of images collected in each condition. Unless stated otherwise, all the data reported in this paper is as reported by the participants in the survey accompanying their drawings and has not been coded.

5.1 Gender

Across all five conditions, only 14 responses described the participant's gender or the gender of the image as "other", and so statistical analyses on data about this group was not conducted (Figure 4). There were also 134 instances when the gender of the drawing or participant is unknown. This occurred when participants described the drawing as not being human, or when participants opted not to provide data about the image's or their own gender.

We therefore used a Pearson's Chi Square test on the remaining drawings, which were described as being male or female by participants who described themselves as either male or female. We

used the null hypothesis that the drawings would be equally distributed across these two genders; that is, that the word was interpreted as being non-gendered, and so would be uniformly distributed independent of participant gender. Figure 3 shows the frequency distributions for all five words by the different genders.

For all five words, images drawn by males were more frequently of males than of females (user: $\chi^2(1, N=71)=39.6$, $\phi=0.71$, $p<.001$; person: $\chi^2(1, N=86)=31.4$, $\phi=0.59$, $p<.001$; participant: $\chi^2(1, N=75)=37.5$, $\phi=0.70$, $p<.001$; researcher: $\chi^2(1, N=64)=42.2$, $\phi=0.76$, $p<.001$; designer: $\chi^2(1, N=61)=15.8$, $\phi=0.48$, $p<.001$). This suggests that when men read these commonly used words, they tend to perceive these people to be men.

Similarly, images drawn by females were more frequently of males than of females for the words "user" ($\chi^2(1, N=49)=4.6$, $\phi=0.29$, $p=.03$) and "researcher" ($\chi^2(1, N=52)=15.1$, $\phi=0.51$, $p<.001$). This finding suggests that, when women read these two words, they tend to perceive these people to be men.

