Gender-Inclusive Design: Sense of Belonging and Bias in Web Interfaces

Danaë Metaxa-Kakavouli[1], **Kelly Wang**[2], **James A. Landay**[1], **Jeff Hancock**[1]
[1]Stanford University, [2]Brown University
{metaxa, landay, jeff.hancock}@stanford.edu, kelly_wang@brown.edu

ABSTRACT
We interact with dozens of web interfaces on a daily basis, making inclusive web design practices more important than ever. This paper investigates the impacts of web interface design on ambient belonging, or the sense of belonging to a community or culture. Our experiment deployed two content-identical webpages for an introductory computer science course, differing only in aesthetic features such that one was perceived as masculine while the other was gender-neutral. Our results confirm that young women exposed to the masculine page are negatively affected, reporting significantly less ambient belonging, interest in the course and in studying computer science broadly. They also experience significantly more concern about others' perception of their gender relative to young women exposed to the neutral page, while no similar effect is seen in young men. These results suggest that gender biases can be triggered by web design, highlighting the need for inclusive user interface design for the web.

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Web-based Interaction

Author Keywords
Gender-inclusive interfaces; web interface psychology; inclusive design; ambient belonging

INTRODUCTION
User interfaces are ubiquitous, used for everything from ordering dinner to applying for jobs. These interfaces are used by a population as diverse as our society itself. How do we effectively design for such a varied audience? Subjective standards in interface design leave room for bias; without understanding and controlling for bias, users—especially those from marginalized identity groups—may experience feelings of exclusion or alienation in online environments.

Psychological biases have been well-studied in the context of women's participation in science, technology, engineering, and mathematics (STEM) [11]. One important factor impacting the representation of women in STEM fields is *ambient belonging*, the feeling of fitting in with a culture or community that is passively elicited by one's surrounding environment [4]. Previous work has shown that this sense of belonging is ascertained quickly and often unconsciously. Ambient belonging can be influenced by subtle cues in the environment, including the layout of a space or physical objects within the space. This has previously been demonstrated in the context of computer science education by surveying college undergraduate participants about their interest in computer science, administering the survey in a room either decorated to evoke stereotypes about computer science (i.e., stacks of soda cans, video games, and a Star Trek poster) or decorated neutrally (i.e., water bottles, nature pictures, and general magazines) [4].

Research on the impact of gender bias in software development and interaction design has shown that virtually all stages of the design process are vulnerable to bias [1, 2, 17]. Can the visual design of a web interface communicate these kinds of unconscious psychological cues and affect ambient belonging? And if so, what is the effect on its users? The literature on ambient belonging has shown these effects can take place not only in physical classrooms, but also in virtual reality representations of a classroom, suggesting that a digital medium can convey these psychological effects [3, 5]. At the same time, research in HCI has shown that men and women evaluate web interfaces differently. For instance, men show a preference for symmetrical website layouts, whereas women's judgments of websites are not affected by symmetry [14]. Studies have also found that women may prefer colorful websites more than men, as well as interfaces with higher-visual complexity, and that this difference is not due to innate differences but rather varies with culture [10].

In addition to preferences in aesthetics, other perceptions of web interfaces can be driven by sociological factors. Linguistic cues can signal ostracism; using masculine gender-exclusive language (such as "he," "him," or "guys" to refer to a group) decreases women's perceived sense of belonging and motivation [12]. Cues such as images or other visual media in an interface can trigger social identity threat and decrease sense of belonging and willingness to participate in or join a community [8]. More broadly, reminding individuals of stigmatization they face through a lack of social connectedness, such as underrepresentation in a field, leads to feelings of belonging uncertainty [15].

Nonetheless, visual cues in a web interface may be too subtle or transient to trigger changes in ambient belonging. A web

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.

