Sex and Gender Analysis Improves Science and Engineering

Cara Tannenbaum[1]*, Robert P. Ellis[2]*, Friederike Eyssel[3]*, James Zou[4]*,
Londa Schiebinger[5^], corresponding author: schiebinger@stanford.edu

[1] Scientific Director, Institute of Gender and Health, Canadian Institutes of Health Research, Professor in the Faculties of Medicine and Pharmacy at the Université de Montréal, Canada. [2] NERC Industrial Innovation Fellow – Sustainable Aquaculture, University of Exeter, UK . [3] Center of Excellence Cognitive Interaction Technology, Universität Bielefeld, Germany. [4] Biomedical Data Science and (by courtesy) of Computer Science and of Electrical Engineering, Stanford University, US; Chan-Zuckerberg Biohub, US. [5] History of Science; Gendered Innovations in Science, Health & Medicine, Engineering, and Environment, Stanford University, US. *Co-first authors.
[^] email: schiebinger@stanford.edu

[Preface]

The goal of sex and gender analysis is to promote rigorous, reproducible and responsible science. Incorporating sex and gender analysis into experimental design has enabled advancements across many disciplines, such as improved heart disease treatment and insights into the societal impact of algorithmic bias. Here we discuss the potential for sex and gender analysis to foster scientific discovery, improve experimental efficiency and enable social equality. We provide a roadmap for sex and gender analysis across scientific disciplines and call upon researchers, funding agencies, peer-reviewed journals and universities to coordinate efforts to implement robust methods of sex and gender analysis.

[Main text]

Integrating sex and gender analysis into the design of research, where relevant, can lead to discovery and improved research methodology. A deeper understanding of the genetic and hormone-mediated basis for sex differences in immunity, for example, promises insights into novel cancer immunotherapies[1]. Evidence that facial recognition systems misclassify gender more often for darker-skinned women compared with lighter-skinned men has led to refinements in computer vision[2]. Understanding sex-based responses to climate change allows better modeling of demographic change among marine organisms and the down-stream impacts for humans[3,4]. Sex or gender analysis can be critical to the interpretation, validation, reproducibility and generalisability of research findings.

The documented importance of sex and gender analysis in research has underwritten policy change at major funding agencies. New policies have been implemented at the Canadian Institutes of Health Research (2010), European Commission (2014), US National Institutes of Health (2016), German Research Foundation (2020), among others. Concurrently, peer-review journals have implemented editorial guidelines to evaluate the rigour of sex and gender analysis as one criterion among many when selecting manuscripts for publication. The goal is to increase transparency, promote inclusion and reset the research default to carefully consider sex and gender, where appropriate.

In this perspective, we discuss how incorporating sex and/or gender analysis into research can improve reproducibility and experimental efficiency, help reduce bias, enable social equality in scientific outcomes and foster opportunities for discovery and innovation. From highlighted examples, we extract decision-tree roadmaps for researchers across disciplines. We consider the limits to sex and gender analysis and offer recommendations to researchers and granting agencies on how to move the field forward. Throughout this perspective we explore how integrating sex and gender analysis into research design has the potential to offer new perspectives, pose new questions and, importantly, enhance social equalities by ensuring that research findings are applicable across the whole of society.

Sex refers to the biological attributes that distinguish organisms as male, female, intersex (ranging from 1:100 to 1:4500 in humans, depending on the criteria used[127,128]) and hermaphrodite (over 30% of non-insect nonhuman animals[129]). In biology, sex describes differences in sexual characteristics within plants or animals that go beyond their reproductive functions to affect appearance, physiology or neuroendocrine, behavioural and metabolic systems. In engineering, sex includes anthropometric, biomechanical and physiological characteristics that may impact the design of products, systems and processes.

Gender refers to psychological, social and cultural factors that shape attitudes, behaviours, stereotypes, technologies and knowledge. Gender includes three related dimensions. *Gender norms* refer to spoken and unspoken rules in the family, workplace, institution or global culture that influence individuals. G*ender identity* refers to how individuals and groups perceive and present themselves within specific cultures. And *gender relations* refer to power relations between individuals with different gender roles and identities[130].

Sex and gender interact in unexpected ways. Pain, for example, exhibits biological sex differences in the physiology of signaling. Pain also incorporates sociocultural components in how symptoms are reported by women, men and gender-diverse people, and how physicians understand and treat pain according to a patient's gender[131].

Reproducibility and efficiency

Reproducibility is a requirement for scientific excellence. One important reason for a lack of reproducibility in experimentation is inconsistency in methodologic reporting, which varies widely across disciplines from biology to chemistry, human-robot interaction, medicine, physics, psychology and beyond[5,6]. Sex- and gender-specific reporting is still limited in a range of scientific disciplines. In preclinical microbiology and immunology, a review of published studies using primary cells from diverse animal species (that is, humans and nonhuman vertebrates) revealed that the majority failed to report the sex of donors from which the cells were isolated[7,8]. In marine science, a review of experimental ocean acidification studies showed that only 3.9% statistically assessed sex-based differences, while only 10.5% accounted for possible sex effects by assessing females and males independently[9]. Similarly, in ecotoxicology, a review of omics studies showed that while most reported sex, only 23% (5 of 22) examined the omics response of each sex to a toxicant[10]. In social robotics, the notion of robot gender, gender-stereotypical domains and their interaction with user gender has only recently become a target of scientific inquiry[11]. A lack of transparency in reporting sex and gender-related variables makes it difficult to reproduce experiments where these variables affect experimental results.

Disaggregating the data

Analysing experimental results by sex and/or gender is critical for improving accuracy and avoiding misinterpretation of data. The common practice of pooling the response of females and males or women and men can mask sex differences. Take for example copepods, small aquatic crustaceans. Failure to disaggregate and analyse data by sex leads to the false interpretation that elevated pCO_2 has no significant biological impact on respiration (FIG 1). Disaggregating data by sex, by contrast, reveals important sex-based differences in the respiration rate of females and males in response to elevated pCO_2 levels[12].

The same is true for human research. Pooling data yields inexact results. In a human-robot experiment, humans were asked to touch or point to anatomical regions on a 23-inch NAO robot. When asked to touch accessible regions (such as feet and hands), there was little reaction; when asked to touch inaccessible regions (such as the robot's plastic buttocks or genitals), human participants had increased heart rate and blood pressure[13]. Equal numbers of women and men were recruited for the experiment, but data were not disaggregated or analysed separately. We know that norms for human social touch vary by participants' age, gender identity and cultural backgrounds—as well as social context and purpose of the touch[14]. If results are not stratified by these variables, opportunities will be missed to provide clearer insights into their impact on human judgments and behaviour.

Variability, sample size, interactions

Scientists have erroneously assumed that females should be excluded from experiments because of the variable nature of the data caused by the reproductive cycle[15]. In fact, research has shown that males exhibit equal or greater variability than females for specific traits due to testosterone fluctuations and other factors, such as animal group caging[16]. Analysis of microarray datasets reveals similar findings that females are no more variable than males on measures of gene expression in both mice and humans[17]. Accounting for sex and gender enhances the likelihood of detecting meaningful effects, elucidating unexplained variability and potentially reducing the overall number of experiments required for determining trends or making groundbreaking discoveries. In a meta-analysis of 11 proteomics datasets from humans and mice, sex explained 13.5% of the observed variation of complex protein abundances and stoichiometry, even more than other environmental factors, such as diet[18].

On the surface it may appear that including females and males, women and men in a study necessitates doubling the number of experimental subjects. This is not always the case: More efficient experimental designs can incorporate both sex and gender while maintaining control over variance[19]. Factorial designs, where two experimental factors with multiple levels are tested, and data are collected across all possible combinations of factors and levels, are one such strategy. This enables the effect of each factor to be tested, in addition to the interaction between the factor levels. For such cases, sample sizes may need to be slightly increased by 14–33% to account for the extra parameter being estimated, but they do not need to be doubled, according to sample size calculators that consider interaction effects[20,21]. Analysing data by sex or gender enhances the likelihood of detecting meaningful effects that, in turn, help reduce confounding, increase reproducibility and reduce the cumulative number of experiments required.