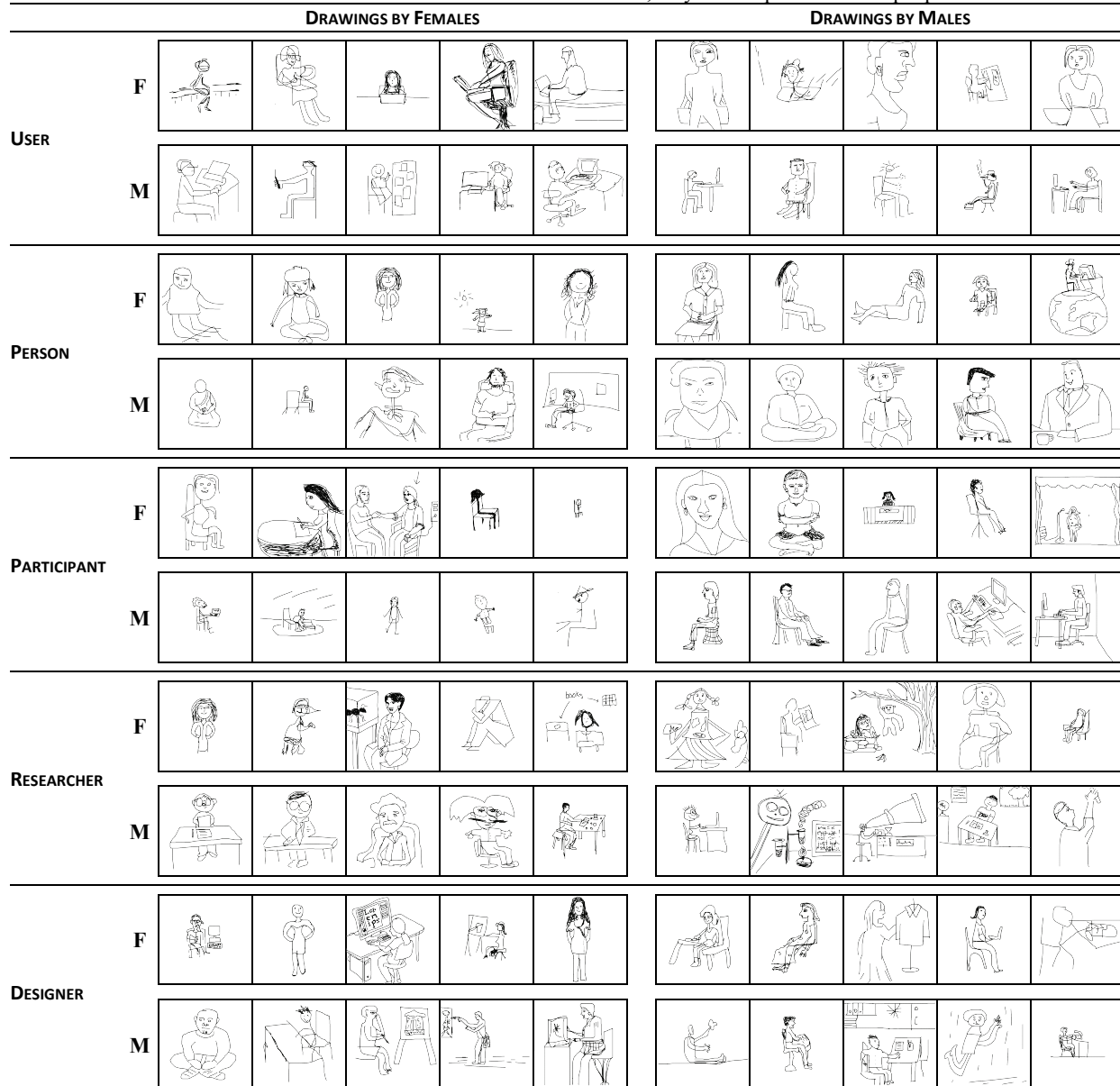


Figure 2: Sample drawings of the top five words used to describe people in CHI 2014 papers. Drawings done by female participants are on the left, and drawings by male participants on the right. Within each word condition, drawings on the top row were identified as being of females, and drawings on the bottom row were identified as being of males.

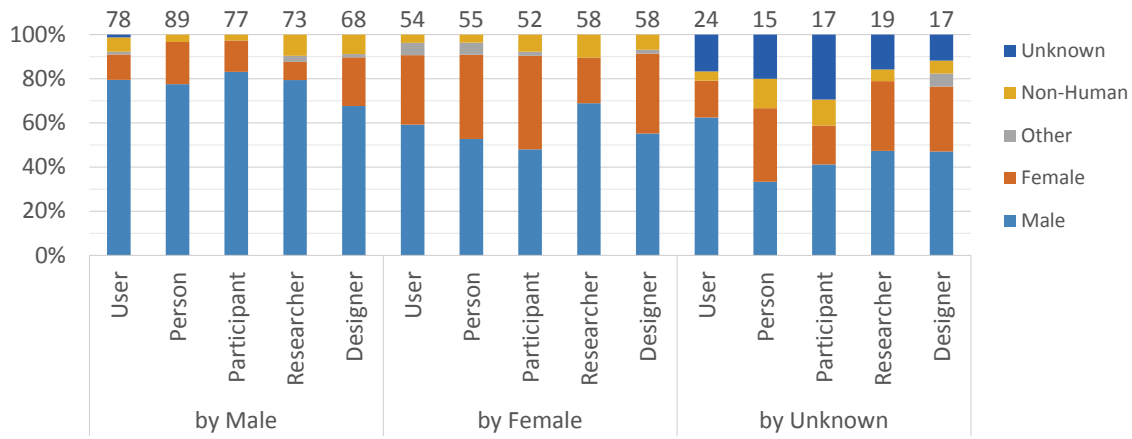


Figure 3: Percentage of drawings described by participants as being of a particular gender. Numbers at the top describe the number of valid responses in that condition. The subject of drawings by males were of males for all words, but the subject of drawings by females were of males only for “user”, and “researcher”. Only data where gender was specified was analyzed.

Promisingly, images drawn by females for the words “person”, “participant”, and “designer” were not more frequently of either men or women (person: $\chi^2(1, N=50)=1.3$, $\phi=0.15$, $p=.26$; participant: $\chi^2(1, N=47)=0.2$, $\phi=0.06$, $p=.66$; designer: $\chi^2(1, N=53)=2.3$, $\phi=0.20$, $p=.13$). While it is not appropriate to accept the null hypothesis (i.e., that this frequency is actually equal to 50%), this finding suggests that this gender bias, if it does exist, is far less pronounced for women for these words than it is for men.