ISBN 978-1-4503-5620-6/18/04 ...$15.00.
https://doi.org/10.1145/3173574.3174188

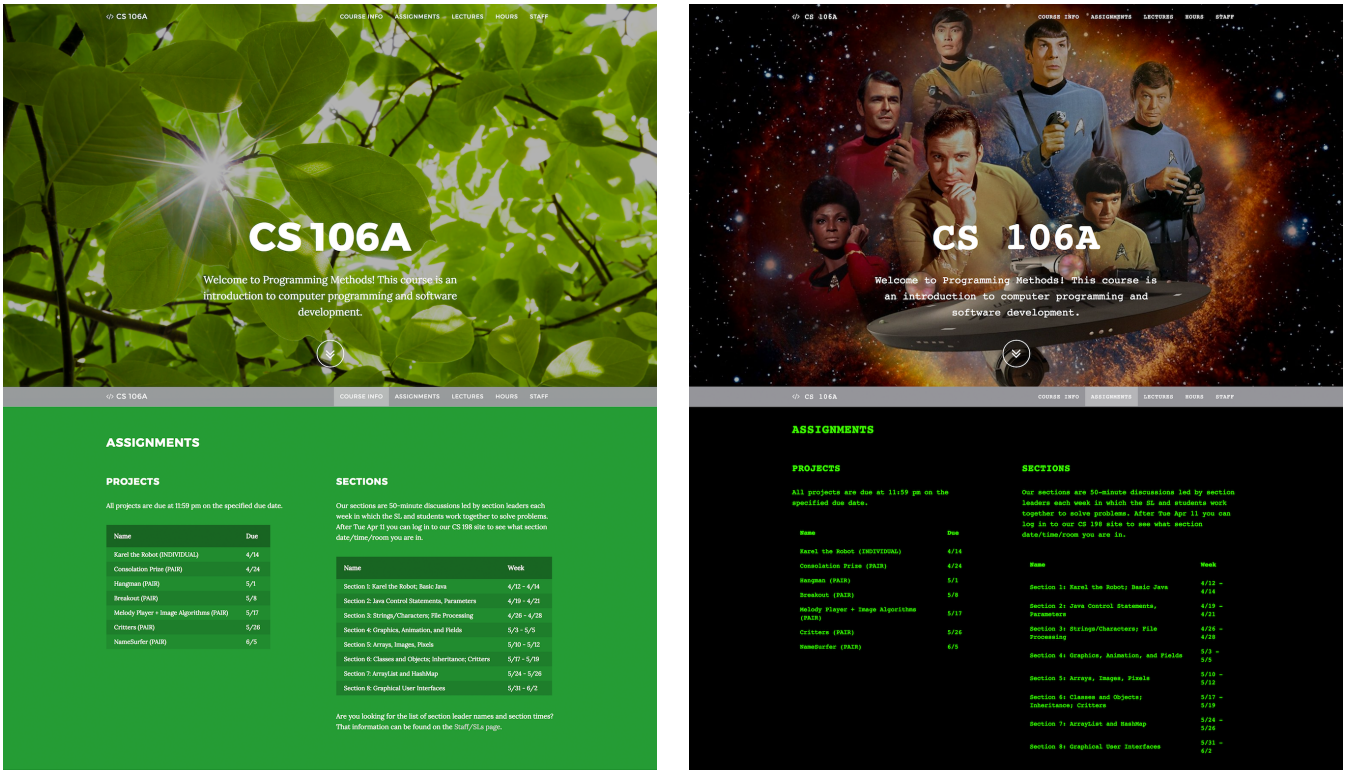

Figure 1. The banner and other content of the gender-neutral interface used nature imagery (left), while the masculine interface included Star Trek imagery and styling evocative of a computer terminal (right).

interface is substantially less immersive, for example, than a virtual reality environment that can provide a sense of presence with multisensory stimuli (e.g., [6, 3]).