Numerous interactions, such as the interaction of the sex of the research subjects, may also influence outcomes. In animal research, female and males are often studied separately in the lab. Yet in the wild, the sexes coexist—and their interactions can influence research results. Recent studies of longevity in the nematode, *Caenorhabditis elegans*, found that the presence of males accelerated aging in individuals of the opposite sex (in this case, hermaphrodites). In other words, hermaphrodites died at a younger age in the presence of males. Researchers traced this “male-induced demise” to pheromones released by males and found it could occur without mating and required only that the hermaphrodites be exposed to the medium in which males were once present[22]. Ignoring such interactions potentially leads to an incomplete understanding of species viability in the wild.

Other interactions focus on the sex of the researcher and potential impacts on research subjects. In social science, it has long been understood that the simple presence of an observer can alter the response of the observed, whether in the field or in laboratory experiments[23]. In quantum mechanics, the act of observation can alter the phenomenon by collapsing the wave function. Similarly, in animal research, experimenter sex can impact research outcomes. A pain study showed that rats and mice did not exhibit pain when a male experimenter was present, as opposed to a woman present in the room or an empty room. Both female and male mice displayed this “male observer” effect, but female mice did so to a greater degree. Researchers determined that the mice responded to male-associated olfactory stimuli[24]. The authors suggest that not controlling for experimenter sex throws into question much prior pain research.

One could proliferate these types of interactions crucial to excellence and discovery in research. One final interaction of note is researcher gender and the type of research conducted. Two new studies provide compelling evidence that in biomedical, clinical and public health research, women in leading positions (first and last author) are more likely to analyse sex and gender in published research[25,26]. This dynamic has not yet been replicated in other research fields, such as computer science, engineering or the physical sciences.

Opportunities for discovery

Ignoring sex and gender analysis can lead to inaccuracies, research inefficiency and difficulties generalizing results. Integrating sex and gender analysis into research can open the door to discovery and innovation.

A prevalent assumption is that sex is a binary trait determined genetically before birth, and that it is fixed across the lifespan[27,28]. Commonly used model organisms in biology, such as mice, *Drosophila melanogaster* and *Caenorhabditis elegans*, reinforce these perceptions. Sex, however, can be highly plastic, and studying interactions with the environment, for example, has led to new understandings of the mechanisms of sex determination within the context of global climate change.

A population's sex ratio influences its resilience to environmental disturbances. The mechanism that determines sex is thus a vital consideration for predicting population viability[29,30]. Enhancing sex analysis capability in a growing number of species, across a wide range of settings, may increase our ability to accurately model climate change impacts.

Climate impacts in the ocean
For species reliant on temperature for sex determination, rapid global warming poses a risk to sex ratios and demographic stability. Turtles are the most widely studied group where sex is determined by temperature. The ability to differentiate female and male juvenile green sea turtles using novel non-invasive endocrine markers has enabled the discovery that global warming negatively skews population sex ratios. Turtles originating from warmer northern Great Barrier Reef sites, for instance, exhibit a female sex ratio of 99%, while cooler southern sites maintain a 68% female juvenile ratio[3]. Similarly, in fish species displaying temperature-dependent sex determination, warming is projected to result in male-skewed populations (up to 3:1 male to female) by the end of the century[28]. Such changes in sex balance can limit mate choice, reduce reproductive capacity and undermine population viability[31,32].

Warming does not occur in isolation, but against a backdrop of anthropogenic disturbances across marine environments, which include habitat destruction, pollution and overfishing. Primary sex differentiation has been shown to respond to a diverse range of these environmental factors in a growing number of species. Hypoxia, for example, has resulted in a higher ratio of males in zebrafish[33]. Similarly, ocean acidification results in 16% more female oysters over a single generational cycle[4], while increased aquatic pH results in more female cichlids[34]. What is increasingly apparent is that alterations in sex ratio—in either direction—will result in populations less resilient to further disturbance and potentially lead to demographic collapse[35,36].

Social organization can also influence population sex ratios. Non-human animals do not have a "gender"; the term gender is reserved for human societies and interactions. Nonetheless, numerous non-human species develop elaborate social organisations, and sex determination can be socially mediated. Clownfish, for example, are protandrous hermaphrodites (they mature as male; some change to female) that live in a strict social hierarchy with a single dominant and highly fecund female at the top who mates with a single large male in the social group; all remaining individuals remain immature juveniles. Removal of the alpha female results in the alpha male changing sex to female, with all subordinates moving up a rung in the social hierarchy[37]. Conversely, many grouper species, a subfamily of long-lived and high-value reef species, are protogynous (they mature as female; some change to male). Large dominant males control groups of females with strong sexual selection resulting in these males achieving the greatest reproductive success. These sequentially hermaphroditic individuals consistently produce more offspring and enjoy greater reproductive success after they have changed sex[36]. Thus, the timing and the direction of sex change are critical species-specific factors determining demographic resilience to disturbance in sex-changing organisms.

A mechanistic understanding of these and other ecologically significant sex-based responses enables more accurate modeling of the impacts of environmental variability (for example, climate change) or anthropogenic disturbance (for example, overfishing) at a population level. Sex-specific impacts of climate change stressors on sex determination mechanisms, particularly in commercially important species, have potentially significant implications for humans with respect to aquatic food production, ecosystem services and biodiversity. Incorporating sex analysis into marine science, and the natural sciences more widely, enhances research excellence and opportunities for discovery.

Targeted human therapeutics

Sex analysis also reveals new opportunities for human drug development. In the areas of pain and depression, the discovery of sex differences in molecular pathways has signaled new directions for targeted therapies[38]. Pain research that uses experimental mouse models of chronic pain shows that male and female mice withdraw from painful stimuli in a similar fashion, except when the contribution of microglial cells are inhibited[39]. Microglia are specialised immune cells located exclusively in the spinal cord and the brain. Microglial cell inhibitors reduce pain sensing in male, but not female mice, underscoring the potential importance of sex-dependent molecular pain pathways. Mouse models of depression also show sexually divergent networks in the brain with distinct patterns of stress-induced gene regulation in males and females[40]. These findings have now been reproduced in human postmortem tissue and may contribute insights into why males and females with major depressive disorder respond differently to antidepressant treatment[40].

Although sex-specific dosages are rare, a few already exist. Such is the case for the drug desmopressin that activates vasopressin receptors in the kidney to regulate water homeostasis. Because the gene for the arginine vasopressin receptor is found on the X chromosome in a region likely to escape X-inactivation, women are more sensitive to the antidiuretic effects of vasopressin than men, who have only one X chromosome and therefore only one copy of the vasopressin receptor gene per cell[41]. As a result, older women taking desmopressin are more likely to experience reduced sodium concentration in the blood than men, which corresponds to a higher incidence of side effects in women. To avoid unnecessary harm, both the European Union and Canada have recommended lower dosages for older women taking desmopressin.

Even cancer immunotherapy is benefitting from a deeper understanding of previously recognized genetic and hormone-mediated sex differences in immunity. Patients with melanoma and lung cancer, who are treated with checkpoint inhibitors, respond differently based on their sex, with a higher proportion of male than female patients achieving successful remission[1]. Designed to outsmart cancer cells' defense tactics, checkpoint inhibitors stimulate NK (natural killer) immune cells to attack tumour cells. NK cells are sensitive to estrogen and testosterone, which may explain these observed sex differences. Understanding the underlying mechanisms will enable us to fine-tune future therapies[42].