Based on this data, we can confirm H1 for drawings by male participants and all words, but can only confirm H1 for drawings by female participants for the words “user” and “researcher”.

5.1.1 Gender Markers

In our analysis of the images, we observed many images with what could be considered gender markers (Figure 5), such as facial hair and ties for men, and dresses and bows for women. We recognize that these markers are not exclusive of gender (women can wear ties and men dresses), but participants identified these images as being male or female as well. In our counts we exclusively used declarations of participants and merely note here that drawings often, but not always, also included gendered markings.

5.2 Ethnicity

The sample in our dataset was not spread sufficiently across all ethnicities to be able to conduct formal analyses of this data (223 Caucasian, 313 South Asian, and 104 not specified, with the remaining 126 spread across conditions such that each ethnicity-condition total was ≤ 10 data points). However, as shown in Figure 6, there was a tendency for Caucasian and South Asian participants to draw people of their own ethnicity in all conditions. It appears as though participants of other ethnicities tend to draw Caucasians in addition to their own ethnicities, though the samples



Figure 4: A selection of images labeled with genders as “other”, colour-coded to indicate the gender of the drawer.

of these populations are perhaps too small to draw conclusions about this data. Thus, we cannot confirm H2.

5.3 Age

We compared the age of participants to the age of the person drawn using paired t-tests. As shown in Figure 7, drawings were of people younger than themselves when they drew “participants” ($t(97)=3.9$, $d=0.39$, $p<.001$) and older than themselves when they drew “researchers” ($t(106)=4.7$, $d=0.46$, $p<.001$). For the remaining three words, the age of the drawings was not significantly different than the age of the participant (user: $t(96)=1.8$, $d=0.19$, $p=.07$; person: $t(122)=1.4$, $d=0.13$, $p=.16$; designer: $t(101)=0.7$, $d=0.07$, $p=.49$). This finding suggests that our participants perceive researchers to be older, participants to be younger, and the remaining words (user, person, and designer) to be of similar age.

5.4 Use of Artifacts

Since we asked participants to draw their condition sitting down, we saw many versions of chairs. We were also able to distinguish additional artifacts in the drawings as participants were prompted to identify them in words. We therefore noted some interesting artifacts such as lab equipment for the “researcher” condition and art easels for the “designer” condition (Figure 8).

5.4.1 Computers

Many drawings in the user condition were of a person sitting down typing at a computer (Figure 9). Interestingly only 6 drawings (3.8%) included images where the participants indicated in their responses to the questionnaire that they had drawn phones or tablets, suggesting that “user” is thought of not only predominantly as male, but also as a desktop computer user.

5.4.2 Drug Use Equipment

One of the interesting aspects of the “user” condition was that 9 drawings (5.8%) were images of drug users (Figure 10). Usually these images were described by our participants as either having needles or drug smoking paraphernalia present within the image. This finding suggests that the word “user”, despite being so ubiquitously used when referencing people who use technology, is still thought of in the context of drug use by a more general population. Thirty years after the term “user” was first appropriated in 1935 to describe drug takers, it was then used as the go-to noun for the emerging computer revolution [37]. Our data suggests that the earlier definition is still alive, and at times will be imagined by

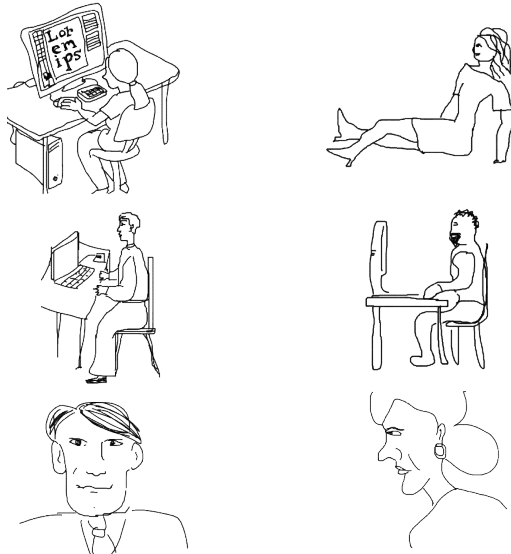


Figure 5: (top) A female “designer” and an image of a “person” wearing a dress and high heels. (middle) Two male “users” and a drawing with the “male” gender marker of a goatee. (bottom) a male and female “person”.

readers. Even though it was a small percentage that drew drug users (5.8%), our study provides evidence that this meaning is still implicated by the use of this word.