To test whether web interface designs communicate ambient belonging we adopted a variation on the procedure used to examine ambient belonging in virtual reality spaces [4, 3]. We exposed college-age web users to one of two web pages for an introductory computer science course. One interface was designed to evoke gender-related stereotypes, while the other was gender-neutral. Having reviewed a website, participants responded to questions corresponding to six measures: their interest in the course, sense of belonging, anticipated success in the course, self-confidence in their computing skills, interest in studying computer science in the long-term, and anxiety about how others in the course would perceive their gender.

Based on prior research in social psychology on ambient belonging, we hypothesized:

H1: Women exposed to a website evoking stereotypical masculinity will feel less sense of belonging and be negatively impacted on measures related to participation in computer science relative to women viewing a neutral website, or men exposed to either website.

Our findings confirm this hypothesis for all six measures. The results highlight that ambient belonging can be triggered by web interface design. Given the prominent gender and racial disparities in computer science, the ubiquity of web interfaces, and the benefits of inclusive design for all users, this work is a first step towards developing a design framework for gender inclusivity.

METHOD

We present participants with one of two webpages for an introductory course, one stereotypically masculine and one gender-neutral, before administering a survey to evaluate six measures including anticipated success in the course, intention to enroll in the course, and sense of belonging in the context of the course.

Webpage Stimuli

Using content from a real introductory computer science course offered in the previous year at Stanford University, two web pages were developed, each containing identical information and page layouts. These pages were "skinned" differently—they differed only in aesthetic aspects: the background image, font, and colors. As shown in Figure 1, the gender-neutral webpage featured a nature theme; the backdrop image was a photograph of tree leaves, background color throughout was green, and font was white in a Helvetica-like style. The masculine webpage instead included a large Star Trek background image, and the rest of the page was a black background with lime green typewriter-style Courier font to imitate the appearance of a computer terminal. Aside from an image of the (male) professor at the bottom of each webpage, neither page contained any other names or images.

Pretest

To verify the gender and domain stereotypes each page was meant to evoke, we recruited a sample of crowdworkers from

Amazon Mechanical Turk, using platform settings to restrict them to 18-25 years of age to be closer in age to the target population. We asked workers to rate the two background images (tree leaves and Star Trek) according to how stereotypical of computer scientists each was, on a scale from 1 (not at all) to 7 (extremely), as well as how masculine or feminine each was, on a scale from 1 (very feminine) to 7 (very masculine).

Sixty crowdworkers (28 female) participated in the pretest. We conducted ANOVAs on the effects of gender and image on perceived CS-stereotypicality, as well as the effects of gender and image on perceived masculinity. The 2 (Condition) x 2 (Gender) ANOVA for CS-stereotypicality revealed a main effect of condition on CS-stereotypicality, $F(1, 59) = 178.586$, $p < 0.001$; the Star Trek themed website was rated higher ($M = 5.90$, $SD = 1.22$) in CS-stereotypicality than the nature themed website ($M = 2.13$, $SD = 1.24$). There was also a main effect of condition on masculinity, $F(1, 59) = 67.925$, $p < 0.001$; the Star Trek themed website was rated higher ($M = 5.23$, $SD = 0.973$) in masculinity than the nature themed website ($M = 3.40$, $SD = 1.03$). The background images had a clear and strong effect: the stereotypical webpage was perceived as significantly more stereotypical of computer scientists and more masculine than the neutral one, whereas the neutral imagery was perceived as gender-neutral and minimally CS-stereotypical, as intended.

Participants

Participants were recruited through Amazon Mechanical Turk (as is commonly done for psychology experiments [9]), using platform features to select for participants between 18 and 25 years of age. To control for possible cultural differences in interface preferences [10], we also restricted the sample to individuals within the United States. Based on the results of a power analysis of Cheryan 2009, we recruited 111 participants (55 self-identified as female, 56 as male, none identifying as non-binary) [4]. One participant was excluded from the analysis for failing initial attention checks, indicating he had not reviewed the webpage. According to self report, racial demographics were 67% white, 12.5% African American, 10.7% Asian, 3.6% Hispanic/Latinx, and 6.2% multiracial or other. At 67%, most participants were current students, with 55% reporting being undergraduate students, 8% graduate students, and 1.8% high school students.