We expect to see an exponential rise in biomedical discoveries now that new computational biology and statistical genetics software facilitates the exploration of X-related expression in complex diseases[43]. Until recently, sex chromosomes were excluded from a majority of genome-wide association studies because of difficulty distinguishing the active from the inactive X chromosome in females, and because of a mismatch in chromosomal size—the X chromosome has 1,669 known genes and the smaller Y chromosome contains only 426[44,45]. Including sex chromosomes in genome-wide association studies, as well as including and analyzing adequate numbers of female and male cells, tissues, animals and humans in research, will broaden our understanding of why women and men are affected differently by certain diseases and how we can tailor life-saving therapies to their specific needs.

Engineering equality

An often neglected but critical component of engineering is to understand the broader social impacts of the technology being developed and to ensure that the new technology enhances social equality by benefiting diverse populations. Human bias and stereotypes can be perpetuated, even amplified, when researchers fail to consider how human preferences and assumptions may consciously or unconsciously be built into science or technology. Gender norms, ethnicity and other biological and social factors shape and are shaped by science and technology in a robust cultural feedback loop[46]. This section discusses examples from product design, artificial intelligence (AI) and social robotics to illustrate how sex and gender analysis can enhance excellence in engineering.

Designing safer products

When products are designed based on the male norm, there is a risk that women and people of smaller stature will be harmed. Motor vehicle safety systems provide one such example. Because male drivers have historically been overrepresented in traffic data, seatbelts and airbags have been designed and evaluated with a focus on the typical male occupant with respect to anthropometric size, injury tolerance and mechanical response of the affected body region. When national automotive crash data from the U.S. were analysed by sex between 1998 and 2008, data revealed that the odds for a belt-restrained female driver to sustain severe injuries were 47% higher than those for a belt-restrained male driver involved in a comparable crash, after controlling for weight and body mass[47]. The subsequent introduction of a virtual female car crash dummy allowed mathematical simulations to account for the effect of acceleration on sex-specific biomechanics, highlighting the need to add a medium-sized female dummy model to regulatory safety testing[48,49]. Beyond automotive safety systems, the importance of anthropometric characteristics, such as the carrying angle of the elbow or the shape and size of the human knee, can be used to guide sex-specific design for artificial joints, limb prostheses and occupational protective gear[50,51].

Reducing gender bias in AI

Alarming examples of algorithmic bias are well documented[52]: When translating gender-neutral language related to Science, Technology, Engineering and Mathematics (STEM) fields, Google Translate defaults to male pronouns[53]. When photos depict a man in the kitchen, automated image captioning algorithms systematically misidentify the individual as a woman[54]. As AI becomes increasingly ubiquitous in everyday lives, such bias, if uncorrected, can amplify social inequities. Understanding how gender operates within the context of the algorithm helps researchers make conscious decisions about how their work functions in society.

Since World War II, medical research has been submitted to stringent review processes aimed at protecting subjects from harm. AI, which has the potential to impact human life at scale, has yet to be so carefully examined. Numerous groups have articulated "principles" for human-centered AI. These include, most importantly, the UN Human Rights Framework that consists of internationally agreed upon human rights laws and standards, as well as the "Asilomar AI Principles", "AI at Google: Our Principles", "Partnership on AI", etc. What we lack are *mechanisms* for technologists to put these principles into practice. Here we delve into a few of such rapidly developing mechanisms for AI.

A first challenge in algorithmic bias is to identify when it is appropriate for an algorithm to use gender information. In some settings, such as the assignment of job ads, it might be desirable for the algorithm to explicitly ignore the individual's gender as well as features such as weight which may correlate with gender but not be directly related to job performance. In other applications, such as image/voice recognition, it might be desirable to leverage gender characteristics to achieve the best accuracy possible across all subpopulations. To date, there is no unified definition of algorithmic fairness[55–57], and the best approach is to understand the nuances of each application domain, make transparent how algorithmic decision-making is deployed and appreciate how bias can arise[58].

Training data is a source of potential bias in algorithms. Certain subpopulations, such as darker-skinned women, are often underrepresented in the data used to train machine-learning algorithms, and efforts are underway to collect more data from such groups[2]. To highlight the issue of underrepresented subpopulations in machine learning data, researchers have designed "nutrition labels" to capture metadata about how the dataset was collected and annotated[59–61]. Useful metadata should summarise statistics on, for example, the sex, gender, ethnicity and geographic location of the participants in the dataset. In many machine learning studies, the training labels are collected via crowdsourcing, and it is also useful to provide metadata about the demographics of crowd labelers.

Another approach to evaluate gender bias in algorithms is counterfactual analysis[62]. Consider Google Search, where men are five times more likely than women to be offered ads for high-paying executive jobs[63]. The algorithm that decides which ad to show inputs features about the individual making the query and outputs a set of ads predicted to be relevant. The counterfactual would test the algorithm *in-silico* by changing the gender of each individual in the data and then studying how predictions change. If simply changing an individual from “woman” to “man” systematically leads to higher paying job ads, then the predictor is, indeed, biased.

Work on debiasing word embeddings is another example of counterfactual analysis[64]. Word embeddings associate each English word with a vector of features so that the geometry between the feature vector captures semantic relations between the words. It is widely used in practice for applications such as sentiment analysis[65], language translation[66] and analysis of electronic health records[67]. Bolukbasi et al. showed that gender stereotypes—for example, men are more likely to be computer scientists—are manifested in the feature vectors of the corresponding words. Whether this association between man and computer is problematic depends on the application of the features. To test, gender-neutral word features were created. For each downstream application, the counterfactual analysis was then performed: The application was run twice, once using the original word features, and once using the gender-neutral features. If the outcome changes, the algorithm is sensitive to gender. In some applications, for example, job search, it might be preferable to use gender-neutral features.

An alternative approach to quantify and reduce gender bias in algorithms is called multi-accuracy auditing[68,69]. In standard machine learning, the objective is to maximise the overall accuracy for the entire population, as represented by the training data. In multi-accuracy, the goal is to ensure that the algorithm achieves good performance not just in the aggregate but also for specific subpopulations, for example, “elderly Asian man”, “Native American woman”, etc. The multi-accuracy auditor takes a complex machine learning algorithm and systematically identifies if the current algorithm makes more mistakes for any subpopulation. In a recent paper, Kim et al. audited the neural network used for facial recognition and identified specific combinations of artificial neurons responding to African American women’s images that are responsible for the misclassifications[70].

The auditor also suggests improvements when it identifies such biases[71]. While achieving equal accuracy across all the demographic groups may not always be feasible, these auditing techniques improve the transparency of the AI systems by quantifying how its performance varies across race, age, sex and intersections of these attributes.

These are just a few of the specific techniques computer scientists are developing to promote gender fairness in algorithms. Some, such as data checks, are relevant across all disciplines that amass and analyse big data. Others are specific to machine learning now being widely deployed across broad swathes of intellectual endeavour from the humanities to the social sciences, biomedicine and judicial systems. In all instances, it is important to be completely transparent where and for what purpose AI systems are used, and to characterize the behaviour of the system with respect to sex and gender[72].

Combatting stereotypes

Analysing gender in software systems is one issue; configuring gender in hardware, such as social robots, is another, and the focus of this section. Until recently, robots were largely confined to factories. Most people never see or interact with these robots; they do not look, sound or behave like humans. But engineers are increasingly designing robots to assist humans as service robots in hospitals, elder care facilities, classrooms, homes, airports and hotels. The field of social human-robot interaction (HRI) examines, among other things, when and how "gendering" robots, virtual agents or chatbots might enhance usability while, at the same time, considering when and how to avoid oversimplifications that may reinforce potentially harmful gender stereotypes[73].

Machines are, in principle genderless. Gender, however, is a core social category in human impression formation that is readily applied to nonhuman entities[74]. Thus, users may consciously or unconsciously gender machines as a function of anthropomorphizing them, even when designers intend to create gender-neutral devices[75–78].