Moving forward, we need to be aware as a community that all of the terminology we use has the potential to not only be misinterpreted due to multiple meanings, but also has the potential to affect how we operate within our own community. One difficulty is that conscious awareness of this process is not a necessary condition for having it occur. We are not suggesting that researchers in any way mean to imply alternate meanings, only that it is a part of using words and we should be aware that it may have these negative implications.

6 DISCUSSION

Our study produced the following results: (1) For all five words, drawings by male participants were significantly more of men than women. (2) For “user” and “researcher”, drawings by female participants were significantly more of men than women. (3) For “person”, “participant”, and “designer”, drawings by female participants were about 50/50. (4) The gender “other” was used to describe only 14 images or participants. (5) Drawings from ethnic



Figure 7: Mean age of drawings and mean age of participants for each word condition.

groups tended to be of a person from their own ethnic group. (6) Drawings of “participants” were younger than themselves, and “researchers” older than themselves.

6.1 Gender

For all five conditions, drawings by male participants were significantly more of men than women. This finding suggests that men, when reading the words that most represent people in our literature, have a tendency to think of these words as being representative of males. Biases are present in all work we do. Significant effort is made to minimize these effects, but we suggest that the imagery implicated by individual people words is not fully known but can be remedied with some careful attention. Although this is a preliminary investigation into this phenomenon with a general population, we find the effect interesting enough to warrant discussion and future investigation in more context-driven environments with domain experts.

For the “user” and “researcher” conditions, drawings by females were significantly more of men than women. That this did not occur for the other three conditions suggests more inherent gender bias in the reception of these two words. There is also a cultural undertone that suggests that the people using technology and that researchers in general are thought of by both genders as being predominantly male. For the “person”, “participant”, and “designer” conditions, female participants drew images that were not significantly different than a 50% split between genders, meaning that, for women, these words may be more gender-balanced.

Our finding of gender imbalances may be reflective of an actual state of gender imbalance within the larger cultural context. Even though we have domain expertise, we are still part of the larger cultural landscape. We hypothesize that adding context may negate some of these effects, but suggest that being aware of our word use and the implications of the words we choose to describe people could have a direct impact on making the discourse of the

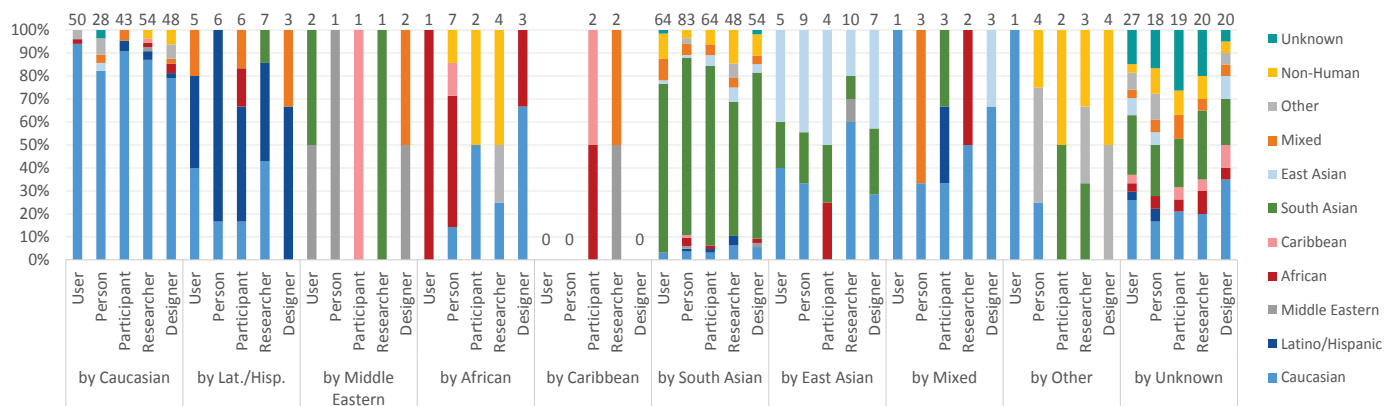


Figure 6: Ethnicity of participants vs. ethnicity of drawn images.