Procedure

Participants were asked to review one of the two course webpages and answer survey questions about six main measures regarding their sense of ambient belonging, perceptions of the class and the discipline of computer-science, and gender-related anxiety. To prevent bias, participants were told that the study was intended to "learn more about young people's attitudes towards studying Computer Science."

After agreeing to participate, participants were redirected to a Qualtrics survey, which randomly assigned them to review one of the two course pages. Random assignment ensured that women in the neutral condition would not significantly differ from the women in the masculine one, without collecting data on CS intentions prior to the experiment (as this could have compromised the manipulation). According to standard experimental methods, individual differences before the manipulation should be equal across the two samples with random assignment.

Participants were then asked three questions concerning webpage content to verify they had viewed the page, as well as a question about who they thought had created the page to probe for suspicion about the study. One male respondent in the masculine condition was excluded from analysis due to incorrect answers to the viewing verification test. Participants then continued to the rest of the survey, answering questions via the main measures described below. Lastly, participants were asked to provide demographic information about gender, race, age, and education level.

Measures

Based on prior literature, we tracked six measures on a 1-7 scale (where 1 meant "not at all" and 7 meant "extremely"). Each was calculated as a composite of 2-4 items in our survey:

1. *Enrollment Intentions*: average of three items asking about how much the participant wanted to take the course ($\alpha = 0.95$);
2. *Ambient Belonging*: average of three items asking how much the participant felt they would fit in or belong in the course ($\alpha = 0.92$);
3. *Anticipated Success*: average of two items asking how they anticipated doing if they were to take the course ($\alpha = 0.90$);
4. *Self Confidence*: average of two items asking how confident they felt in their computer science abilities ($\alpha = 0.88$);
5. *Future CS Study Intentions*: average of three items asking how interested in learning computer science and programming the participant felt ($\alpha = 0.90$);
6. *Gender-related Anxiety*: average of three items asking whether the participant felt that their gender would impact their performance, or would impact other students' interpretation of it (e.g., "How much do you worry that if you performed poorly in this course, others would attribute your poor performance to your gender?") ($\alpha = 0.83$).

These measures were drawn directly from prior work on sense of belonging, so their validity has been established in prior contexts [4, 3].

In addition to these continuous measures, we asked participants to answer an open-ended response question about enrollment intentions: "Would you take this course? Why or why not?" As a manipulation check after these measures, we surveyed participants on perceived masculinity and perceived CS-stereotypicality of the course.

RESULTS

Analyzing our data with planned contrast analysis, we find widespread support for our hypothesis. Relative to all other participants, women in the masculine condition were negatively impacted across all six measures.