Anthropomorphizing technologies may help users engage more effectively with them, which poses the question: Are there benefits to tapping into the power of social stereotypes by building gender into virtual agents[79–83], chatbots[84] or social robots[11,85,86]? For example, if roboticists deploy female carebots in female-typical roles, such as nursing, do users better comply with the robot's requests to take daily medication or to exercise? Does gendering robots or virtual agents facilitate interaction or boost objective outcomes like performance[11, 80–91]? Will personalising robots or chatbots by gender increase consumer acceptance and, even, sales figures? Systematic empirical research is needed to address these open research issues.

What features lead humans to gender a robot? To date, experimental research designed to analyse robot gender has manipulated gender in a number of ways, including: 1) by choosing a male or female name to label the robot[87–92], 2) by color-coding the robot[93,94], 3) by manipulating visual indicators of gender (for example, face, hairstyle or lip color[94,95]), 4) by adding a male vs. female voice, or low vs. high pitch, respectively[87–92,94,96,97], 5) by designing a gendered personality[87,98], or 6) by deploying robots in gender-stereotypical domains, such as a male-voiced robot for security and a female-voiced robot in a healthcare role[95]. Other aspects, such as movement or gesture, that may potentially gender a robot still require empirical research[85,86].

But there are dangers here. As soon as designers or users assign gender to a machine, stereotypes follow. Designers of robots and artificial intelligence do not simply create products that reflect our world, they also (perhaps unintentionally) reinforce and validate certain gender norms considered appropriate for men, women or gender-diverse people. Eliciting gendered perceptions of technologies implies actively designing human gender biases, including binary constructions of gender as "male" vs. "female", into machines. From a social psychological viewpoint, this can contribute to stereotypical gender norms in society[95]. Even though this might not seem to be relevant from an engineering point of view, social psychological research would suggest that a robot with a female appearance, for example, may perpetuate ideas of women as nurturing and communal, traits stereotypically associated with women[95]. Thus, a female robot may be deemed socially warm and particularly suitable for stereotypically female tasks, such as elderly care, or it might be openly sexualized and objectified as revealed in abusive commentary on video clips of female robots in recent qualitative research[99]. Similarly, virtual personal assistants with female names, voices and stereotypical, submissive behaviours, such as Siri or Alexa, represent heteronormative ideas about females and thereby indirectly contribute to the discrimination of women in society[100,101]. An interesting development in this regard is the genderless voice, Q, recently developed in Denmark to overcome such bias[102].

Questions in this area abound. How, for example, do user attributes, that is, age or gender, interact with different robot design features? How do robots enhance or harm real-world attitudes and behaviours related to social equality? How does built robot "gender" elicit different responses across cultures? More experimental, laboratory and longitudinal field research is needed to test whether, and how, a machine's gendered, gender-diverse or gender-neutral appearance or behaviour impacts human affect, cognition and behaviour. It is likely that even social robots designed to be genderless or gender-neutral elicit gender attributions due to the relatively automatic nature of anthropomorphizing humanoid robots. It is also likely that when potential end users are offered the option to select a digital assistant's gender, their choice will be driven by their own gender identity and gender-related attitudes and stereotypes. Addressing these research questions and issues remain important to shed light on the psychological, social and ethical implication of implicit or explicit design choices for novel technologies.

Developing technologies that enhance, or at least do not harm, social equality will require novel configurations of researchers. Much lip service has been paid to the need for interdisciplinary research, consisting of humanists, legal experts, technologists and social scientists, especially in the fields of human-centered AI. The historic development of universities, however, has artificially separated human knowledge into disciplines over the course of the nineteenth and twentieth centuries that may not support current research needs. Research institutions now need to develop new, robust mechanisms to bring together social analysis and engineering in way that rigorously address the emerging needs of society[103].

Pathway to integration into research design

To reap the full potential of sex and gender analysis for discovery and innovation, it is important to integrate sex and gender analysis, where relevant, into the design of research from the very beginning. Much science and engineering is path-dependent: once research has been designed, it becomes difficult to change. It is also important to understand that sex and gender are categories of analysis or "variables" (or controls) to be incorporated into the research process and need not be the main focus of the research. Nor will sex and gender analysis be relevant to all research. As the decision trees, analyzing sex (FIG 2) and analyzing gender (FIG 3) indicate, where researchers have consider sex and/or gender but judge this analysis not relevant for a specific hypothesis, they may rule it out. Moreover, if researchers expect sex or gender to be important but find no significant differences, this may represent a result worthy of publication. Reporting where sex or gender sameness, overlap or no difference is found may represent an important finding.

In this perspective, we have highlighted the need and promise for designing sex and gender analysis into research through specific case studies and examples. From these, we extracted key considerations for analysing sex (FIG 2) and for analysing gender (FIG 3). These are generic recommendations that work across disciplines. But more related study is needed in the next five years. First, through interdisciplinary work, researchers need to sharpen and standardize generic approaches to sex and gender analysis that generalize across fields. Second, through discipline-specific work, researchers need to craft state-of-the-art analytics for study design and data analysis in their own subfields. The European Commission is currently funding an expert group that seeks to tailor sex and gender methods of analysis to field-specific protocols[104].

Future challenges

We do not yet have results for sex and gender analysis in the physical sciences, such as basic chemistry, pure physics, geology or astronomy. Much work has analysed gender gaps in participation and gender bias in the culture of these fields, but attention has yet to turn to how the research itself may respond to gender analysis. As research in the physical sciences becomes more applied, sex and gender analysis become more relevant—for example, in the chemistry of aerosols, sex differences govern rates of inhalation, and gender differences influence rates of exposure[105].

A number of methodological challenges remain for the field of sex and gender analysis itself. While advances have been made in methods for analysing sex, we lack non-invasive methods of sex determination in numerous non-model organisms, where sexual morphological dimorphism is not easily detected. Technological advances through the development of novel genetic[106], metabolomic[107] and endocrine[3] markers of organism sex are needed for non-model species at all stages of development, an endeavor that will be aided by the innovation and increased affordability of omics approaches. Attention will also need to be paid to translation of evidence from animal species to humans, since in many cases, molecular sex differences observed in humans may not be mirrored in nonhuman mammals[108].

While sex as a biological variable in the sciences and engineering is increasingly well understood[109], the same cannot be said for gender as a cultural variable. Gender is complex and multidimensional (Facebook introduced 58 gender categories in 2014[110]) and applications in technical fields often require collaboration with social scientists to understand the relevant aspects of gender for specific projects. Even in health research, we lack systematic measures for assessing how gender relates to health because gender does not reduce easily to variables that can be manipulated statistically. Two recent studies have attempted to remedy this. The first employed a binary gender index (masculinity vs. femininity) constructed from seven variables and found that the incidence of recurrence and death 12 months after diagnosis in young adults of acute coronary syndrome was associated with gender and specifically not biological sex[111]. A second study under development at Stanford University seeks to better capture the multidimensionality of gender by identifying theoretically robust gender-related variables relevant for health research. This study is based on U.S. data, and new variables tailored to specific cultural settings need to be identified. Developing measures of gender is clearly an area where more research is needed.

Other methodological challenges include going beyond the binary—female and male, women and men—in both sex and gender analysis. Take for instance the Gender API algorithm that allows social scientists to understand, for example, gender differences in research patterns. The algorithm identifies only binaries: female/male; woman/man. In the US 0.6% of the population, or nearly 2 million people, identify as transgender[112], and more than fifteen countries offer a third sex category on legal documents, birth certificates, passports and the like. Research needs to keep pace with social change. Or take the lack of research addressing how hermaphroditic animals respond to environmental change. In simultaneous hermaphrodites, where reproductively mature individuals have both male and female gametes, there is a need to consider the role of male or female tissues in determining whole organism response. Conversely, in sequential hermaphrodites that change sex, there is a need to consider whether an organism responds as a female or a male to environmental stress during the sex change process, given that this process is dynamic, with behavioral, endocrine and genetic systems switching sex on dramatically different timescales[113].