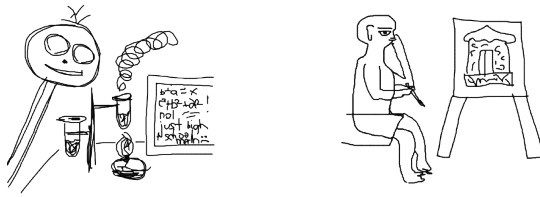


Figure 8: (left) A drawing of a researcher with lab equipment and (right) a designer sitting and painting at an easel.



Figure 9: (left) A “user” sitting at a computer and (right) a simplistic but still recognizable picture of a computer.



Figure 10: Drawings of drug users from the “user” condition.

HCI community more inclusive. It may be that consideration of the variety of complex personas from many different demographics results in more inclusive impressions of people.

It should also be noted that although only 14 data points (0.3%) used “other” to describe the participant’s or the image’s gender, the considerations for this gendered language description cannot and should not only be inclusive of males and females. Moreover, the categories provided in our questionnaire of “male”, “female”, and “other”, may be insufficient to describe gender, and perhaps both language in CHI papers and our methods of gender data collection should aim to be more sensitive to implying a binary [38].

Our hypothesis for this work was that words we use in HCI research to describe people were inherently gendered. Our data shows that, in words that we intend to be universal, there exists a gender bias in a general population of participants. “User” is our most common word for describing people in CHI papers and we should be aware that words like these can bring with them unintended implications. Future work should test these phenomena in context-driven situations to observe whether the effects hold.

6.2 Ethnicity

The majority of our drawings had participants that self-identified as Caucasian and South Asian, and the ethnicity of the drawing was typically the same as the participant’s. The data itself was not spread sufficiently across all ethnicities to draw conclusions, but future work could investigate whether people words are thought of as being more of one ethnicity than another. This would be a surprising result because of the diversity of the HCI community, but our data provides some (very weak) evidence that some non-Caucasian ethnicities would more frequently draw Caucasians.

6.3 Age

In the “researcher” condition, drawings were of someone older than participants, suggesting that this word not only holds a gender bias, but has age connotations as well. Moreover, in the “participant” condition, drawings were of people younger than participants, indicating a different kind of interpretation, which perhaps

indicates a belief that this group is less mature than one’s self. Further study would help elucidate the meaning of these findings.

7 STUDY LIMITATIONS

A limitation of our study is that we only tested the reception of the words by Mechanical Turkers in a context free environment. We recognize that the Mechanical Turk platform has several limitations; the integrity of our data is not guaranteed, and our sample is limited to Mechanical Turk workers (i.e., those that know of its existence). However, in past work collecting drawings from Mechanical Turk it has been found that “despite cultural, age, and geographic differences, the representations had some striking similarities in the way they conveyed the meanings of the words” [34], which indicates potential in providing an overall base line.

Also, since we investigated words without context, we want to be clear that these perceptions may not reflect those of a particular word in the context of a paper. For example, the word might even be used in the context of a gender discussion. To address possibilities of ambiguity around the drawings’ contents, we follow the suggestion to collect “more in-depth survey data to complement and support the image data captured” [34]. We think our approach provides interesting evidence that we hope will trigger considerable future work to further explore these issues. We are particularly interested in the contextual issue in our own future work.

8 FUTURE WORK

We do not have a prescriptive set of instructions on how to solve this problem. In the early stages of this work, we predicted that the word “user” in particular was more inherently biased and began the practice of replacing it with words like “person” or “human”, but our data suggests that this will not completely avoid the bias in question, so it is a partial solution at best. While the practice of substituting words such as “user” for the words “person” or “people” will still bring with it some male bias for a general population, it is perceived by females as being more equally distributed, and may therefore have the effect of being more inclusive. We thus suggest that substituting “person” for the word “user”, and (when appropriate) “designer” for the word “researcher” can begin to address these problems of gender imbalance at the point of reception. We believe the HCI community has always been forward thinking, and an active discussion of language choice in our literature, in light of this new evidence, can lead to greater equality. As a community, we need to increase our awareness of this issue to fuel creative solutions to this problem going forward.