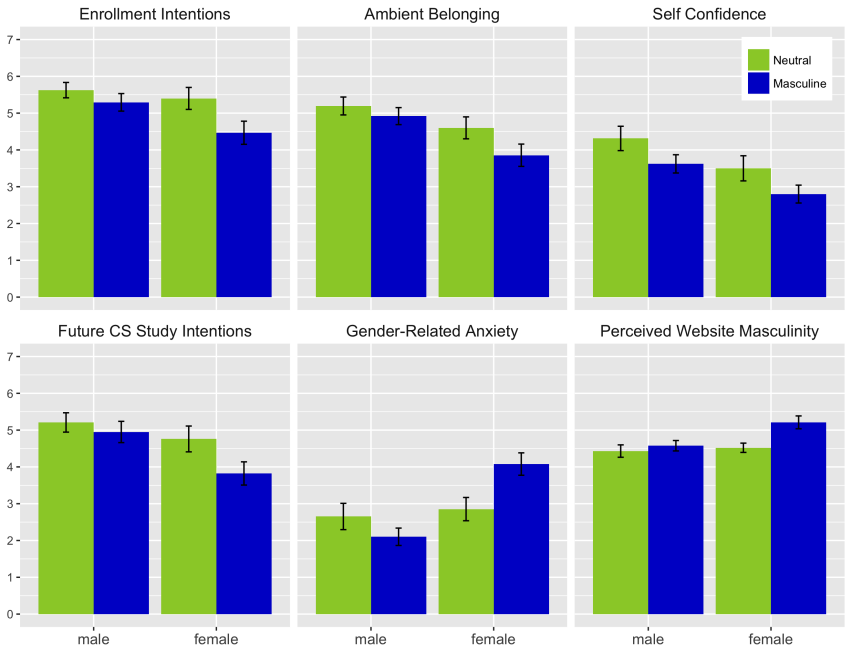

Figure 2. Male versus female responses by website condition, neutral (in green, left) versus masculine (in blue, right) for each of the primary dependent variables and the manipulation check (perceived website masculinity). Higher scores represent higher response for that dependent variable.

Randomization Success. In order to allow us to draw causal conclusions about the results, participants were assigned to view one of the two interfaces randomly. Our results verified that this randomization procedure was successful; women in the two conditions were not significantly different in age, race, or education level.

Analytic Approach. For all six quantitative measures we performed a planned contrast analysis using contrast coefficients calculated from the results reported in Cheryan 2011 to directly investigate our hypothesis [3]. These coefficients were -3 and 1 for women in the masculine and neutral conditions, and 1 and 1 for men in the masculine and neutral conditions, coding the hypothesis that women in the masculine condition should be negatively impacted relative to all other participants.

We find evidence across all measures that confirms our hypothesis: women in the masculine website condition were negatively impacted by that interface relative to women in the neutral condition or men in either condition, and this negative effect was not observed for the male participants. In contrast with members of all other groups, women in the masculine condition were less likely to want to enroll in the course, $t(108) = 3.20$, $p = 0.001$, $d = 0.62$; were less likely to feel they would belong in that class, $t(108) = 3.47$, $p < 0.001$, $d = 0.67$; anticipated lesser success if they were to take the course, $t(108) = 2.41$, $p = 0.01$, $d = 0.47$; reported lower levels of confidence in their technical abilities relative to others who might take the introductory course, $t(108) = 3.18$, $p = 0.001$, $d = 0.61$; expressed less interest in learning computer science or programming in the long term, $t(108) = 3.33$, $p < 0.001$, $d = 0.64$; and were much more anxious about the way their gender would be perceived in the course, $t(108) = -4.57$, $p < 0.001$, $d = -0.88$.

Our manipulation check shows similar results to the main measures. As Figure 2 (bottom-right) shows, male participants rated both web pages as roughly equal in masculinity and nearly gender-neutral. Female participants rated the neutral page similarly, but in contrast, women in the masculine condition rated the page as significantly more masculine.

While our statistical test focused on women in the masculine condition relative to others, Figure 2 reveals that men were also scoring slightly lower on each of our primary measures, but these differences did not achieve statistical significance.

Participant Open Responses

We coded the open response question, "Would you take this class? Why or why not?" for positive, negative, or equivocating language. Between two reviewers, inter-rater agreement was very high ($\kappa = 0.94$). As reported in Table 1, fewer participants of both genders expressed intention to enroll in the masculine condition course; this effect was especially pronounced for women, who were over 20% less likely to want to enroll. In contrast, the gender-neutral interface resulted in higher enrollment intention rates for all participants.