Additional challenges include accounting for other social variables, such as age, race and geographic location, and how these intersect with sex and/or gender. Sex or gender cannot be isolated from other characteristics, and we need model systems and intersectional methods to understand these interrelationships[114]. An intersectional approach in human research underscores the importance of unmasking and rectifying overlapping and interdependent systems of discrimination often built into knowledge, programs and policies. Benefits for global health, for example, will only be achieved when unbiased decision-making about resources takes into account the lived experiences of women with multiple identity characteristics who simultaneously suffer from race, class, education, economic and cultural power imbalance in accessing food and water, digital technology and healthcare services[115].

Science policy

Policy is one driver of discovery and innovation that can enable sex and gender analysis in science and technology. To push forward rigorous sex and gender analysis, interlocking policies need to be implemented by three pillars of academic research: funding agencies, peer-reviewed journals and universities (FIG 4).

Government-led funding agencies have taken the lead by asking applicants to explain how sex and gender analysis is relevant to their proposed research, or to explain that it is not (for a list of agencies and policies, see supplement, section 1). The Canadian Institutes of Health Research showed robust uptake after mandating applicants to declare whether sex and/or gender were accounted for in proposals and to justify exclusion in 2010. Their evaluation revealed that from 2010–2011 the proportion of funded proposals incorporating sex and/or gender analysis nearly doubled[116,117].

The second pillar, peer-reviewed journals, have developed editorial policies requiring sex or gender analysis to ensure excellence in papers selected for publication (for a list of journals and policies, see supplement, section 2). Uptake has been swift in health and medicine. *The Lancet*, for example, adopted such guidelines in 2016, followed quickly by the International Committee of Medical Journal Editors[118]. *Cell* Press's Structured, Transparent, Accessible Reporting Methods (STAR) has required transparent reporting of the sex distribution of donor cells, also since 2016. Importantly, the widely adopted Sex and Gender Equity in Reporting (SAGER) guidelines require that data be disaggregated by both sex and gender[119]. While biomedical journals have moved rapidly, we are not aware of any engineering or computer science conferences or journals with such guidelines.

Pillars one and two need the support of a third pillar: universities. Both granting agencies and journals may have policies in place, but researchers and evaluators by and large lack expertise in sex and gender analysis. The European Commission, with policies in place since 2014, has found that only one in seven funded research proposals incorporated sex and gender analysis and has correlated this low proportion to an "absence of training on gender issues"[120]. Similarly, an analysis of animal research in the neurosciences showed that in 2014 only about 14 percent of peer-reviewed articles considered sex as a biological variable[121].

Universities need to step up and incorporate sex and gender analysis as a conceptual tool into the sciences and engineering curricula. Numerous universities offer gender analysis in the humanities and social sciences, but not in core natural science and engineering courses. Efforts have been made in medicine—the Charité in Berlin, Germany, for instance, has successfully integrated sex and gender analysis throughout all six years of medical training from early basic science to later clinical modules[122]. But this is a rare example, and universities must do more to prepare the scientific workforce for the future.

Several initiatives have endeavoured to fill this gap. Gendered Innovations, a global, collaborative project initiated from Stanford University in 2009 and supported by the European Commission and the U.S. National Science Foundation, has developed practical methods of sex and gender analysis for natural scientists and engineers, and provided case studies as concrete illustrations of how sex and gender analysis lead to discovery and innovation[123]. The World Health Organization has developed a gender responsive assessment tool[124]. The Organization for the Study of Sex Differences has advanced sex and gender analysis methods in the life and health sciences[125]. The Canadian Institutes of Health Research (CIHR) have developed online training modules for integrating sex and gender analysis into biomedical research[126]. These initiatives should now be mainstreamed into university education.

Much work remains to be done to systematically integrate sex and gender analysis into relevant domains of science and technology—from strategic considerations for establishing research priorities to guidelines for establishing best practices in formulating research questions, designing methodologies and interpreting data. To make real headway in the next decade, researchers, funding agencies, peer-reviewed journals and universities need to coordinate efforts to develop and standardize methods of sex and gender analysis.

But eyes have been opened, and by integrating sex and gender analysis into their work, researchers can enhance excellence and social responsibility in science and engineering.

References

1. Conforti, F. et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol.* **19**, 737–746 (2018).
2. Buolamwini, J. & Gebru, T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proc. Mach. Learn. Res.* **81**, 1–15 (2018). **This paper demonstrates that commercial gender identification algorithms misclassify darker-skinned females at a higher rate than the rest of the population, which is an example of how algorithmic bias intersects with gender and race.**
3. Jensen, M. P. et al. Environmental warming and feminization of one of the largest sea turtle populations in the world. *Curr. Biol.* **28**, 154–159 (2018).
4. Parker, L. M. et al. Ocean acidification but not warming alters sex determination in the Sydney rock oyster, *Saccostrea glomerata*. *Proc. R. Soc. B* **285**, http://dx.doi.org/10.1098/rspb.2017.2869 (2018).
5. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
6. American Society of Cell Biology Survey on Reproducibility http://www.ascb.org/wp-content/uploads/2015/11/final-survey-results-without-Q11.pdf (2014).
7. Shah, K., McCormack, C. E. & Bradbury, N. A. Do you know the sex of your cells? *Am. J. Physiol. Cell Physiol*. **306**, C3–C18 (2014). **This review demonstrates that the sex of cells used in experiments can impact the cell's biology and provides a table outlining the sex of cell lines that have appeared in *AJP-Cell Physiology* in recent years.**
8. Potluri, T., Engle, K., Fink, A. L., Vom Steeg, L. G. & Klein, S. L. Sex reporting in preclinical microbiological and immunological research. *MBio* **8**, http://dx.doi.org/10.1128/mBio.01868-17 (2017).
9. Ellis, R. P. et al. Does sex really matter? Explaining intraspecies variation in ocean acidification responses. *Biol. Lett.* **13**, http://dx.doi.org/10.1098/rsbl.2016.0761 (2017).
10. Bahamonde, P. A., Feswick, A., Isaacs, M. A., Munkittrick, K. R. & Martyniuk, C. J. Defining the role of omics in assessing ecosystem health: perspectives from the Canadian environmental monitoring program. *Environ. Toxicol. Chem.* **35**, 20–35 (2016).
11. Nomura, T. Robots and gender. *Gend. Genome* **1**, 18–25 (2017). **This paper presents an overview of research on the role of sex and gender in the context**

of human-robot interaction research and summarizes the state-of-the art in this area from an engineering perspective.

12. Cripps, G., Flynn, K. J. & Lindeque, P. K. Ocean acidification affects the phyto-zoo plankton trophic transfer efficiency. *PLoS One* **11**, https://doi.org/10.1371/journal.pone.0151739 (2016). **This paper discusses how sex analysis reveals significant effects of ocean acidification on the respiration rate in marine copepods, with males and females showing differential baseline respiration rates and responding to elevated CO_2 in opposing directions.**
13. Li, J., Ju, W. & Reeves, B. Touching a mechanical body: tactile contact with intimate parts of a humanoid robot is physiologically arousing. *J. Hum. Robot Interact.* **6**, 188–130 (2017).
14. Suvilehto, J. T., Glerean, E., Dunbar, R. I. M., Hari, R. & Nummenmaa, L. Topography of social touching depends on emotional bonds between humans. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13811–13816 (2015).
15. Becker, J. B., Prendergast, B. J. & Liang, J. W. Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biol. Sex Differ.* **7**, https://doi.org/10.1186/s13293-016-0087-5 (2016).
16. Prendergast, B. J., Onishi, K. G. & Zucker, I. Female mice liberated for inclusion in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* **40**, 1–5 (2014).
17. Ioth, Y. & Arnold, A. P. Are females more variable than males in gene expression? Meta-analysis of microarray datasets. *Biol. Sex Differ*. **6**, https://doi.org/10.1186/s13293-015-0036-8 (2015).
18. Romanov, N. et al. Disentangling genetic and environmental effects on the proteotypes of individuals. *Cell* **16**. 1308–1318 (2019).
19. Beery, A. K. Inclusion of females does not increase variability in rodent research studies. *Curr. Opin. Behav. Sci*. **23**,143–149 (2018).
20. Buch, T. et al. Benefits of a factorial design focusing on inclusion of female and male animals in one experiment. *J. Mol. Med.* **97**, 871–877 (2019).
21. Miller, L. et al. Considering sex as a biological variable in preclinical research. *FASEB J.* **31**, 29–34 (2017).
22. Maures, T. J. et al. Males shorten the life span of C. elegans hermaphrodites via secreted compounds. *Science* **343**, 541–544 (2014).
23. Chapman, C. D., Benedict, C. & Schiöth, H. B. Experimenter gender and replicability in science. *Sci. Adv.* **4**, https://doi.org/10.1126/sciadv.1701427 (2018).
24. Sorge, R. E. et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* **11**, 629–632 (2014).
25. Nielsen, M. W., Andersen, J. P., Schiebinger, L. & Schneider, J. W. One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nat. Hum. Behav.* **1**, 791–796 (2017).
26. Sugimoto, C. R., Ahn, Y. Y., Smith, E., Macaluso, B. & Larivière, V. Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. *Lancet* **393**, 550–559 (2019).
27. Ainsworth, C. Sex redefined. *Nature* **518**, 288–291 (2015).