While our study addresses how individual words used to describe people in CHI papers are received by non-experts, we did not connect actual CHI papers and researchers to these phenomena nor do we suggest that there is malicious intent present within the HCI community in using these words. Future work will focus on whether use of these words within the community follows similar patterns to our data, and will try to investigate the other side of this issue, unintentional gender biases in research. There are methodological questions that make these next steps challenging:

How do you choose a paper that is representative of the community as a whole?

How do you then choose a single paragraph?

With all of the varied work within the community does this phenomenon exist across the entire spectrum of work?

These are difficult questions and we ask the community to engage in ongoing debate with regards to equality and gender issues within HCI research. In our future work, we would also like to connect language to other work in the community that is more technological, such as how large screens level the gender playing field [31].

We wish to identify what the gender norms may be within our own field. This motivation aligns with feminist HCI, assisting with revealing unspoken values within our research methods. While some authors [2, 7, 10] have already brought attention to a problematic relationship between humans and technology through exploration of unconventional themes, we hope to support their illustrative work with our rich set of images and data.

9 CONCLUSION

Our Mechanical Turk study shows that non-gendered words used in CHI papers to describe people are received by general readers with a gender bias. In HCI, the idea of thinking about humans is already a central tenet of our work and we have highlighted how the terminology we use to describe people does not always reflect these design goals. We suggest that community members be aware that their work has the potential to be received in this way and that they should substitute the words “user” and “researcher”, with the words “person” and “designer” respectively when appropriate. Although this does not address the whole problem, which we think is culturally present outside of our work, it does take a positive step forward in trying to make the HCI community an inclusive one, sensitive to the needs and wants of all of our people.