Most respondents indicated interest in the class due to the material and the importance of programming skills in today's workforce, and did not mention the interface itself, underscoring that web design is often processed unconsciously. Among those whose comments included opinions about the website, the masculine design was perceived generally unfavorably by women, who found it "juvenile"; one woman reported that it made her "question the professor's judgment". Men,

	Men		Women	
	Neutral	Masculine	Neutral	Masculine
% Intent	87.5	78.1	80	63.3
% No intent	12.5	21.9	20	36.7

Table 1. Coded responses for open-ended intention to enroll question by condition and gender.

meanwhile, were more likely to project positive associations onto the presumed webpage creator, describing the professor as "fun," "interesting," and "relatable". These responses are consistent with the gender effect on ambient belonging.

In contrast, all participants responded positively to the neutral interface. Participants described the course page as "interesting and unique," imbuing them with confidence since it was "set up neatly and easy to follow. It makes me feel like I know what I am getting into." The aesthetics and ease of use were perhaps a pleasant surprise to participants with less optimistic expectations for computer science. One female participant who viewed the neutral course page not only found it "very attractive and easy to navigate," and mistakenly (despite seeing the professor's name and photograph on the page) intuited that the professor was a woman, further increasing her interest in enrolling.

DISCUSSION

The results reported in this paper suggest that webpage design can unconsciously impact downstream user psychology, including attitudes and intentions. Our findings have implications for web users, who increasingly use the internet for purposes ranging from social engagement and community-building to education and employment. Biases that reduce a user's sense of belonging could unconsciously discourage her from taking STEM courses, applying for a job, or voicing her opinion online.

Through aesthetic choices, as well as language and imagery, web designers have the power to unconsciously communicate strong signals to their users that may have unintended consequences for user engagement in the future. In the accessibility literature, design choices have repeatedly been shown to discriminate against disabled users [16]. Systemic technological biases against rural users and users of lower socioeconomic status have also been shown in the context of social media algorithms [7] and the sharing economy [13]. The consequences of biased design are particularly severe for vulnerable and marginalized populations. In contrast, since improvements need not be made at the expense of the dominant group's perception—indeed men in our study were slightly more positively affected by the gender-inclusive design across all or our measures— inclusive design is truly inclusive: it benefits all users.

Limitations

While these results suggest purely aesthetic differences can have powerful unconscious impacts on web users, we focus on one particular (albeit important) type of bias, specifically in the context of computer science education. Additionally, for logistical as well as ethical reasons, this study is done in a purely hypothetical context; asking participants to evaluate a course as if they were considering enrolling in it, rather than experimenting with students enrolling in an actual class, may affect the generalizability of our results. We also acknowledge that further work must be done to examine issues of web interface psychology in a fully intersectional framework. The sample of this study is lacking in racial diversity, which may interact with gender biases in multidimensional ways, and we examine the experience of gender as it relates to women, with participants self-reporting only two genders. Lastly, we note that due to the limitations of our survey distribution our participants are young people aged 18-25, rather than using exclusively undergraduates and filtering out computer science students, as in the prior literature.

Future Work

Given the importance of inclusivity in web interface design, future research must conduct further experiments to identify the most important dimensions in website design for a wider range of psychological biases, such as stereotype threat and social identity threat, that may inhibit the inclusiveness of a web interface. Future work can rely on extant social psychological theory as a guide to identify attributes of interest (including colorfulness, complexity, imagery, and language [10, 14, 12, 8]) that can be used to empirically determine the relative weight of such design factors for various biases in real computer science webpages. The long-range goals of such an effort should be to develop a comprehensive understanding of design principles for gender inclusivity.

CONCLUSION

Much like there are well-understood principles from psychology governing user engagement or behavior in design, we argue for developing a comprehensive understanding of the psychology of inclusivity and bias in web interface design. This work begins to develop such a framework by examining ambient belonging in the context of gender bias. Through a controlled experiment exposing participants to masculine and gender-neutral interfaces, we demonstrate that gender-biased design choices can significantly negatively impact women users of those web interfaces, whereas gender-neutral design was perceived positively by individuals of all genders. This highlights the potential consequences of non-inclusive design, and the importance of developing a systemic understanding of bias for inclusivity in web design.