28. Bachtrog, D. et al. Sex determination: why so many ways of doing it? *PLoS Biology* **12**, e1001899.
29. Ospina-Alvarez, N. & Piferrer, F. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS One* **3**, https://doi.org/10.1371/journal.pone.0002837 (2008).
30. Munday, P. L., Buston, P. M. & Warner, R. R. Diversity and flexibility of sex-change strategies in animals. *Trends Ecol. Evol.* **21**, 89–95 (2006).
31. Parker, K. A direct method for estimating northern anchovy, Engraulis mordax, spawning biomass. *Fish. Bull.* **78**, 541–544 (1980).
32. Barneche, D. R., Robertson, D. R., White, C. R. & Marshall, D. J. Fish reproductive-energy output increases disproportionately with body size. *Science* **360**, 642–645 (2018).
33. Shang, E. H. H., Yu, R. M. K. & Wu, R. S. S. Hypoxia affects sex differentiation and development, leading to a male-dominated population in zebrafish (Danio rerio). *Environ. Sci. Technol.* **40**, 3118–3122 (2006).
34. Oldfield, R. G. Genetic, abiotic and social influences on sex differentiation and the evolution of sequential hermaphroditism. *Fish Fish.* **6**, 93–110 (2005).
35. Kindsvater, H. K., Reynolds, J. D., Sadovy de Mitcheson, Y. & Mangel, M. Selectivity matters: rules of thumb for management of plate-sized, sex-changing fish in the live reef food fish trade. *Fish fish.* **18**, 821–836 (2017).
36. Benvenuto, C., Coscia, I., Chopelet, J., Sala-Bozano, M. & Mariani, S. Ecological and evolutionary consequences of alternative sex-change pathways in fish. *Sci. Rep.* **7**, https://doi.org/10.1038/s41598-01afri7-09298-8 (2017). **This paper demonstrates that the direction of sex-change in sequentially hermaphroditic fish is a critical variable determining demographic stability and resilience, with significant implications concerning the resilience of these populations to anthropogenic disturbance (for example, overfishing).**
37. Casas, L. et al. Sex change in clownfish: molecular insights from transcriptome analysis. *Sci. Rep.* **6**, https://doi.org/10.1038/srep35461 (2016).
38. Labonté, B. et al. Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).
39. Sorge, R. E. et al. Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nat. Neurosci*. **18**, 1081–1083 (2015).
40. Tannenbaum, C. & Day, D. Age and sex in drug development and testing for adults. *Pharmacol. Res.* **121**, 83–93 (2017).
41. Juul, K. V., Klein, B. M., Sandström, R., Erichsen L. & Nørgaard, J. P. Gender difference in antidiuretic response to desmopressin. *Am. J. Physiol. Renal. Physiol*. **300**, F1116–F1122 (2011).
42. Giefing-Kröll, C., Berger, P., Lepperdinger, G. & Grubeck-Loebenstein, B. How sex and age affect immune responses, susceptibility to infection and response to vaccination. *Aging Cell* **14**, 309–321 (2015).
43. Gao, F. et al. XWAS: a software toolset for genetic data analysis and association studies of the X chromosomes. *J. Hered.* **106**, 666–671 (2015).

44. Wise, A. L., Gyi, L. & Manolio, T. A. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.* **92,** 643–647 (2013).
45. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019). **This review discusses new techniques for sex analysis in genome-wide association studies, epigenetic studies and X-inactivation.**
46. Schiebinger, L. *Women and Gender in Science and Technology*, 4 vols. (Routledge, London, 2014).
47. Bose, D., Segui-Gomez, M. & Crandall, J. R. Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk. *Am. J. Public Health* **101**, 2368–2373 (2011).
48. Linder, A., Holmqvist, K. & Svensson, M. Y. Average male and female virtual dummy model (BioRID and EvaRID) simulations with two seat concepts in the Euro NCAP low severity rear impact test configuration. *Accid. Anal. Prev.* **114**, 62–70 (2018).
49. Linder, A. & Svedberg, W. Review of average sized male and female occupant models in European regulatory safety assessment tests and European laws: gaps and bridging suggestions. *Accid. Anal. Prev.* **127**, 156–162 (2019).
50. Falys, C. G., Schutkowski, H. & Weston, D. A. The distal humerus—a blind test of Rogers' sexing technique using a documented skeletal collection. *J. Forensic Sci.* **50**, 1289–1293 (2005).
51. Conley, S., Rosenberg, A. & Crowninshield, R. The female knee: anatomic variations. *J. Am. Acad. Orthop. Surg.* **15**, S31–S36 (2007).
52. Zou, J. & Schiebinger, L. AI can be sexist and racist—it’s time to make it fair. *Nature* **559**, 324–326 (2018).
53. Prates, M., Avelar, P. & Lamb, L. Assessing gender bias in machine translation—a case study with google translate. *Neural Comput. Appl.* https://doi.org/10.1007/s00521-019-04144-6 (2019).
54. Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K. W. Men also like shopping: reducing gender bias amplification using corpus-level constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (ed. Palmer, M., Hwa, R. & Riedel, S.) 2979–2989 (ACL, Stroudsburg, PA, 2017).
55. Corbett-Davies, S. & Goel, S. The measure and mismeasure of fairness: a critical review of fair machine learning. Preprint at https://arxiv.org/abs/1808.00023 (2018).
56. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214–226 (ACM, New York, 2012). **This paper develops a mathematical framework to study the fairness notion that similar individuals should be treated similarly by algorithms.**
57. Zemel, R., Wu, Y., Swersky, K., Pitassi, T. & Dwork, C. Learning fair representations. *Proc. Mach. Learn. Res.* **28**, 325–333 (2013).
58. Barocas, S. & Selbst, A. D. Big data's disparate impact. *Calif. Law Rev.* **104**, 671–732 (2016). **This paper studies what it means for algorithms to create**

disparate impact and to discriminate through the lens of American antidiscrimination law, and it discusses the legal challenges involved.