REFERENCES

- [1] S. Bardzell, E. Churchill, J. Bardzell, J. Forlizzi, R. Grinter, and D. Tatar, “Feminism and interaction design,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, 2011, pp. 1–4.
- [2] S. Bardzell, S. Gross, J. Wain, A. Toombs, and J. Bardzell, “The significant screwdriver: care, domestic masculinity, and interaction design,” in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, 2011, pp. 371–377.
- [3] S. Bardzell and J. Bardzell, “Towards a feminist HCI methodology: social science, feminism, and HCI” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 675–684.
- [4] S. Bardzell, “Feminist HCI: taking stock and outlining an agenda for design,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 1301–1310.
- [5] L. Beckwith and M. Burnett, “Gender: An important factor in end-user programming environments?,” in *Visual Languages and Human Centric Computing, 2004 IEEE Symposium on*, 2004, pp. 107–114.
- [6] M. M. Burnett, L. Beckwith, S. Wiedenbeck, S. D. Fleming, J. Cao, T. H. Park, V. Grigoreanu, and K. Rector, “Gender pluralism in problem-solving software,” *Interact. Comput.*, vol. 23, no. 5, pp. 450–460, Sep. 2011.
- [7] L. Buttrick, C. Linehan, B. Kirman, and D. O’Hara, “Fifty shades of CHI: the perverse and humiliating human-computer relationship,” 2014, pp. 825–834.
- [8] J. Margolis and A. Fisher, *Unlocking the clubhouse: women in computing*. Cambridge, Mass.: MIT Press, 2002.
- [9] G. Kannabiran, S. Bardzell, and J. Bardzell, “Designing (for) desire: a critical study of technosexuality in HCI,” in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, 2012, pp. 655–664.
- [10] A. Light, “HCI as heterodoxy: Technologies of identity and the queering of interaction with computers,” *Interact. Comput.*, vol. 23, no. 5, pp. 430–438, Sep. 2011.
- [11] L. Hemphill, I. Erickson, D. Ribes, and I. Mergel, “Feminism and social media research,” 2014, pp. 319–322.
- [12] M. Muller, “Feminism asks the ‘Who’ questions in HCI,” *Interact. Comput.*, vol. 23, no. 5, pp. 447–449, Sep. 2011.
- [13] B. Vasilescu, A. Capiluppi, and A. Serebrenik, “Gender, Representation and Online Participation: A Quantitative Study,” *Interact. Comput.*, vol. 26, no. 5, pp. 488–511, Sep. 2014.
- [14] S. Padilla, T. S. Methven, D. W. Corne, and M. J. Chantler, “Hot topics in CHI: trend maps for visualising research,” 2014, pp. 815–824.
- [15] S. M. Dray, A. N. Peters, A. M. Brock, A. Peer, A. Druiin, S. Gitau, J. Kumar, and D. Murray, “Leveraging the Progress of Women in the HCI Field to Address the Diversity Chasm,” in *CHI ’13 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2013, pp. 2399–2406.
- [16] A. Forte, J. Antin, S. Bardzell, L. Honeywell, J. Riedl, and S. Stierch, “Some of All Human Knowledge: Gender and Participation in Peer Production,” in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, New York, NY, USA, 2012, pp. 33–36.
- [17] A. Dillon and C. Watson, “User analysis in HCI—the historical lessons from individual differences research,” *Int. J. Hum.-Comput. Stud.*, vol. 45, no. 6, pp. 619–637, 1996.
- [18] J. Grudin and J. Pruitt, “Personas, participatory design and product development: An infrastructure for engagement,” in *PDC 2002*, Malmo, Sweden, 2002, pp. 144–152.
- [19] J. Wajcman, *Feminism confronts technology*. University Park, Pa.: Pennsylvania State University Press, 1991.
- [20] T. Winograd, *Understanding computers and cognition: a new foundation for design*. Norwood, N.J.: Ablex Pub. Corp, 1986.
- [21] J. Butler, *Undoing gender*. New York ; London: Routledge, 2004.
- [22] S. Sontag, *On photography*. New York: Picador USA : Farrar, Straus and Giroux, 2001.
- [23] A. H. Brafman, *The Language of Drawings: A New Finding in Psychodynamic Work*. London: Karnac Books, 2012.
- [24] B. Türkcan, “Semiotic approach to the analysis of children’s drawings,” *Educ. Sci. Theory Pract.*, vol. 13, no. 1, pp. 600–607, 2013.
- [25] S. Wright, “‘I Mean, the Queen’s Fierce and the King’s Not’: Gendered Embodiment in Children’s Drawings,” *Int. J. Early Child.*, vol. 46, no. 3, pp. 391–406, Nov. 2014.
- [26] D. W. Chambers, “Stereotypic images of the scientist: The draw-a-scientist test,” *Sci. Educ.*, vol. 67, no. 2, pp. 255–265, Apr. 1983.
- [27] N. S. Dalton, “Neurodiversity & HCI,” in *CHI’13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 2295–2304.
- [28] B. Coleman, *Hello avatar: rise of the networked generation*. Cambridge, Mass: MIT Press, 2011.
- [29] T. A. M. Carnegie, “Interface as Exordium: The Rhetoric of Interactivity,” *Comput. Compos.*, vol. 26, no. 3, pp. 164–173, Sep. 2009.
- [30] P. Dourish, *Divining a digital future: mess and mythology in ubiquitous computing*. Cambridge, Mass: MIT Press, 2011.
- [31] M. Czerwinski, D. S. Tan, and G. G. Robertson, “Women take a wider view,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2002, pp. 195–202.
- [32] I. Chalmers, “People are ‘participants’ in research. Further suggestions for other terms to describe ‘participants’ are needed,” *BMJ*, vol. 318, no. 7191, p. 1141, Apr. 1999.
- [33] P. M. Boynton, “People should participate in, not be subjects of, research,” *BMJ*, vol. 317, no. 7171, p. 1521, Nov. 1998.
- [34] S. McMaster, “New Approaches to Image-based Research and Visual Literacy: A Pilot Project Report,” *New Horiz. Vis. Lit. Sel. Read. Int. Vis. Lit. Assoc.*, pp. 122–132, 2012.
- [35] C. Cohn, “Sex and death in the rational world of defense intellectuals,” *Signs*, pp. 687–718, 1987.
- [36] M. Schmettow, M. L. Noordzij, and M. Mundt, “An implicit test of UX: individuals differ in what they associate with computers,” in *CHI’13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 2039–2048.
- [37] “user, n.1.” *OED Online*. Oxford University Press, 2014.
- [38] M. Balarajan, M. Gray, and M. Mitchell, *Monitoring equality: developing a gender identity question*. 2011.