ACKNOWLEDGMENTS

We would like to thank Sapna Cheryan for providing research materials and feedback that made this work possible, as well as members of the Stanford HCI Group and other friends who pilot-tested and proof-read this work.

REFERENCES

1. Shaowen Bardzell. 2010. Feminist HCI: Taking Stock and Outlining an Agenda for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1301–1310. DOI: **http://dx.doi.org/10.1145/1753326.1753521**

2. Margaret Burnett, Anicia Peters, Charles Hill, and Noha Elarief. 2016. Finding Gender-Inclusiveness Software

Issues with GenderMag: A Field Investigation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2586–2598. DOI: http://dx.doi.org/10.1145/2858036.2858274

3. Sapna Cheryan, Andrew N. Meltzoff, and Saenam Kim. 2011. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education* 57, 2 (2011), 1825–1835. DOI: http://dx.doi.org/10.1016/j.compedu.2011.02.004

4. Sapna Cheryan, Victoria C. Plaut, Paul G. Davies, and Claude M. Steele. 2009. Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology* 97, 6 (2009), 1045–1060. DOI: http://dx.doi.org/10.1037/a0016239

5. Katheryn R. Christy and Jesse Fox. 2014. Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education* 78, Supplement C (2014), 66 – 77. DOI:http://dx.doi.org/https://doi.org/10.1016/j.compedu.2014.05.005

6. James J. Cummings and Jeremy N. Bailenson. 2016. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2 (2016), 272–309. DOI: http://dx.doi.org/10.1080/15213269.2015.1015740

7. Isaac Johnson, Connor McMahon, Johannes Schöning, and Brent Hecht. 2017. The Effect of Population and "Structural" Biases on Social Media-based Algorithms: A Case Study in Geolocation Inference Across the Urban-Rural Spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1167–1178. DOI: http://dx.doi.org/10.1145/3025453.3026015

8. Mary C. Murphy, Claude M. Steele, and James J. Gross. 2007. Signaling Threat. *Psychological Science* 18, 10 (2007), 879–885. DOI:http://dx.doi.org/10.1111/j.1467-9280.2007.01995.x PMID: 17894605.

9. Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010), 411–419. https://ssrn.com/abstract=1626226

10. Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 11–20. DOI: http://dx.doi.org/10.1145/2556288.2557052

11. Steven J. Spencer, Claude M. Steele, and Diane M. Quinn. 1999. Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology* 35, 1 (1999), 4 – 28. DOI:http://dx.doi.org/https://doi.org/10.1006/jesp.1998.1373

12. Jane G. Stout and Nilanjana Dasgupta. 2011. When He Doesn't Mean You: Gender-Exclusive Language as Ostracism. *Personality and Social Psychology Bulletin* 37, 6 (2011), 757–769. DOI: http://dx.doi.org/10.1177/0146167211406434 PMID: 21558556.

13. Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Toward a Geographic Understanding of the Sharing Economy: Systemic Biases in UberX and TaskRabbit. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 21 (April 2017), 40 pages. DOI: http://dx.doi.org/10.1145/3058499

14. Alexandre N. Tuch, Javier A. Bargas-Avila, and Klaus Opwis. 2010. Symmetry and Aesthetics in Website Design: It's a Man's Business. *Comput. Hum. Behav.* 26, 6 (Nov. 2010), 1831–1837. DOI: http://dx.doi.org/10.1016/j.chb.2010.07.016

15. G. M. Walton and G. L. Cohen. 2007. A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology* 92, 1 (2007), 82–96.

16. Sue Watling. 2011. Digital exclusion: coming out from behind closed doors. *Disability & Society* 26, 4 (2011), 491–495. DOI: http://dx.doi.org/10.1080/09687599.2011.567802

17. Gayna Williams. 2014. Are You Sure Your Software is Gender-neutral? *interactions* 21, 1 (Jan. 2014), 36–39. DOI:http://dx.doi.org/10.1145/2524808