59. Gebru, T. et al. Datasheets for datasets. Preprint at https://arxiv.org/abs/1803.09010 (2018).
60. Holland, S., Hosny, A., Newman, S., Joseph, J. & Chmielinski, K. The dataset nutrition label: a framework to drive higher data quality standards. Preprint at https://arxiv.org/abs/1805.03677 (2018).
61. Bender, E. M. & Friedman, B. Data statements for NLP: toward mitigating system bias and enabling better science. Preprint at https://openreview.net/forum?id=By4oPeX9f (2018).
62. Kusner, M., Loftus, J., Russell, C. & Silva, R. Counterfactual fairness. In *Advances in Neural Information Processing Systems 30* (ed. Lee, D. D., von Luxburg, U., Garnett, R., Sugiyama, M. & Guyon, I.) 4066–4076 (Curran, Red Hook, NY, 2017).
63. Datta, A., Tschantz, M. C. & Datta, A. Automated experiments on ad privacy settings. *Proc. Privacy Enhancing Technol. 2015* **1**, 92–112 (2015).
64. Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V. & Kalai, A. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in Neural Information Processing Systems 29* (ed. Lee, D. D., von Luxburg, U., Garnett, R., Sugiyama, M. & Guyon, I.) 4349–4357 (Curran, Red Hook, NY, 2016).
65. Maas, A. L. et al. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (ed. Lin, D., Matsumoto, Y. & Mihalcea, R.) 142–150 (ACL, Stroudsburg, PA, 2011).
66. Zou, W. Y., Socher, R., Cer, D. & Manning, C. D. Bilingual word embeddings for phrase-based machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (ed. Yarowsky, D., Baldwin, T., Korhonen, A., Livescu, K. & Bethard, S.) 1393–1398 (ACL, Stroudsburg, PA, 2013).
67. Nie, A. et al. DeepTag: inferring diagnoses from veterinary clinical notes. *ngj Digit. Med.* **1**, https://doi.org/10.1038/s41746-018-0067-8 (2018).
68. Hébert-Johnson, U., Kim, M. P., Reingold, O. & Rothblum, G. N. Multicalibration: calibration for the (computationally-identifiable) masses. *Proc. Mach. Learn. Res.* **80**, 1939–1948 (2018).
69. Kearns, M., Neel, S., Roth, A. & Wu, Z. S. Preventing fairness gerrymandering: auditing and learning for subgroup fairness. *Proc. Mach. Learn. Res.* **80**, 2564–2572 (2018).
70. Kim, M. P., Ghorbani, A. & Zou, J. Multiaccuracy: black-box post-processing for fairness in classification. Preprint at https://arxiv.org/abs/1805.12317 (2018).
71. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. Preprint at https://arxiv.org/abs/1702.08608 (2017).
72. West, S. M., Whittaker, M. & Crawford, K. Discriminating systems: gender, race and power in AI. *AI Now Institute* https://ainowinstitute.org/discriminatingsystems.html (2019).

73. Wang, Y. & Young, J. E. Beyond pink and blue: gendered attitudes towards robots in society. In *Proceedings of the Second Conference on Gender and IT Appropriation*, 49–54 (European Society for Socially Embedded Technologies, Siegen, Ger., 2014).
74. Fiske, S. T. Stereotyping, prejudice, and discrimination. In *The Handbook of Social Psychology* (ed. Gilbert, D. T., Fiske, S. T. & Lindzey, G.) 357–411 (McGraw-Hill, New York, 1998).
75. Epley, N., Waytz, A. & Cacioppo, J. T. On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* **114**, 864–886 (2007). **This paper presents a psychological model of anthropomorphism, specifying three factors that elicit anthropomorphic inferences about nonhuman entities.**
76. von Zitzewitz, J., Boesch, P. M., Wolf, P. & Riener, R. Quantifying the human likeness of a humanoid robot. *Int. J. Soc. Robot.* **5**, 263–276 (2013).
77. Lemaignan, S., Fink, J., Dillenbourg, P. & Braboszcz, C. The cognitive correlates of anthropomorphism. In *2014 Human-Robot Interaction Conference: A Bridge between Robotics and Neuroscience* (2014).
78. Nass, C. & Moon, Y. Machines and mindlessness: social responses to computers. *J. Soc. Issues* **56**, 81–103 (2002).
79. Gulz, A. & Haake, M. Challenging gender stereotypes using virtual pedagogical characters. In *Gender Issues in Learning and Working with Information Technology: Social Constructs and Cultural Contexts* (ed. Goodman, S., Booth, S. & Kirkup, G.) 113–132 (IGI Global, Hershey, PA, 2010).
80. Krämer N. C. et al. Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Comput. Educ.* **99**, 1–13 (2016).
81. Baylor, A. L. The design of motivational agents and avatars. *Educ. Tech. Res. Dev.* **59**, 291–300 (2011).
82. Arroyo, I., Woolf, B. P., Royer, J. M. & Tai, M. Affective gendered learning companions. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (ed. Dimitrova, V. et al.) 41–48 (IOS Press, Amsterdam, 2009).
83. Baylor, A. L. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philos. Trans. R. Soc. B* **364**, 3559–3565 (2009).
84. McDonnell, M. & Baxter, D. Chatbots and gender stereotyping. *Interact. Comput.* https://doi.org/10.1093/iwc/iwz007 (2019).
85. Alesich, S. & Rigby, M. Gendered robots: implications for our humanoid future. *IEEE Technol. Soc. Mag.* **36**, 50–59 (2017).
86. Søraa, R. A. (2017). Mechanical genders: how do humans gender robots? *Gend. Technol. Dev.* **21**, 99–115 (2017).
87. Kraus, M., Kraus, J., Baumann, M. & Minker, W. Effects of gender stereotypes on trust and likability in spoken human-robot interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (ed. Calzolari, N. et al.) 112–118 (European Language Resources Association, Paris, 2018).

88. Crowell, C. R, Scheutz, M., Schermerhorn, P., & Villano, M. Gendered voice and robot entities: perceptions and reactions of male and female subjects. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems,* 3735–3741 (IEEE Press, Piscataway, NJ, 2009).
89. Alexander, E., Bank, C., Yang, J. J., Hayes, B. & Scassellati, B. Asking for help from a gendered robot. In *Proceedings of the Annual Meeting of the Cognitive Science Society* **36**, 2333–2338 (2014).
90. Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F. & André, E. Keep an eye on the task! How gender typicality of tasks influence human-robot interactions. *Int. J. Soc. Robot.* **6**, 417–427 (2014).
91. Reich-Stiebert, N. & Eyssel, F. (Ir)relevance of gender? On the influence of gender stereotypes on learning with a robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (ed. Mutlu, B. & Tscheligi, M.) 166–176 (ACM, New York, 2017).
92. Tay, B., Jung, Y. & Park, T. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Comput. Hum. Behav.* **38**, 75–84 (2014).
93. Jung, E. H., Waddell, T. F. & Sundar, S. S. Feminizing robots: user responses to gender cues on robot body and screen. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3107–3113 (ACM, New York, 2016).
94. Powers, A. et al. Eliciting information from people with a gendered humanoid robot. In *Proceedings of the 2005 IEEE International Workshop on Robots and Human Interactive Communication*, 158–163 (IEEE, 2005).
95. Eyssel, F. & Hegel, F. (S)he's got the look: gender stereotyping of robots. *J. Appl. Soc. Psychol.* **42**, 2213–2230 (2012).
96. Siegel, M., Breazeal, C. & Norton, M. I. Persuasive robotics: the influence of robot gender on human behavior. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems,* 2563–2568 (IEEE Press, Piscataway, NJ, 2009).
97. Eyssel, F., Kuchenbrandt, D., Hegel, F. & de Ruiter, L. Activating elicited agent knowledge: how robot and user features shape the perception of social robots. In *Proceedings of the 21st IEEE international Symposium on Robot and Human Interactive Communication*, 851–857 (IEEE, 2012).
98. Kittmann, R. et al. Let me introduce myself: I am Care-O-bot 4, a gentleman robot. In *Mensch und Computer 2015 - Proceedings* (De Gruyter Oldenbourg, Stuttgart, 2015).
99. Strait, M. K., Aguillon, C., Contreras, V. & Garcia, N. The public's perception of humanlike robots: online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. In *26th IEEE International Symposium on Robot and Human Interactive Communication*, 1418–1423 (IEEE, 2017).
100. Loideain, N. N. & Adams, R. From Alexa to Siri and the GDPR: the gendering of virtual personal assistants and the role of EU data protection law. *King's College London Dickson Poon School of Law Legal Studies Research Paper Series*, http://dx.doi.org/10.2139/ssrn.3281807 (2018).

101. West, M., Kraut, R. & Chew, H. E. *I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education* https://unesdoc.unesco.org/ark:/48223/pf0000367416 (EQUALS/UNESCO, 2019).
102. Meet Q: The First Genderless Voice. https://www.genderlessvoice.com/.
103. Nielsen, M. W., Bloch, C. W. & Schiebinger L. Making gender diversity work for scientific discovery and innovation. *Nat. Hum. Behav.* **24**, 726–734 (2018).
104. H2020 Expert Group to update and expand Gendered Innovations/Innovation through Gender. *European Commission* http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3601&NewSearch=1&NewSearch=1 (2019).
105. Lorenz, C. et al. Nanosized aerosols from consumer sprays: experimental analysis and exposure modeling for four commercial products. *J. Nanopart. Res*. **13**, 3377–3391 (2011).
106. Hines, A. et al. Comparison of histological, genetic, metabolomics, and lipid-based methods for sex determination in marine mussels. *Anal. Biochem.* **369**, 175–186 (2007).
107. Ellis, R. P. et al. ^{1}H NMR metabolomics reveals contrasting response by male and female mussels exposed to reduced seawater pH, increased temperature, and a pathogen. *Environ. Sci. Technol.* **48**, 7044–7052 (2014).
108. Naqvi, S., Godfrey, A. K., Hughes, J. F., Goodheart, M. L. & Mitchell, R. N. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, https://doi.org/10.1126/science.aaw7317 (2019).
109. Clayton, J. A. Studying both sexes: a guiding principle for biomedicine. *FASEB J.* **30**, 519–524 (2016).
110. Bivens, R. The gender binary will not be deprogrammed: ten years of coding gender on Facebook. *New Media Soc.* **19**, 880–898 (2017).
111. Pelletier, R., Ditto, B. & Pilote, L. A composite measure of gender and its association with risk factors in patients with premature acute coronary syndrome. *Psychosom. Med*. **77**, 517–526 (2015).
112. Flores, A. R., Herman, J. L., Gates, G. G., & Brown, T. N. T. How many adults identify as transgender in the United States. *The Williams Institute* https://williamsinstitute.law.ucla.edu/research/how-many-adults-identify-as-transgender-in-the-united-states/ (2016).
113. Lamm, M. S., Hui, L., Gemmell, N. J. & Godwin, J. R. The need for speed: neuroendocrine regulation of socially-controlled sex change. *Integr. Comp. Biol.* **55**, 307-322 (2015).
114. Rice, C., Harrison, E. & Friedman, M. Doing justice to intersectionality in research. *Cult. Stud. Crit. Methodol.* https://doi.org/10.1177/1532708619829779 (2019).
115. Heise et al. Gender inequality and restrictive gender norms: framing the challenges to health. *Lancet* http://dx.doi.org/10.1016/S0140-6736(19)30652-X (2019).
116. Johnson, J., Sharman, Z., Vissandjée, B. & Stewart, D. E. Does a change in health research funding policy related to the integration of sex and gender have an impact? *PLoS One* **9**, https://doi.org/10.1371/journal.pone.0099900 (2014).

117. Duchesne, A., Tannenbaum, C. & Einstein, G. Funding agency mechanisms to increase sex and gender analysis. *Lancet* **389**, 699 (2017).
118. Schiebinger, L., Leopold, S. S. & Miller, V. M. Editorial policies for sex and gender analysis. *Lancet* **388**, 2841–2842 (2016).
119. Heidari, S., Babor, T. F., De Castro, P., Tort, S. & Curno, M. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Res. Integr. Peer Rev.* **1**, https://doi.org/10.1186/s41073-016-0007-6 (2016).
120. Directorate-General for Research and Innovation. *Interim Evaluation: Gender Equality as a Crosscutting Issue in Horizon 2020* (European Commission, Brussels, 2017.)
121. Will, T. R. et al. Problems and progress regarding sex bias and omission in neuroscience research. *eNeuro* **4**, https://doi.org/10.1523/ENEURO.0278-17.2017 (2017).
122. Ludwig, S. et al. A successful strategy to integrate sex and gender medicine into a newly developed medical curriculum. *J. Women's Health* **24**, 996–1005 (2015).
123. Schiebinger, L.et al. Gendered Innovations in Science, Health & Medicine, Engineering, and Environment. www.genderedinnovations.stanford.edu (2011–2019).
124. World Health Organization. *Gender Mainstreaming for Health Managers: A Practical Approach.* (World Health Organization, Geneva, 2011) https://www.who.int/gender-equity-rights/knowledge/health_managers_guide/en.
125. Organization for the Study of Sex Differences. https://www.ossdweb.org.
126. Online Training Modules: Integrating Sex & Gender in Health Research. *Canadian Institutes of Health Research* http://www.cihr-irsc.gc.ca/e/49347.html (2017).
127. Arboleda, V. A., Sandberg, D. E. & Vilain, E. DSDs: genetics, underlying pathologies and psychosexual differentiation. *Nat. Rev. Endocrinol.* **10**, 603–615 (2014).
128. Huges, I. A., Houk, C., Ahmed, S. F. & Lee, P. A. Consensus statement on management of intersex disorders. *J. Pediatr. Urol.* **2**, 148–162 (2006).
129. Jarne, P. & Auld, J. R. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evol.* **60**, 1816-1824 (2006).
130. Terms: gender. *Gendered Innovations*. http://genderedinnovations.stanford.edu/terms/gender.html (2011-2019).
131. Boerner, K. E. et al. Conceptual complexity of gender and its relevance to pain. *Pain* **159**. 2137–2141 (2018).
132. Mauvais-Jarvis, F., Arnold, A. P. & Reue, K. A guide for the design of pre-clinical studies on sex differences in metabolism. *Cell Metab.* **25**, 1216–1230 (2017).

Acknowledgements We thank H. LeBlanc for her assistance. R.E. acknowledges financial support from a NERC Industrial Innovation Fellowship (NE/R013241/1). J.Z. is supported by a Chan-Zuckerberg Investigator Award and NIH P30AG059307.

Author Contributions L.S. conceptualised the paper and invited R.E., F.E., C.T., and J.Z. to collaborate. L.S. and C.T. structured and drafted the article, R.E. wrote the

marine science section, F.E. wrote the social robots section, L.S. wrote the introductory and policy sections, C.T. wrote the health and medicine sections, and J.Z. wrote the machine learning section, all authors commented and revised. R.E. conceived and developed figure 1, C.T. conceived and developed figures 2 and 3, and contributed to figure 1, L.S. contributed to figure 3 and developed figure 4.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.S. (schiebinger@stanford.edu).

Reviewer Information *Nature* thanks….

Fig. 1 Hazards of pooling data from both sexes. *Pooling data across sex not only assumes no difference between males and females, but subsequently prevents researchers from testing for the dependency of an experimental response on the sex of a subject. The theoretical examples reveal that pooling (green circles) masks important male (orange triangles) and female (blue squares) differences in baseline data, treatment response and sex x treatment interactions—any one of which leads to misinterpretation in results. Experimental data demonstrate one example where pooling would have masked both the sex difference in the respiration rate of copepods, as well as the response of this variable to elevated* pCO_2*. Theoretical example generated using hypothetical data; experimental data taken from Cripps et al.*[12].

Fig. 2 Sex analysis and reporting in science & engineering. *This decision tree represents a cognitive process for analyzing sex. A "no" indicates no further analysis is necessary. A "yes" suggests a next step.*

Fig. 3 Gender analysis and reporting in science & engineering. *This decision tree represents a cognitive process for analyzing gender. A "no" indicates no further analysis is necessary. A "yes" suggests a next step.*

Fig. 4 Three pillars of science & engineering infrastructure*. To reap the benefits of sex and gender analysis, the pillars of science infrastructure must develop and implement coordinate policies.*